

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 15, NOMOR 1, JUNI 2023 ISSN 2086 – 4132

Pemodelan Dampak Pandemi Covid-19 Terhadap PDRB Sektor Penyediaan Akomodasi Dan Makan Minum Provinsi Bali (Implementasi Model Arima Intervensi)

MUHAMMAD ZIYAD AHMAD, ERNI TRI ASTUTI

Perancangan Prototipe Web Diseminasi Sensus Pertanian 2023 Dengan *Responsive Web Design*

FATURROKHMAN, FARID RIDHO

Peramalan Kasus Harian Monkeypox Dunia Dengan Pendekatan *Support Vector Regression*

MARCEL LAVERDA SUBIYANTO, YULIA AMANDA, MUHAMMAD NADHIL FACHRIAN, AFRIANI, ACHMAD YAZID BUSTHOMI ROHIM, NUR CHAMIDAH

Penerapan *Model-Based Clustering* Pada Pengelompokan Saham Berdasarkan Rasio Keuangan

IRENA SEKAR DWI HASNIDA, ROSITA KUSUMAWATI

Pengembangan Laboratorium Statistika Berbasis Android

YUSRIANTI HANIKE, SITI NURJANNAH

*Generating Synthetic Training Datasets Using Conditional Generative Adversarial Network To Improve Images Segmentation*

IFFATI UZMA, RANI NOORAENI, TAKDIR

Karakteristik Spasio temporal Angin Kencang Di Pulau Sulawesi Berbasis Model Reanalisis Era5

ACHMAD FAHRUDIN RAIS, REZKY YUNITA, HIJRAH KULE MUSGAMY, RAHMAT TAULADANI, AHMAD KOSASIH



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.

<b>Editor in Chief:</b>	Rani Nooraeni, SST., M.Stat.
<b>Managing Editor:</b>	Dr. Fitri Kartiasih, SST., S.E., M.Si.
<b>Editor:</b>	Prof. Dr. Hardius Usman, S.Si., M.Si. Prof. Setia Pramana, S.Si., M.Sc, Ph.D Dr. Achmad Syahrul Choir Lutfi Rahmatuti Maghfiroh, SST., M.T.
<b>Guest Editor:</b>	Firman M. Firmansyah, M.Sc., Ph.D Muhammad Rausyan Fikri, ST, M.Eng
<b>Assistant Editor dan IT:</b>	Geri Yesa Ermawan, S.Tr.Stat. Muhammad Luqman, S.Tr.Stat.
<b>Peer Reviewers:</b>	Dr. Eng. Arie Wahyu Wijayanto Dr. Eng. Anang Yunarso Dr. Eng. Lya Hulliyyatus Suadaa, SST, M.T Dr. Sarni Maniar Berliana Siti Mariyah, M.T Siskarossa Ika Oktora, SST, M.Stat Ibnu Santoso, SST, MT Nori Wilantika, S.S.T., M.T.I.

## **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 15, Nomor 1, Juni 2023 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, Juni 2023

Ketua Dewan Redaksi,

**Rani Nooraeni, SST., M.Stat.**

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 15, NOMOR 1, JUNI 2023

## DAFTAR ISI

<b><u>Pengantar Redaksi</u></b> .....	iii
<b><u>Daftar Isi</u></b> .....	iv
<b><u>Abstrak</u></b> .....	v-xii
<b>Pemodelan Dampak Pandemi Covid-19 Terhadap PDRB Sektor Penyediaan Akomodasi Dan Makan Minum Provinsi Bali (Implementasi Model Arima Intervensi)</b> <u>Muhammad Ziyad Ahmad, Erni Tri Astuti</u> .....	1-14
<b>Perancangan Prototipe Web Diseminasi Sensus Pertanian 2023 Dengan <i>Responsive Web Design</i></b> <u>Faturrokhman, Farid Ridho</u> .....	15-26
<b>Peramalan Kasus Harian Monkeypox Dunia Dengan Pendekatan <i>Support Vector Regression</i></b> <u>Marcel Laverda Subiyanto, Yulia Amanda, Muhammad Nadhil Fachrian, Afriani, Achmad Yazid Busthomi Rohim, Nur Chamidah</u> .....	27-36
<b>Penerapan <i>Model-Based Clustering</i> Pada Pengelompokan Saham Berdasarkan Rasio Keuangan</b> <u>Irena Sekar Dwi Hasnida, Rosita Kusumawati</u> .....	37-50
<b>Pengembangan Laboratorium Statistika Berbasis Android</b> <u>Yusrianti Hanike, Siti Nurjannah</u> .....	51-60
<b>Generating Synthetic Training Datasets Using Conditional Generative Adversarial Network to Improve Images Segmentation</b> <u>Iffati Uzma, Rani Nooraeni, Takdir</u> .....	67-72
<b>Karakteristik Spasiotemporal Angin Kencang Di Pulau Sulawesi Berbasis Model Reanalisis Era5</b> <u>Achmad Fahrudin Rais, Rezky Yunita, Hijrah Kule Musgamy, Rahmat Tauladani, Ahmad Kosasih</u> .....	73-78

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

DDC: 315.98

Muhammad Ziyad Ahmad, Erni Tri Astuti

Pemodelan Dampak Pandemi Covid-19 Terhadap PDRB Sektor Penyediaan Akomodasi Dan Makan Minum Provinsi Bali (Implementasi Model Arima Intervensi)

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 1 – 14

#### Abstrak

Pandemi Covid-19 yang terjadi menyebabkan penurunan perekonomian di Indonesia. Hal ini juga terjadi di Bali terutama pada sektor penyediaan akomodasi dan makan minum yang merupakan salah satu kontributor terbesar dalam PDRB Provinsi Bali. Oleh karena itu, penelitian ini bertujuan untuk mengetahui waktu terjadi dan seberapa besar dampak pandemi Covid-19 terhadap sektor penyediaan akomodasi dan makan minum. Penelitian ini menggunakan metode analisis intervensi dengan fungsi *step*. Hasil penelitian menunjukkan bahwa pandemi Covid-19 memberikan dampak negatif secara langsung yang signifikan terhadap PDRB sektor ini pada triwulan I 2020. Dampak negatif pandemi Covid-19 yang ditimbulkan hampir mencapai 50 persen. Dampak pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum permanen hingga akhir tahun 2021. Penelitian ini diharapkan dapat mampu membantu pemerintah dan pelaku usaha sektor terkait untuk melakukan tindakan yang diharapkan mampu mengurangi dampak dari pandemi Covid-19.

Kata kunci: Covid-19, analisis intervensi, akomodasi dan makan minum

DDC: 315.98

Faturrokhman, Farid Ridho

Perancangan Prototipe Web Diseminasi Sensus Pertanian 2023 Dengan *Responsive Web Design*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 15 – 26

#### Abstrak

Setelah pelaksanaan cacah lengkap dalam kegiatan Sensus Pertanian 2023, BPS (Badan Pusat Statistik) berkewajiban untuk menyajikan data yang telah diperoleh melalui diseminasi data, salah satunya menggunakan situs web. Namun, web diseminasi sensus sebelumnya, ST2013, hanya mendapatkan hasil *usability testing* sebesar 66,19% dan pengguna perangkat seluler kesulitan dalam melihat informasi di dalam web. Maka dari itu, diperlukan web diseminasi ST2023 yang memiliki desain responsif di semua ukuran layar. Untuk mengatasi masalah tersebut, peneliti melakukan penelitian untuk merancang Prototipe Web Diseminasi Sensus Pertanian 2023 yang responsif menggunakan teknik *Responsive Web Design*. Prototipe kemudian dievaluasi menggunakan *usability testing* dan mendapatkan hasil sebesar 81,64%. Uji coba responsif dan kompatibilitas pada berbagai browser juga telah dilaksanakan dan mendapatkan hasil bahwa prototipe sudah responsif di semua ukuran layar dan tampil dengan baik di sebagian besar browser di perangkat yang berbeda.

Kata kunci: Prototipe, Diseminasi, Sensus Pertanian 2023, *Responsive Web Design*

DDC: 315.98

Marcel Laverda Subiyanto, Yulia Amanda, Muhammad Nadhil Fachrian, Afriani, Achmad Yazid Busthomi Rohim, Nur Chamidah

Peramalan Kasus Harian Monkeypox Dunia Dengan Pendekatan *Support Vector Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 27 – 36

Abstrak

*Monkeypox* saat ini menjadi perhatian masyarakat global. Maka, penting untuk mengetahui perkembangan jumlah kasus *monkeypox* kedepannya. Pada penelitian ini dilakukan peramalan kasus harian *monkeypox* menggunakan metode *Support Vector Regression* (SVR) dengan Fungsi *Kernel Radial Basis Function* (RBF). Data yang digunakan adalah data sekunder berupa deret waktu harian mulai 29 Mei sampai 20 Oktober 2022. Untuk memperoleh parameter optimal pada model SVR, peneliti menggunakan algoritma *grid search* untuk memprediksi data *testing* secara akurat. Berdasarkan hasil pencarian parameter optimal menggunakan algoritma *grid search* diperoleh parameter optimalnya adalah  $C = 2^{0,6}$ ,  $\gamma = 2^{0,2}$ , dan  $\varepsilon = 0,09$ . Nilai RMSE untuk hasil peramalan menggunakan model SVR dengan parameter optimal pada data *training* dan *testing* sebesar 352,3 dan 809,7.

Kata kunci: Monkeypox, Peramalan, SVR, RBF, *Grid Search*

DDC: 315.98

Irena Sekar Dwi Hasnida, Rosita Kusumawati

Penerapan *Model-Based Clustering* Pada Pengelompokan Saham Berdasarkan Rasio Keuangan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 37 – 50

Abstrak

Untuk meminimalkan kerugian dengan tingkat keuntungan tertentu, investor perlu memilih saham potensial agar keuntungan yang diperoleh optimal. Penelitian ini bertujuan untuk mengetahui penerapan *Model-Based Clustering* (MBC) dalam mengelompokkan perusahaan berdasarkan kinerja keuangan saham. Indikator keuangan yang digunakan yaitu data rasio likuiditas, profitabilitas, dan solvabilitas tahun 2020 untuk perusahaan yang terdaftar pada indeks LQ45. Pemilihan model dan jumlah komponen akan dilakukan berdasarkan nilai *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL). Dari proses *clustering*, terbentuk 6 *cluster* dengan nilai BIC dan ICL berturut-turut sebesar 709,3757 dan -709,376, dan model optimal terpilih VEV (*Variable volume, Equal shape, Variable orientation*). Berdasarkan nilai rata-rata setiap rasio, *cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terendah. *Cluster* 6 memiliki kemampuan yang tinggi dibandingkan perusahaan *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban keuangannya jangka pendek.

Kata kunci: *Model-Based Clustering*, pengelompokan saham, rasio keuangan.

DDC: 315.98

Yusrianti Hanike, Siti Nurjannah

Pengembangan Laboratorium Statistika Berbasis Android

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 51 – 60

Abstrak

Pentingnya praktikum sebagai wadah dalam penjelasan teoritis memiliki dukungan yang kuat dalam proses transfer pengetahuan, maka perlu diejawantahkan pada proses pembelajaran dalam bentuk yang mudah diakses oleh mahasiswa. Pengembangan laboratorium berbasis statistika adalah salah satu solusi untuk memudahkan mahasiswa dalam mengakses *software* statistika yang notabene sulit dan memiliki kapasitas yang besar. Tujuan pengembangan laboratorium statistika berbasis *Android* untuk mengimplementasikan keilmuan statistika dan didukung dengan uji kepraktisan dan keefektifan aplikasi pada mahasiswa dan dosen. Metode penelitian yang dilakukan melalui lima tahap yakni penilaian, perancangan, pengembangan, implementasi, dan evaluasi. Sampel penelitian diambil dari dua program studi (prodi) di dua perguruan tinggi di Ambon yakni prodi Statistika Universitas Pattimura dan prodi Ekonomi Syariah Institut Agama Islam Negeri (IAIN) Ambon. Hasil penelitian menunjukkan Laboratorium Statistika berbasis *Android* memiliki persentase yang cukup signifikan terhadap kelayakan, kepraktisan dan keefektifan praktikum statistika dengan penilaian kelayakan sebesar 71%, kepraktisan sebesar 79% dan 75% menunjukkan keefektifan penggunaan aplikasi.

Kata kunci: Praktikum, Laboratorium Statistika, Uji Kelayakan, Uji Kepraktisan, Uji Keefektifan.

DDC: 315.98

Iffati Uzma, Rani Nooraeni, Takdir

Generating Synthetic Training Datasets Using Conditional Generative Adversarial Network To Improve Images Segmentation

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 60– 72

Abstrak

Dataset pelatihan dalam jumlah terbatas dalam penelitian pembelajaran mendalam dapat memengaruhi keakuratan model yang dihasilkan. Situasi ini dapat menyebabkan overfit, sehingga model tidak dapat bekerja dengan baik. Conditional Generative Adversarial Network (CGAN) diperkenalkan untuk menghasilkan data sintetik dengan mempertimbangkan kondisi tertentu. Penelitian ini bertujuan untuk menghasilkan dataset training sintetik tambahan untuk meningkatkan akurasi model segmentasi objek citra. Pertama, kami mengevaluasi akurasi generator dataset berbasis CGAN terhadap beberapa dataset terbuka. Kemudian, kami menerapkan generator untuk melatih dua model segmentasi objek, yaitu FCN dan CNN U-Net. Evaluasi kami menunjukkan bahwa CGAN dapat menghasilkan dataset sintetik dengan baik. Kumpulan data yang kompleks memerlukan iterasi pelatihan yang lebih banyak. Ini juga meningkatkan kehilangan dan akurasi validasi kedua model segmentasi, meskipun metrik lainnya masih memerlukan peningkatan lebih lanjut.

Kata kunci: *deep learning*, data sintetik, cGAN, U-net, segmentasi

DDC: 315.98

Achmad Fahrudin Rais, Rezky Yunita,  
Hijrah Kule Musgamy, Rahmat Tauladani,  
Ahmad Kosasih

Karakteristik Spasiotemporal Angin  
Kencang Di Pulau Sulawesi Berbasis  
Model Reanalisis Era5

Jurnal Aplikasi Statistika & Komputasi  
Statistik, Volume 15, Nomor 1, Juni 2023,  
hal 73 – 78

#### Abstrak

Pola spasiotemporal angin kencang di Pulau Sulawesi dieksplorasi dalam tulisan ini menggunakan data reanalisis ERA5 periode 1979-2020. Metode yang digunakan adalah *roteted empirical orthogonal function* (REOF) dan fourier harmonik. Hasil yang didapatkan adalah bahwa frekuensi dan magnitudo kecepatan angin terbesar berada di Sulawesi Selatan dan sebagian Sulawesi Tenggara dengan puncak di bulan Agustus dan Januari, pola dominan lainnya adalah di Sulawesi Utara dan Gorontalo dengan pola puncak yang sama yaitu Agustus dan Januari dan di pegunungan tengah Sulawesi dengan puncak pada bulan Januari.

Kata kunci: angin kencang, ERA5, REOF



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

DDC: 315.98

Muhammad Ziyad Ahmad, Erni Tri Astuti

Pemodelan Dampak Pandemi Covid-19 Terhadap PDRB Sektor Penyediaan Akomodasi Dan Makan Minum Provinsi Bali (Implementasi Model Arima Intervensi)

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 1 – 14

*Abstract*

*The Covid-19 pandemic has caused a decline in the economy in Indonesia. This also happened in Bali, especially in the accommodation and food service sector, which is one of the largest contributors to the GRDP of the Province of Bali. Therefore, this study aims to determine when it occurred and how much impact the Covid-19 pandemic had on the accommodation and food service sector. This study uses an intervention analysis method with a step function. The results showed that the Covid-19 pandemic had a significant direct negative impact on the GRDP of this sector in the first quarter of 2020. The negative impact of the Covid-19 pandemic caused almost 50 percent. The impact of the Covid-19 pandemic on the GRDP of the accommodation and food service sector is permanent until the end of 2021. This research is expected to be able to help the government and related sector business actors to take actions that are expected to reduce the impact of the Covid-19 pandemic.*

*Keywords: Covid-19, intervention analysis, accommodation and food service*

DDC: 315.98

Faturrokhman, Farid Ridho

Perancangan Prototipe Web Diseminasi Sensus Pertanian 2023 Dengan *Responsive Web Design*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 15 – 26

*Abstract*

*After carrying out the complete enumeration in the 2023 Agricultural Census, BPS is obliged to present the data that has been obtained through data dissemination, one of which uses a website. However, the previous census web dissemination, ST2013, only got usability testing results of 66.19% and mobile device users had difficulty viewing information on the web. Therefore, we need a ST2023 web dissemination that has a responsive design on all screen sizes. To overcome this problem, researchers conducted a study to design a responsive 2023 Agricultural Census Dissemination Web Prototype using Responsive Web Design techniques. The prototype was then evaluated using usability testing and got a result of 81.64%. Responsiveness and compatibility tests on various browsers have also been carried out and got the results that the prototype is responsive on all screen sizes and performs well in most browsers on different devices.*

*Keywords: Prototype, Dissemination, 2023 Agricultural Census, Responsive Web Design*

DDC: 315.98

Marcel Laverda Subiyanto, Yulia Amanda, Muhammad Nadhil Fachrian, Afriani, Achmad Yazid Busthomi Rohim, Nur Chamidah

Peramalan Kasus Harian Monkeypox Dunia Dengan Pendekatan *Support Vector Regression*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 27 – 36

*Abstract*

*Monkeypox is a disease that becomes public health emergency of international concern. It is important to know the development of the number of monkeypox cases in the future. In this study, forecasting monkeypox daily cases was carried out using the Support Vector Regression (SVR) method with the Kernel Radial Basis Function (RBF). The data used is secondary data from May 29<sup>th</sup> to October 20<sup>th</sup> 2022. To obtain optimal parameters in the SVR model, researcher uses a grid search algorithm to accurately predict the testing data. Based on the search results for optimal parameters using the grid search algorithm, the optimal parameters are  $C = 2^{0.6}$ ,  $\gamma = 2^{0.2}$ , and  $\varepsilon = 0.09$ . RMSE values for forecasting using the SVR model with optimal parameters on training and testing data are 352.3 and 809.7.*

*Keywords: Monkeypox, Forecasting, SVR, RBF, Grid Search*

DDC: 315.98

Irena Sekar Dwi Hasnida, Rosita Kusumawati

Penerapan *Model-Based Clustering* Pada Pengelompokan Saham Berdasarkan Rasio Keuangan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 37 – 50

*Abstract*

*To minimize losses at a certain profit level, investors need to choose potential stocks so that the profits are optimal. This study aims to determine the application of model-based clustering (MBC) in classifying companies based on stock performance. The financial indicators used are data on liquidity ratios, profitability, and solvency for 2020 for companies listed on the LQ45 index. The selection of the model and the number of components will be based on the Bayesian Information Criterion (BIC) and Completed Likelihood Criterion (ICL) values. From the clustering process, six clusters were formed with BIC and ICL values of 709.3757 and -709.376, respectively, and the optimal model was selected as VEV (variable volume, equal shape, variable orientation). Based on the average value of each ratio, cluster 6 is the best cluster because it has the majority of the best liquidity and profitability ratios and the lowest solvency ratio. Cluster 6 has a high ability, compared to other cluster companies, to finance the company's operational activities and fulfill its short-term financial obligations.*

*Keywords: Model-Based Clustering, stock clustering, financial ratios.*

DDC: 315.98

Yusrianti Hanike, Siti Nurjannah

Pengembangan Laboratorium Statistika Berbasis Android

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 51 – 60

*Abstract*

*The importance of practicum as a forum for theoretical explanations has strong support in the knowledge transfer process, so it needs to be embodied in the learning process in a form that is easily accessible to students. The development of a statistics-based laboratory is one of the solutions to make it easier for students to access statistical software which is difficult and has a large capacity. The aim of developing an Android-based statistics laboratory is to implement statistical science and is supported by practicality and effectiveness tests for students and lecturers. The research method is carried out through five stages, namely assessment, design, development, implementation, and evaluation. The research samples were taken from two study programs (prodi) at two tertiary institutions in Ambon, namely the Statistics study program at Pattimura University and the Sharia Economics study program at the State Islamic Institute (IAIN) Ambon. The results showed that the Android-based Statistics Laboratory had a significant percentage of the feasibility, practicality and effectiveness of statistical practicum with a feasibility rating of 71%, practicality of 79% and 75% indicating the effectiveness of using the application.*

*Keywords: Practicum, Statistics Laboratory, Feasibility Test, Practicality Test, Effectiveness Test*

DDC: 315.98

Iffati Uzma, Rani Nooraeni, Takdir

Generating Synthetic Training Datasets Using Conditional Generative Adversarial Network To Improve Images Segmentation

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 15, Nomor 1, Juni 2023, hal 60– 72

*Abstract*

*A limited amount of training datasets in deep learning research could impact the accuracy of the resulting models. This situation can cause overfit, so the model cannot work correctly. Conditional Generative Adversarial Network (CGAN) was introduced to generate synthetic data by considering certain conditions. This study aims to generate additional synthetic training datasets to improve the accuracy of the object segmentation model of images. Firstly, we evaluated CGAN-based dataset generator accuracy against several open datasets. Then, we applied the generator to train two object segmentation models, i.e., FCN and CNN U-Net. Our evaluation shows that CGAN can generate synthetic datasets well. Complex datasets require more training iterations. It also improves both segmentation models' validation loss and accuracy, although other metrics still need further improvement.*

*Keywords: deep learning, synthetic data, cGAN, U-net, segmentation*

DDC: 315.98

Achmad Fahrudin Rais, Rezky Yunita,  
Hijrah Kule Musgamy, Rahmat Tauladani,  
Ahmad Kosasih  
Karakteristik Spasiotemporal Angin  
Kencang Di Pulau Sulawesi Berbasis  
Model Reanalisis Era5

Jurnal Aplikasi Statistika & Komputasi  
Statistik, Volume 15, Nomor 1, Juni 2023,  
hal 73 – 78

*Abstract*

*The spatiotemporal pattern of wind gusts over Sulawesi Island was explored in this paper using ERA5 reanalysis for the period 1979-2020. The methods used were rotated empirical orthogonal function (REOF) and Fourier harmonics. The results obtained were that the highest frequency and magnitude of wind speed were in South Sulawesi and parts of Southeast Sulawesi with peaks in August and January, the other dominant pattern were in North Sulawesi and Gorontalo with the same peak pattern, namely August and January and in the central mountains of Sulawesi with a peak in January.*

*Keywords:* angin kencang, ERA5, REOF

# PEMODELAN DAMPAK PANDEMI COVID-19 TERHADAP PDRB SEKTOR PENYEDIAAN AKOMODASI DAN MAKAN MINUM PROVINSI BALI (IMPLEMENTASI MODEL ARIMA INTERVENSI)

Muhammad Ziyad Ahmad<sup>1</sup>, Erni Tri Astuti<sup>2</sup>

<sup>1</sup>Politeknik Statistika STIS

e-mail: 211810478<sup>1</sup>@stis.ac.id

*Received: November 15, 2022; Revised: June 27, 2023; Accepted: June 28, 2023; Published: June 30, 2023*

## Abstrak

Pandemi Covid-19 yang terjadi menyebabkan penurunan perekonomian di Indonesia. Hal ini juga terjadi di Bali terutama pada sektor penyediaan akomodasi dan makan minum yang merupakan salah satu kontributor terbesar dalam PDRB Provinsi Bali. Oleh karena itu, penelitian ini bertujuan untuk mengetahui waktu terjadi dan seberapa besar dampak pandemi Covid-19 terhadap sektor penyediaan akomodasi dan makan minum. Penelitian ini menggunakan metode analisis intervensi dengan fungsi *step*. Hasil penelitian menunjukkan bahwa pandemi Covid-19 memberikan dampak negatif secara langsung yang signifikan terhadap PDRB sektor ini pada triwulan I 2020. Dampak negatif pandemi Covid-19 yang ditimbulkan hampir mencapai 50 persen. Dampak pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum permanen hingga akhir tahun 2021. Penelitian ini diharapkan dapat mampu membantu pemerintah dan pelaku usaha sektor terkait untuk melakukan tindakan yang diharapkan mampu mengurangi dampak dari pandemi Covid-19.

**Kata kunci:** Covid-19, analisis intervensi, akomodasi dan makan minum

## Abstract

*The Covid-19 pandemic has caused a decline in the economy in Indonesia. This also happened in Bali, especially in the accommodation and food service sector, which is one of the largest contributors to the GRDP of the Province of Bali. Therefore, this study aims to determine when it occurred and how much impact the Covid-19 pandemic had on the accommodation and food service sector. This study uses an intervention analysis method with a step function. The results showed that the Covid-19 pandemic had a significant direct negative impact on the GRDP of this sector in the first quarter of 2020. The negative impact of the Covid-19 pandemic caused almost 50 percent. The impact of the Covid-19 pandemic on the GRDP of the accommodation and food service sector is permanent until the end of 2021. This research is expected to be able to help the government and related sector business actors to take actions that are expected to reduce the impact of the Covid-19 pandemic.*

**Keywords:** Covid-19, intervention analysis, accommodation and food service

## PENDAHULUAN

*Severe Acute Respiratory Syndrome Coronavirus* (SARS-CoV-2) (Gorbalenya et al., 2020) merupakan virus yang menyebabkan kasus pandemi penyakit pernafasan akut yang disebut sebagai *Coronavirus disease 2019* (Covid-19) (Stawicki et al., 2020). Virus ini dapat menular dari manusia ke manusia (Chan et al., 2020). Virus ini pertama kali ditemukan di Wuhan, Tiongkok pada bulan Desember 2019. World Health Organization (WHO) mendeklarasikan kondisi merebaknya virus SARS-CoV-2 ini sebagai *Public Health Emergency of International Concern* (PHEIC) atau darurat kesehatan masyarakat yang menjadi perhatian Internasional pada tanggal 30 Januari 2020 (World Health Organization, 2020a). WHO kemudian mendeklarasikan merebaknya virus ini menjadi pandemi pada tanggal 11 Maret 2020 (World Health Organization, 2020b). Di Indonesia sendiri, kasus Covid-19 pertama kali ditemukan pada tanggal 2 Maret 2020 (Kompas.com, 2020a). Pemerintah selanjutnya berusaha mengantisipasi penyebaran Covid-19 dengan menerapkan kebijakan Pembatasan Sosial Berskala Besar (PSBB) melalui pengesahan Peraturan Pemerintah No. 21 Tahun 2020 (Kompas.com, 2020b).

Merebaknya Covid-19 di Indonesia lalu diikuti kebijakan pembatasan kegiatan yang dilakukan pemerintah membuat perekonomian Indonesia menjadi terdampak. Menurut data Badan Pusat Statistik (BPS), pertumbuhan ekonomi Indonesia (y-o-y) anjlok pada triwulan II tahun 2020 yaitu sebesar -5,32 persen. Penurunan pertumbuhan ekonomi Indonesia terus berlanjut pada triwulan III tahun 2020 yaitu sebesar -3,49 persen sehingga Indonesia resmi memasuki jurang resesi. Penurunan pertumbuhan ekonomi ini dapat dilihat juga berdasarkan sektor-sektor penyusun produk domestik bruto (PDB). Berdasarkan lapangan usahanya, sebanyak 10 dari 17 mengalami penurunan pada triwulan II 2020 dan triwulan III 2020 secara y-o-y (BPS, 2020).

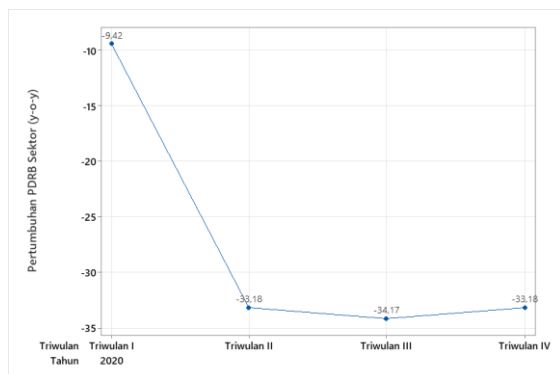
Salah satu sektor yang terkena dampak penurunan yang cukup besar pada masa

pandemi adalah sektor penyediaan akomodasi dan makan minum. Berdasarkan data BPS sektor ini mengalami penurunan sebesar -21,91 persen pada triwulan II 2020 dan penurunan sebesar -11,81 persen pada triwulan III 2020 (y-o-y) (BPS, 2020). Hal ini juga diperkuat dengan hasil Survei Dampak Covid-19 terhadap Pelaku Usaha yang dilaksanakan pada Juli 2020, bahwa sektor penyediaan akomodasi dan makan minum merupakan sektor usaha paling terdampak Covid-19 dengan sebanyak 92,47 persen pelaku usaha sektor tersebut mengalami penurunan pendapatan.

Provinsi Bali merupakan salah satu provinsi yang mengandalkan sektor pariwisata. Provinsi Bali merupakan salah satu provinsi yang mengandalkan sektor pariwisata. Menurut Kepala Asosiasi Perusahaan Perjalanan Wisata Indonesia (ASITA) Bali, 80 persen warga Bali tergantung kepada sektor Pariwisata (ABC News, 2020). Sebelum terjadinya pandemi Covid-19, Provinsi Bali memberikan kontribusi sebesar hampir 40 persen dari keseluruhan wisatawan mancanegara yang berkunjung ke Indonesia. Jumlah wisatawan mancanegara yang berkunjung ke Bali meningkat setiap tahunnya dengan kunjungan sekitar 6,3 juta wisatawan pada tahun 2019 (Bappenas, 2021). Namun sejak pandemi Covid-19 terjadi, pariwisata di Bali mengalami penurunan. Penurunan ini dapat ditunjukkan dari turunnya kunjungan wisatawan ke Bali. Wisatawan mancanegara yang datang ke Bali sebelum pandemi Covid-19 terjadi sebesar ratusan ribu tiap bulannya. Namun ketika pandemi Covid-19 terjadi, kunjungan wisatawan mancanegara mengalami penurunan yang sangat drastis. Pada bulan April 2020, penurunan wisatawan mancanegara sebesar -99,92 persen atau dari 477.969 orang menjadi 106 orang secara y-o-y. Penurunan sangat drastis terus terjadi selama pandemi Covid-19, menyebabkan kunjungan wisatawan ke Bali yang biasanya berkisar di ratusan ribu tiap bulannya menjadi hanya puluhan atau ratusan orang saja tiap bulannya (BPS Provinsi Bali, 2021a). Penurunan juga terjadi pada kunjungan wisatawan domestik. Sebelum terjadinya

pandemi Covid-19 wisatawan domestik yang datang ke Bali dapat mencapai lebih dari 700 ribu orang. Bahkan ketika musim libur, angka kunjungan wisatawan domestik dapat mencapai lebih dari 1 juta orang. Namun ketika pandemi Covid-19 terjadi, kunjungan wisatawan domestik mengalami penurunan dengan penurunan yang cukup besar terjadi pada April 2020 sebesar -78 persen atau dari 795.997 orang menjadi 175.120 orang (BPS Provinsi Bali, 2021c).

Menurut Klasifikasi Leiper dalam Pitana dan Diarta (2009) sektor pariwisata disusun oleh beberapa sektor termasuk sektor penyediaan akomodasi dan makan minum. Berdasarkan data BPS Provinsi Bali, pada periode 2010-2021, sektor penyediaan akomodasi dan makan minum menyumbang sekitar 20 persen PDRB Provinsi Bali (BPS Provinsi Bali, 2022). Hal ini menunjukkan bahwa sektor penyediaan akomodasi dan makan minum merupakan sektor andalan di Provinsi Bali. Namun sejak pandemi Covid-19, sektor ini terus mengalami penurunan. Hal ini ditunjukkan oleh grafik di bawah ini



Gambar 1. Pertumbuhan PDRB Sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali Tahun 2020

Gambar 1 menunjukkan bahwa pertumbuhan sektor penyediaan akomodasi dan makan minum di Provinsi Bali menurun sepanjang tahun 2020 secara y-o-y. Masalah ini salah satunya disebabkan oleh kunjungan wisatawan mancanegara ke Bali mulai turun drastis pada bulan April 2020 sebesar -99,92 persen atau dari 477.969 orang menjadi 106 orang secara y-o-y (BPS Provinsi Bali, 2021a) dan kunjungan

wisatawan domestik ke Bali pada bulan April 2020 turun sebesar -78 persen atau dari 795.997 orang menjadi 175.120 orang (BPS Provinsi Bali, 2021c). Kondisi penurunan yang terus menerus terjadi membuat suatu penelitian untuk menghitung dampak akibat pandemi Covid-19 perlu dilakukan. Hal ini agar pemerintah sebagai pengambil kebijakan dapat menentukan kebijakan yang tepat untuk meminimalisasi dampak yang terjadi. Penelitian seperti ini juga bermanfaat bagi pelaku usaha agar dapat menentukan strategi yang tepat agar usaha yang dijalankan dapat terus bertahan.

Penelitian tentang dampak pandemi Covid-19 terhadap suatu variabel tertentu menggunakan metode ARIMA intervensi telah banyak dilakukan. Penelitian yang dilakukan oleh Rianda (2021) bertujuan untuk menganalisis dan meramalkan jumlah penumpang pesawat melalui Bandara Soekarno-Hatta akibat pandemi Covid-19. Metode yang digunakan adalah ARIMA intervensi. Hasil dari penelitian ini menunjukkan bahwa dampak penurunan akibat pandemi Covid-19 langsung terjadi secara signifikan sebesar 43,48%. Manoj dan Philip (2021) melakukan penelitian yang bertujuan untuk mengetahui dampak pandemi Covid-19 terhadap pasar saham di negara terpilih yaitu India, Tiongkok, Singapura, dan Amerika Serikat. Data yang digunakan adalah data harga penutupan saham negara tersebut pada Maret 2020 sampai Juli 2020. Metode yang digunakan adalah ARIMA dengan analisis intervensi. Hasil dari penelitian ini menunjukkan bahwa seluruh indeks saham berjalan pada tren negatif, dan tingkat pemulihannya juga sangat lambat bahkan setelah mencapai nilai pengembalian (*return*) yang paling kecil. Temuan tentang dampak menunjukkan bahwa India menerima guncangan (*shock*) negatif tertinggi sementara Tiongkok paling sedikit. Praporn dan Laipaporn (2021) melakukan penelitian yang bertujuan untuk mengetahui dampak dari suatu insiden terhadap kedatangan turis asal Malaysia ke Provinsi Songkhla di Thailand. Data yang digunakan merupakan data kedatangan turis asal

Malaysia ke Provinsi Songkhla di Thailand pada Januari 2004 sampai Desember 2019. Metode yang digunakan adalah SARIMA dengan analisis intervensi. Model intervensi menunjukkan bahwa empat kejadian, yaitu pemberontakan, kerusakan politik, bencana alam, dan wabah penyakit memberikan pengaruh yang signifikan terhadap jumlah kedatangan. Sebagian besar kejadian intervensi menurunkan jumlah kedatangan secara signifikan dan menurunkan pendapatan pariwisata sekitar 370 hingga 1.031 juta Baht. Namun, penelitian ini secara mengejutkan menemukan bahwa dua insiden pemberontakan yang terjadi di Narathiwat, provinsi tetangga Songkhla, secara signifikan menyebabkan peningkatan sementara kedatangan turis Malaysia ke Songkhla dan menyebabkan peningkatan pendapatan pariwisata Songkhla sekitar 289 juta baht. Choe et al. (2021) melakukan penelitian yang bertujuan untuk mengetahui dampak wabah Middle East Respiratory Syndrome (MERS) terhadap turis yang masuk ke Korea Selatan. Data yang digunakan adalah data kedatangan turis internasional dengan periode Januari 2009 hingga Desember 2015. Metode yang digunakan adalah ARIMA dengan analisis intervensi dan metode lainnya. Hasil dari penelitian ini menunjukkan bahwa ARIMA Intervensi adalah metode terbaik dalam memodelkan data dan dampak wabah MERS terjadi pada Juni hingga September 2015, dampak total dari wabah terhadap kedatangan turis ke Korea Selatan diestimasi sebesar -1.968.765 turis dengan kerugian \$3,1 miliar. Ilmiah dan Oktora (2021) melakukan penelitian yang bertujuan untuk mengetahui dampak penangkapan bibit lobster dan kebijakan larangan ekspor terhadap ekspor lobster Indonesia. Data yang digunakan merupakan data nilai ekspor dengan periode Januari 2012 hingga April 2020. Metode yang digunakan adalah ARIMA dengan analisis intervensi. Hasil dari penelitian ini menunjukkan bahwa dua intervensi, yakni Peraturan Menteri No.1 Tahun 2015 tentang pelarangan penangkapan bibit lobster dan Peraturan Menteri No.56 Tahun 2016 tentang

kebijakan larangan penangkapan dan ekspor bibit lobster memiliki dampak signifikan terhadap nilai ekspor lobster Indonesia dengan dampak yang terjadi secara permanen.

Berdasarkan latar belakang dan permasalahan yang telah dipaparkan, peneliti tertarik untuk meneliti dampak pandemi Covid-19 terhadap pendapatan sektor penyediaan akomodasi dan makan minum Provinsi Bali. Penelitian ini bertujuan untuk mengetahui perkembangan pendapatan sektor penyediaan akomodasi dan makan minum, kapan dan seberapa lama dampak tersebut terjadi, dan seberapa besar dampak yang ditimbulkan pandemi Covid-19 terhadap pendapatan sektor akomodasi dan makan minum di Provinsi Bali.

## METODE PENELITIAN

### 1. Analisis Intervensi

Suatu data runtun waktu dapat dipengaruhi oleh kejadian atau kondisi khusus tertentu seperti hari libur, aksi mogok kerja, perubahan kebijakan, kebijakan lingkungan, dan kejadian yang sejenis. Kejadian eksternal ini disebut dengan intervensi (Box, Reinsel, dan Ljung, 2016:481). Metode ini pertama kali diperkenalkan oleh Box dan Tiao (1975) (Wei, 2006). Menurut Box, Reinsel, dan Ljung (2016), kejadian intervensi diasumsikan terjadi pada suatu titik waktu  $T$  dari data runtun waktu. Kepentingan dari analisis ini adalah untuk mengetahui apakah terdapat bukti adanya efek/perubahan dari suatu data runtun waktu dengan kejadian terkait. Makridakis, Wheelwright, dan Hyndman (1997:419) menyatakan bahwa analisis intervensi sangat berguna apabila peneliti ingin mengukur dampak persis dari suatu intervensi atau jika peneliti ingin meramalkan runtun data yang memungkinkan untuk efek intervensi. Menurut Box, Reinsel, dan Ljung (2016:482), terdapat dua tipe kejadian intervensi, yaitu:

#### a) Fungsi *step*

Fungsi *step* merupakan fungsi yang digunakan untuk merepresentasikan



dampak dari sebuah intervensi yang diekspektasikan akan memiliki dampak yang permanen setelah waktu  $T$  hingga batas tertentu. Fungsi *step* dinyatakan dengan persamaan berikut:

$$S_t^{(T)} = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases} \quad (1)$$

b) Fungsi *pulse*

Fungsi *pulse* merupakan fungsi yang digunakan untuk merepresentasikan dampak dari sebuah intervensi yang bersifat temporer/semesta yang akan hilang setelah waktu  $T$ . Fungsi *pulse* dinyatakan dengan persamaan berikut:

$$P_t^{(T)} = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases} \quad (2)$$

dimana  $T$  adalah waktu terjadinya intervensi. Penggunaan fungsi *step* atau *pulse* dapat dilihat dari plot residualnya. Apabila plot residual data setelah intervensi melewati batas dalam periode waktu yang lama maka dapat digunakan fungsi *step* namun apabila hanya dalam periode waktu yang singkat maka dapat menggunakan fungsi *pulse*.

Menurut (Box, Reinsel, dan Ljung, 2016) dan (Lee, Suhartono, dan Sanugi, 2010), secara umum persamaan ARIMA intervensi dapat dituliskan sebagai berikut :

$$Y_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t + \frac{\theta(B)}{\phi(B)(1-B)^d} \varepsilon_t \quad (3)$$

$$\omega_s(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s \quad (4)$$

$$\delta_r(B) = \delta_0 - \delta_1 B - \dots - \delta_r B^r \quad (5)$$

dengan:

$Y_t$  : variabel amatan pada waktu ke- $t$

$I_t$  : variabel intervensi

$b$  : waktu terjadi intervensi

$\omega_s$  : parameter ordo  $s$

$\delta_r$  : parameter ordo  $r$

$B$  : *backshift operator*

$\frac{\theta(B)\varepsilon_t}{\phi(B)(1-B)^d}$  : model ARIMA tanpa pengaruh intervensi

Ordo  $b$ ,  $s$ , dan  $r$  ditentukan dengan melihat plot residual dari model ARIMA sebelum intervensi terlebih dahulu. Lee, Suhartono, dan Sanugi (2010) menyatakan bahwa ordo  $b$ ,  $s$ , dan  $r$  bisa diketahui dan ditentukan melalui grafik residual model ARIMA sebelum intervensi dengan batasan  $\pm 3\hat{\sigma}$ , dimana  $\hat{\sigma}$  merupakan Root Mean Square Error (RMSE) dari model ARIMA sebelum adanya efek intervensi. Selain itu,

Panjaitan, Prahutama, dan Sudarno (2018) juga menyatakan bahwa batasan sebesar  $\pm 2\hat{\sigma}$  dapat pula digunakan untuk menentukan ordo  $b$ ,  $s$ , dan  $r$ . Ordo  $b$  diketahui dari nilai residual pertama yang melewati batasan RMSE dikarenakan ordo  $b$  merupakan waktu pertama kali efek intervensi berdampak. Ordo  $s$  dapat dilihat dari berapa lama waktu setelah  $b$  untuk mencapai kestabilan setelah intervensi. Ordo  $r$  bernilai 1 apabila terdapat lag pada plot residual dan bernilai 0 apabila plot residual berpola eksponensial menurun. Lee, Suhartono, dan Sanugi (2010) menyatakan bahwa penghitungan dampak dari intervensi tidak bisa dilakukan secara langsung dari model intervensi yang terbentuk. Hal ini dikarenakan data masih dalam bentuk transformasi. Penghitungan besar dari dampak intervensi dapat dilakukan menggunakan angka hasil ramalan (*forecast*) model ARIMA yang telah diperoleh sebelumnya. Setelah angka ramalan diperoleh maka langkah selanjutnya adalah mengurangi hasil dari ramalan model intervensi dengan hasil ramalan model sebelum adanya intervensi. Inversi data perlu dilakukan sebelum menghitung dampak apabila data yang digunakan model merupakan data hasil transformasi.

## 2. Cakupan Penelitian

Penelitian ini memiliki maksud untuk melakukan analisis intervensi yaitu dampak dari pandemi Covid-19 terhadap pendapatan sektor penyediaan akomodasi dan makan minum di Provinsi Bali menggunakan data runtun waktu. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Bali. Satuan data asli adalah juta rupiah. Namun dalam penelitian ini data akan ditampilkan dalam satuan triliun rupiah. Variabel yang digunakan dalam penelitian adalah variabel independen dan variabel dependen. Variabel dependen yang digunakan dalam penelitian ini adalah PDRB sektor penyediaan akomodasi dan makan minum atas dasar harga konstan 2010 Provinsi Bali dengan periode triwulan I 2010 sampai

triwulan IV 2021. Total pengamatan berdasarkan periode data yang digunakan adalah sebesar 48 series. Variabel independen yang digunakan dalam penelitian ini adalah variabel intervensi. Variabel intervensi yang digunakan dalam penelitian ini adalah periode triwulan I 2020 di mana Covid 19 pertama kali terdeteksi di Indonesia.

#### Metode Analisis

Metode analisis yang digunakan dalam penelitian ini untuk menjawab tujuan penelitian adalah analisis deskriptif dan analisis inferensia. Analisis deskriptif yang digunakan dalam penelitian ini adalah penggambaran data secara umum. Gambaran umum data dapat digunakan sebagai salah satu cara untuk menjawab tujuan penelitian dan mendukung analisis inferensia. Gambaran umum data akan ditampilkan menggunakan diagram garis. Analisis inferensia yang digunakan adalah ARIMA dengan metode Box-Jenkins serta analisis intervensi. Analisis intervensi dapat digunakan untuk mengetahui signifikansi (berpengaruh atau tidaknya) pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum di Provinsi Bali dan bagaimana serta seberapa besar dampak dari intervensi terhadap PDRB sektor penyediaan akomodasi dan makan minum di Provinsi Bali.

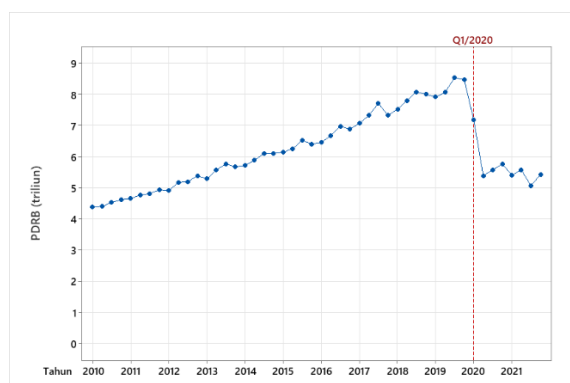
Tahapan untuk melakukan analisis ARIMA dengan intervensi dalam penelitian ini dilakukan sebagai berikut:

1. Data dikelompokkan menjadi dua kelompok berdasarkan waktu terjadinya intervensi
2. Data I merupakan data sebelum terjadinya intervensi, yaitu triwulan I 2010 hingga triwulan IV 2019 ( $t = 1, 2, \dots, 40$ ) sebanyak 40 series data ( $n_1 = 40$ )
3. Data II merupakan data yang dimulai saat terjadinya intervensi sampai akhir periode data, yaitu triwulan I 2020 hingga triwulan IV 2021 ( $t = 41, 42, \dots, 48$ ) sebanyak 8 series data ( $n_2 = 8$ )
4. Data I digunakan untuk pemodelan ARIMA/SARIMA sebelum intervensi dengan metode Box-Jenkins

5. Identifikasi, dan pemilihan model ARIMA/SARIMA terbaik menggunakan *Akaike Information Criterion* (AIC) dan *Schwarz Bayesian Criterion* (SBC), lalu dilakukan pengujian diagnostik (*white noise* dan normalitas residual) model ARIMA/SARIMA
6. Peramalan menggunakan model ARIMA/SARIMA sebelum intervensi terpilih hingga periode data II. Data ramalan adalah  $\hat{Y}_t$
7. Penghitungan nilai residual data II ( $Y_t^*$ )
8. Plot residual data II dengan batas  $\pm 3\sigma$
9. Penentuan ordo  $b$ ,  $s$ , dan  $r$  model intervensi berdasarkan plot residual yang telah dibuat
10. Estimasi parameter, pengujian signifikansi menggunakan uji parsial, dan pemilihan model intervensi terbaik menggunakan AIC dan SBC
11. Pengujian diagnostik (*white noise* dan normalitas residual) terhadap model intervensi yang terpilih
12. Apabila langkah ke-11 terpenuhi maka model intervensi terbaik telah terbentuk. Namun, apabila langkah ke-11 belum terpenuhi maka dilakukan pengulangan kembali dari langkah ke-9

## HASIL DAN PEMBAHASAN

### 1. Gambaran Umum PDRB Sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali



Gambar 2. PDRB ADHK Sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali Tahun 2010-2021

Berdasarkan Gambar 2, sebelum terjadinya pandemi Covid-19, PDRB

Sektor Penyediaan Akomodasi dan Makan Minum di Provinsi Bali memiliki tren yang cenderung naik setiap periode waktu. Untuk setiap tahunnya, PDRB mencapai nilai tertinggi mayoritas terjadi pada triwulan III. Hal ini dapat dipahami dikarenakan pada triwulan III terdapat libur akhir semester di sekolah dan universitas yang biasanya bersamaan dengan libur Idul Fitri untuk wisatawan domestik dan libur musim panas wisatawan mancanegara.

Pandemi Covid-19 yang mulai merebak di dunia pada akhir tahun 2019 hingga awal tahun 2020 dan mulai masuk ke Indonesia pada Maret 2020 membuat PDRB sektor ini mengalami penurunan yang drastis. Secara q-to-q, PDRB sektor ini mengalami penurunan sebesar -15,25% dari triwulan sebelumnya. Secara y-o-y, PDRB sektor juga mengalami penurunan sebesar -9,32%. Kondisi PDRB yang cenderung menurun terus terjadi hingga akhir tahun 2021.

## 2. Pemodelan ARIMA Sebelum Intervensi

Pemodelan ARIMA sebelum intervensi merupakan pemodelan menggunakan data sebelum terjadinya intervensi pandemi Covid-19 yaitu data dengan periode triwulan I 2010 sampai triwulan IV 2019. Model ARIMA sebelum intervensi menggunakan model ARIMA klasik atau model yang biasa dikenal dengan ARIMA Box-Jenkins. Metode ini memiliki lima tahapan yakni, identifikasi model, estimasi parameter, pemilihan model terbaik, pemeriksaan diagnostik, dan peramalan. Tahapan ini baru dapat dilakukan apabila data telah memenuhi asumsi stasioneritas dalam ragam dan rata-rata.

Pengecekan stasioneritas data dalam ragam dilakukan menggunakan transformasi Box-Cox. Data dikatakan stasioner dalam ragam apabila nilai  $\lambda = 1$ . Berdasarkan plot Box-Cox didapatkan nilai  $\lambda$  sebesar 0,46. Hal ini menunjukkan bahwa data belum stasioner dalam ragam dan memerlukan transformasi. Transformasi yang digunakan sesuai dengan metode transformasi Box-Cox adalah  $\frac{1}{\sqrt{Y_t}}$

dikarenakan nilai lambda yang mendekati 0,5. Berdasarkan plot Box-Cox, diperoleh nilai  $\lambda$  data hasil transformasi sebesar 0,93. Hal ini menunjukkan bahwa data hasil transformasi telah memenuhi asumsi stasioneritas karena nilai  $\lambda$  yang mendekati 1.

Langkah yang dilakukan selanjutnya adalah melakukan pengecekan stasioneritas data dalam rata-rata. Pengecekan stasioneritas dalam rata-rata dilakukan menggunakan uji ADF (*Augmented Dickey-Fuller*). Data dikatakan stasioner dalam rata-rata apabila nilai *p-value* kurang dari tingkat signifikansi ( $\alpha$ ) yang ditentukan. Hasil uji ADF pada level menghasilkan nilai *p-value* sebesar 0,9863. Nilai tersebut lebih besar dari tingkat signifikansi yang ditentukan yaitu 0,1 sehingga dengan tingkat signifikansi 10% keputusannya adalah gagal tolak  $H_0$ . Kesimpulan yang didapat adalah data memiliki akar unit atau data belum stasioner dalam rata-rata. Plot data menunjukkan kemungkinan adanya efek musiman dilihat dari data yang mengalami kenaikan pada setiap triwulan III. Oleh karena itu langkah selanjutnya adalah melakukan diferensiasi musiman dengan ordo 4 ( $D=4$ ). Hasil uji ADF setelah dilakukan diferensiasi musiman menghasilkan nilai *p-value* sebesar 0,1993. Nilai tersebut lebih besar dari tingkat signifikansi yang ditentukan yaitu 0,1 sehingga dengan tingkat signifikansi 10% keputusannya adalah gagal tolak  $H_0$ . Kesimpulan yang didapat adalah data memiliki akar unit atau data belum stasioner dalam rata-rata. Langkah selanjutnya adalah melakukan diferensiasi nonmusiman menggunakan ordo 1 ( $d=1$ ). Hasil uji ADF setelah dilakukan diferensiasi musiman dan diferensiasi nonmusiman menghasilkan nilai *p-value* sebesar 0,02186. Nilai tersebut lebih kecil dari tingkat signifikansi yang ditentukan yaitu 0,1 sehingga dengan tingkat signifikansi 10% keputusannya adalah tolak  $H_0$ . Kesimpulan yang didapat adalah data tidak memiliki akar unit atau data stasioner dalam rata-rata.

Tabel 1. Model Tentatif ARIMA Box-Jenkins

Model ARIMA	Estimasi Parameter	AI C	SB C
ARIMA(1,1,1)(1,1,1) <sup>4</sup>	$\phi_1 = -0,1897$ $\theta_1 = ,79$ $\Phi_1 = -0,4606$ $\Theta_1 = 0,3705$ $\theta_1 = -0,7655^*$	-	-
ARIMA(1,1,0)(1,1,0) <sup>4</sup>	$\phi_1 = -0,5412^*$ $\Phi_1 = ,03$ $\theta_1 = -0,2592$	293	286
ARIMA(1,1,1)(0,1,0) <sup>4</sup>	$\phi_1 = -0,1163$ $\theta_1 = ,82$ $\Phi_1 = -0,5571$	293	289
ARIMA(0,1,1)(0,1,1) <sup>4</sup>	$\theta_1 = -0,61133$ $\theta_1 = -0,4196^*$	296	291

Keterangan: “\*” secara menunjukkan signifikansi pada tingkat 10%

Berdasarkan plot ACF dan PACF, model ARIMA dibentuk. Tabel 1 menunjukkan bahwa model SARIMA(0,1,1)(0,1,1)<sup>4</sup> memiliki nilai AIC, SBC dan RMSE yang paling kecil di antara model tentatif lainnya. Selain itu, model ini juga memiliki signifikansi terbaik di antara model lainnya. Oleh karena itu, model ini dipilih menjadi model terbaik.

Tabel 2. Uji Diagnostik Model ARIMA Terpilih

Parameter	Uji Ljung-Box			Uji Jarque-Bera		
Esti masi	P-value	L	Chi-squa re	P- valu e	Chi-squa re	P- valu e
$\hat{\theta}_1 = -0,61133$	<0,0001	6	1,6298	0,9504	0,2423	0,8859
$\hat{\theta}_1 = -0,4196$	0,08424	18	10,234	0,924		
		24	17,052	0,8464		

Tabel 2 menunjukkan bahwa hasil pengujian Ljung-Box di seluruh lag menghasilkan p-value yang lebih besar dari 0,1 (p-value <  $\alpha$ ), sehingga dapat

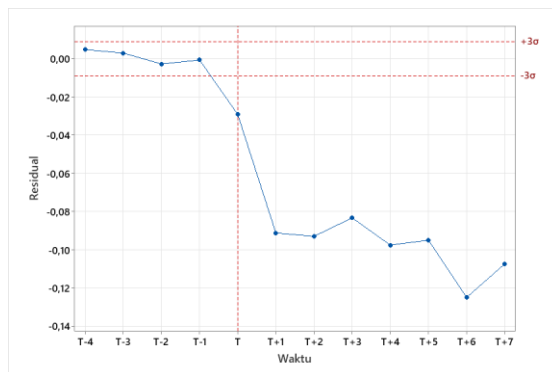
disimpulkan bahwa tidak terdapat autokorelasi pada setiap lag residual dan model ARIMA(0,1,1)(0,1,1)<sup>4</sup> telah memenuhi asumsi white noise pada residual. Uji Jarque-Bera menghasilkan p-value yang lebih besar dari 0,1 (p-value <  $\alpha$ ), sehingga dapat disimpulkan bahwa residual dari model ARIMA(0,1,1)(0,1,1)<sup>4</sup> telah memenuhi asumsi normalitas. Persamaan model ARIMA(0,1,1)(0,1,1)<sup>4</sup> dapat dituliskan sebagai berikut:

$$Y_t^* = \frac{(1+0,61133B)(1+0,4196B^4)}{(1-B)(1-B^4)} \varepsilon_t \quad (6)$$

dengan  $Y_t^* = \frac{1}{\sqrt{Y_t}}$ .

### 3. Pemodelan ARIMA Intervensi Pandemi Covid-19

Pembentukan model ARIMA Intervensi dimulai dengan membuat menentukan ordo intervensi. Ordo intervensi (b, s, dan r) dapat diketahui dan diidentifikasi melalui plot residual yang melewati batas  $\pm 2\sigma$  atau  $\pm 3\sigma$ . Nilai residual dapat diperoleh melalui hasil pengurangan nilai data aktual saat intervensi terjadi dengan data hasil peramalan menggunakan model ARIMA sebelum intervensi.



Gambar 3. Plot residual Plot Residual Intervensi Pandemi Covid-19

**Error! Reference source not found.** memperlihatkan bahwa residual pertama kali melewati batas  $\pm 3\sigma$  tepat pada waktu T atau pada triwulan I 2020. Hal ini berarti dampak dari intervensi pandemi Covid-19 terjadi pada periode waktu yang bersamaan dengan masuknya pandemi Covid-19 ke Indonesia. Hal ini menunjukkan bahwa ordo  $b$  sebagai waktu tunda dapat diperkirakan bernilai 0. Setelah itu, dapat dilihat bahwa penurunan setelah terjadinya

intervensi terjadi pada  $T+1$ . Hal ini menunjukkan bahwa ordo  $s$  dapat diperkirakan bernilai 1. Selanjutnya dalam plot juga dapat dilihat bahwa residual tidak memiliki pola tertentu. Hal ini menunjukkan bahwa ordo  $r$  dapat diperkirakan bernilai 0. Berdasarkan orde yang telah ditentukan, dibentuk model intervensi tentatif dengan cara *trial and error* (percobaan) untuk mendapatkan model ARIMA intervensi terbaik.

Tabel 3. Model ARIMA Intervensi Tentatif

Model	Estimasi Parameter	AIC	SBC
ARIMA(0,1,1)(0,1,1) <sup>4</sup>			
b=0, s=1, r=0	$\hat{\theta}_1 = 0,10637$ $\hat{\theta}_1 = 0,89472^*$ $\hat{\omega}_0 = 0,02366^*$ $\hat{\omega}_1 = 0,02008^*$	- 249,508	- 242,463
b=0, s=2, r=0	$\hat{\theta}_1 = 0,49788^*$ $\hat{\theta}_1 = 0,99986^*$ $\hat{\omega}_0 = 0,03373^*$ $\hat{\omega}_2 = 0,03982^*$	- 268,533	- 261,488
b=0, s=3, r=0	$\hat{\theta}_1 = 0,79888^*$ $\hat{\theta}_1 = 0,82486^*$ $\hat{\omega}_0 = 0,0331^*$ $\hat{\omega}_3 = 0,04963^*$	- 290,204	- 283,159
b=0, s=4, r=0	$\hat{\theta}_1 = 0,39692^*$ $\hat{\theta}_1 = 0,99983^*$ $\hat{\omega}_0 = 0,02600^*$ $\hat{\omega}_4 = 0,03664^*$	- 252,262	- 245,217

Keterangan: “\*” menunjukkan signifikansi pada tingkat 10%

Tabel 3 menunjukkan bahwa menunjukkan bahwa ARIMA intervensi dengan ordo  $b$  bernilai 0, ordo  $s$  bernilai 3, dan ordo  $r$  bernilai 0 merupakan model terbaik. Hal ini ditunjukkan oleh nilai AIC, SBC, dan RMSE yang dimiliki model tersebut merupakan nilai yang terkecil di

antara model tentatif lainnya. Selanjutnya akan dilakukan pengujian diagnostik terhadap model untuk mengetahui apakah model terpilih memenuhi asumsi independensi dan normalitas residual.

Tabel 4. Uji Diagnostik Model ARIMA Intervensi Terpilih

Estimasi Parameter	Uji Ljung-Box		Uji Jarque-Bera			
	P-value	La g	Chi-Square	P-value	Chi-Square	P-value
$\hat{\theta}_1 = 0,79888$	<0,001	6	4,73	0,3165	0,3729	0,8299
$\hat{\theta}_1 = 0,82486$	0,0001	12	5,82	0,8300		
$\hat{\omega}_0 = 0,03310$	<0,001	18	7,92	0,9513		
$\hat{\omega}_3 = 0,04963$	<0,001	24	8,55	0,9954		

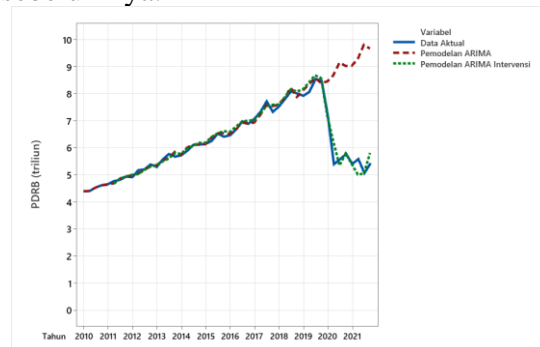
Tabel 4 menunjukkan bahwa hasil pengujian Ljung-Box di seluruh *lag* menghasilkan *p-value* yang lebih besar dari 0,1 (*p-value* <  $\alpha$ ), sehingga dapat disimpulkan bahwa tidak terdapat autokorelasi pada setiap lag residual dan model ARIMA intervensi terpilih telah memenuhi asumsi *white noise* pada residual. Uji Jarque-Bera menghasilkan *p-value* yang lebih besar dari 0,1 (*p-value* <  $\alpha$ ), sehingga dapat disimpulkan bahwa residual dari model ARIMA intervensi terpilih telah memenuhi asumsi normalitas. Persamaan model ARIMA(0,1,1)(0,1,1)<sup>4</sup> dengan ordo  $b$  sama dengan 0,  $s$  sama dengan 3, dan  $r$  sama dengan 0 dapat dituliskan sebagai berikut:

$$Y_t^* = [(0,0331 + 0,04963B^3)B^0]S_t + \frac{(1 - 0,79888B)(1 - 0,82486B^4)}{(1-B)(1-B^4)} \varepsilon_t \quad (7)$$

dengan  $S_t$  merupakan fungsi intervensi pandemi Covid-19.  $S_t$  bernilai 0 untuk periode sebelum pandemi Covid-19 (triwulan I 2010 sampai dengan triwulan IV 2019) dan bernilai 1 untuk triwulan I 2020 hingga akhir periode data.

Parameter  $\omega_b$  yang signifikan pada model terpilih menunjukkan bahwa pandemi Covid-19 berdampak signifikan terhadap PDRB sektor Penyediaan Akomodasi dan Makan Minum di Provinsi Bali. Parameter  $\omega_b$  yang memiliki ordo 0 ( $b=0$ ) menunjukkan bahwa dampak pandemi Covid-19 pertama kali dirasakan pada triwulan I 2020 atau periode Januari-Maret 2020. Penurunan PDRB dari sektor ini secara umum disebabkan wisatawan mancanegara yang datang ke Bali mengalami penurunan yang drastis yaitu penurunan sebesar -31,91 persen (m-to-m) dan -16,81 persen (y-o-y) pada Februari 2020 (BPS Provinsi Bali, 2020a) serta -56,89 persen (m-to-m) dan -65,11 persen pada bulan Maret 2020 (BPS Provinsi Bali, 2020b). Salah satu sumber terbesar penurunan turis mancanegara disebabkan oleh kebijakan larangan masuk bagi turis yang berasal dari wilayah Republik Rakyat Tiongkok (RRT) daratan yang diterapkan pemerintah Indonesia dari wilayah Tiongkok mulai 2 Februari 2020 (CNBC Indonesia, 2020). Hal ini tentu sangat berdampak bagi sektor ini karena wisatawan mancanegara yang berasal dari Tiongkok sebelum pandemi mencapai 20 persen dari total wisatawan mancanegara yang datang ke Bali (BPS Provinsi Bali, 2021b). Selain itu masuknya Covid-19 ke Indonesia pada Maret 2020 (Kompas.com, 2020a) turut menurunkan jumlah wisatawan domestik. Wisatawan domestik turun pada Maret 2020 turun sebesar -21 persen dari bulan Februari 2020 dan turun sebesar -30,5 persen dari bulan Maret 2019 (BPS Provinsi Bali, 2021c).

Parameter  $\omega_s$  memperlihatkan adanya fluktuasi (penurunan atau peningkatan) pada data periode ke-s setelah periode ke-b. Berdasarkan Berdasarkan, dapat dilihat bahwa terjadi peningkatan PDRB sektor penyediaan akomodasi dan makan minum pada T+3 yaitu pada triwulan IV 2020. Kenaikan ini terjadi karena kunjungan wisatawan domestik yang kembali mengalami kenaikan pada periode Oktober-Desember 2020 dibandingkan triwulan sebelumnya.



Gambar 4. Perbandingan Data Aktual, Data ARIMA Sebelum Intervensi, dan Data ARIMA Intervensi PDRB Sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali

Gambar 4 menunjukkan bahwa model ARIMA intervensi telah mampu menggambarkan data PDRB sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali selama tahun 2010-2021 dengan baik. Hal ini dapat dilihat dari plot hasil model ARIMA intervensi yang mendekati data aktual. Pernyataan tersebut juga diperkuat dengan nilai MAPE dari model ARIMA intervensi sebesar 2,16 persen.

#### 4. Penghitungan Dampak Intervensi Pandemi Covid-19

Tabel 5. Besar Dampak Intervensi Pandemi Covid-19 terhadap PDRB Sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali

Waktu	Periode	ARIMA	ARIMA Intervensi	Besar Dampak	Persentase (%)
T	2020q1	8,451	7,165	-1,286	-15,22
T+1	2020q2	8,680	6,222	-2,458	-28,32
T+2	2020q3	9,151	5,341	-3,810	-41,63
T+3	2020q4	9,022	5,784	-3,238	-35,89
T+4	<b>2021q1</b>	<b>9,029</b>	<b>5,391</b>	<b>-3,638</b>	<b>-40,29</b>

T+5	2021q2	9,282	5,007	-4,275	-46,06
T+6	2021q3	9,804	5,041	-4,763	-48,58
T+7	2021q4	9,661	5,806	-3,855	-39,90

Tabel 5 menunjukkan bahwa pandemi Covid-19 memiliki dampak negatif terhadap PDRB sektor Penyediaan Akomodasi dan Makan Minum Provinsi Bali. Dampak pertama kali dirasakan pada triwulan I 2020 dengan besar dampak sebesar -15,22 persen atau sebesar Rp1,286 triliun. Setelah itu dampak yang terjadi terus mengalami kenaikan. Dampak terbesar terjadi pada triwulan III 2021 yaitu sebesar -48,58 persen atau sebesar Rp4,763 triliun. Hal ini terjadi karena pada bulan Juli 2021, Indonesia mengalami puncak gelombang Covid-19 yang disebabkan oleh varian Delta (Kompas.com, 2021), wisatawan domestik yang turun setelah sempat naik di periode sebelumnya, dan tidak adanya wisatawan mancanegara yang datang ke Bali (0 kedatangan) pada bulan Juli-September 2021 (BPS Provinsi Bali, 2021a).

## KESIMPULAN

Berdasarkan hasil dan pembahasan mengenai dampak pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum di Provinsi Bali, diperoleh bahwa PDRB sektor penyediaan akomodasi dan makan minum Provinsi Bali selama tahun 2010-2019 secara umum mengalami peningkatan. Namun, pandemi Covid-19 yang terjadi sejak tahun 2020 mengakibatkan dampak yang negatif terhadap pendapatan sektor tersebut. Dampak negatif pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum Provinsi Bali langsung terjadi pada triwulan I 2020. Dampak ini terjadi secara permanen hingga akhir periode data. Dampak negatif pandemi Covid-19 terhadap PDRB sektor penyediaan akomodasi dan makan minum Provinsi Bali cukup signifikan dengan dampak terbesar hampir mencapai 50 persen. Hasil penelitian juga menunjukkan adanya tindakan yang perlu dilakukan oleh

pihak terkait. Pemerintah dapat memprioritaskan normalisasi pariwisata di Provinsi Bali karena sektor penyediaan akomodasi makan minum merupakan salah satu tulang punggung perekonomian Bali. Pemerintah juga dapat memberikan bantuan untuk pelaku usaha terdampak untuk mengurangi dampak yang dirasakan. Pemerintah juga dapat menggunakan model intervensi untuk mengestimasi dampak atas kejadian sejenis di masa depan. Selain itu, di masa depan diperlukan adanya sektor andalan lain agar perekonomian Bali tidak terlalu bertumpu pada sektor ini saja.

## DAFTAR PUSTAKA

- ABC News. (2020). Ekonomi Bali yang Tergantung Pariwisata Terpukul Setelah Larangan Turis Asing Akibat Virus Corona. Diakses pada 19 Juli 2022 dari <https://www.abc.net.au/indonesian/2020-04-06/pariwisata-bali-melesu-karena-corona/12124850>
- Bappenas. (2021). *Peta Jalan Ekonomi Kerthi Bali Menuju Bali Era Baru: Hijau, Tangguh dan Sejahtera*. Jakarta: Badan Perencanaan Pembangunan Nasional.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis Forecasting and Control* (Fifth edit). New Jersey: John Wiley & Sons.
- BPS. (2020). [Seri 2010] Laju Pertumbuhan PDB Seri 2010 (Persen), 2020. Retrieved January 5, 2022, from Produk Domestik Bruto website: <https://www.bps.go.id/indicator/11/104/2/-seri-2010-laju-pertumbuhan-pdb-seri-2010.html>
- BPS Provinsi Bali. (2020a). Perkembangan Pariwisata Provinsi Bali Februari 2020. Retrieved June 13, 2022, from <https://bali.bps.go.id/pressrelease/2020/04/01/717327/perkembangan->

- pariwisata-provinsi-bali-februari-2020.html
- BPS Provinsi Bali. (2020b). Perkembangan Pariwisata Provinsi Bali Maret 2020. Retrieved June 13, 2022, from <https://bali.bps.go.id/pressrelease/2020/05/04/717328/perkembangan-pariwisata-provinsi-bali-maret-2020.html>
- BPS Provinsi Bali. (2021a). Banyaknya Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk (Orang), 2021. Retrieved June 13, 2022, from <https://bali.bps.go.id/indicator/16/106/2/banyaknya-wisatawan-mancanegara-bulanan-ke-bali-menurut-pintu-masuk.html>
- BPS Provinsi Bali. (2021b). Banyaknya Wisatawan Mancanegara yang Datang Langsung ke Bali Menurut Kebangsaan, 2014-2020. Retrieved June 13, 2022, from <https://bali.bps.go.id/statictable/2018/02/09/27/banyaknya-wisatawan-mancanegara-ke-bali-menurut-kebangsaan-2014-2020.html>
- BPS Provinsi Bali. (2021c). Kunjungan Wisatawan Domestik ke Bali tahun 2004-2021. Retrieved June 13, 2022, from <https://bali.bps.go.id/statictable/2018/02/09/29/kunjungan-wisatawan-domestik-ke-bali-per-bulan-2004-2018.html>
- BPS Provinsi Bali. (2022). Distribusi PDRB Triwulanan Provinsi Bali Menurut Lapangan Usaha (Persen). Retrieved January 5, 2022, from <https://bali.bps.go.id/indicator/52/165/3/distribusi-pdrb-triwulanan-provinsi-bali-menurut-lapangan-usaha.html>
- Chloe, Y., Wang, J., & Song, H. J. (2021). The impact of the Middle East Respiratory Syndrome coronavirus on inbound tourism in South Korea toward sustainable tourism. *Journal of Sustainable Tourism*, 29(7), 1117–1133. <https://doi.org/10.1080/09669582.2020.1797057>
- CNBC Indonesia. (2020). Turis Dari China Dilarang Masuk dan Transit di Indonesia. Retrieved June 13, 2022, from <https://www.cnbcindonesia.com/news/20200202161713-4-134587/turis-dari-china-dilarang-masuk-dan-transit-di-indonesia>
- Fuk-Woo Chan, J., Yuan, S., Kok, K.-H., Kai-Wang To, K., Chu, H., Yang, J., ... Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395, 2020. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., ... Viruses, C. S. G. of the I. C. on T. of. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4), 536–544. <https://doi.org/10.1038/s41564-020-0695-z>
- Ilmiah, R. D., & Oktora, S. I. (2021). ARIMA Intervention Model for Measuring the Impact of the Lobster Seeds Fishing and Export Ban Policy on the Indonesian Lobster Export. *Journal of Physics: Conference Series*, 2123(1). <https://doi.org/10.1088/1742-6596/2123/1/012011>
- Kompas.com. (2020a). Fakta Lengkap Kasus Pertama Virus Corona di Indonesia. Retrieved January 5, 2022, from <https://nasional.kompas.com/read/2020/03/03/06314981/fakta-lengkap-kasus-pertama-virus-corona-di-indonesia?page=all>
- Kompas.com. (2020b). Pembatasan Sosial Berskala Besar Berhak Batasi Orang Keluar Masuk Suatu Daerah. Retrieved January 5, 2022, from <https://nasional.kompas.com/read/2020/04/01/11054771/pembatasan-sosial-berskala-besar-berhak-batasi-orang-keluar-masuk-suatu>
- Kompas.com. (2021). [KALEIDOSKOP 2021] Varian Delta yang Menggila,



- Pelajaran Penting di Bulan Juli. Retrieved June 17, 2022, from <https://nasional.kompas.com/read/2021/12/28/09235191/kaleidoskop-2021-varian-delta-yang-menggila-pelajaran-penting-di-bulan-juli?page=all>.
- Lee, M. H., Suhartono, & Sanugi, B. (2010). Multi input intervention model for analyzing the impact of the asian crisis and terrorist attacks on tourist arrivals in bali. *Applied Mathematical Sciences*, 7(133–136), 6715–6727. <https://doi.org/10.12988/ams.2013.310598>
- Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1997). *Forecasting: Methods and Applications*. New York: Wiley.
- Manoj, J., & Philip, B. (2021). Assessing the Immediate Impact of Covid-19 on the Stock Markets of Selected Countries. *Empirical Economics Letters*, 20(2), 1–8.
- Pitana, I. G., & Diarta, I. K. S. (2009). *Pengantar Ilmu Pariwisata Angkasa* (Cet. 1). Yogyakarta: Andi.
- Praporm, C., & Laipaporn, J. (2021). The Intervention Analysis of the Interrupted Incidents' Impacts on Malaysian Tourist Arrivals to Songkhla Province in Thailand. *Journal of Environmental Management and Tourism* Vol 12 No 6 (2021), 12(6), 1513–1522. Diakses pada <https://journals.aserspublishing.eu/jemt/article/view/6431>
- Rianda, F. (2021). Pemodelan Intervensi Untuk Menganalisis dan Meramalkan Jumlah Penumpang Pesawat di Bandara Soekarno-Hatta Akibat Pandemi Covid-19. *Seminar Nasional Official Statistics*, 2021(1), 283–292. <https://doi.org/10.34123/semnasoffstat.v2021i1.857>
- Stawicki, S. P., Jeanmonod, R., Miller, A. C., Paladino, L., Gaieski, D. F., Yaffee, A. Q., ... Garg, M. (2020). The 2019-2020 Novel Coronavirus (Severe Acute Respiratory Syndrome Coronavirus 2) Pandemic: A Joint American College of Academic International Medicine-World Academic Council of Emergency Medicine Multidisciplinary COVID-19 Working Group Consensus Paper. *Journal of Global Infectious Diseases*, 12(2), 47–93. [https://doi.org/10.4103/jgid.jgid\\_86\\_20](https://doi.org/10.4103/jgid.jgid_86_20)
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed). Boston: Pearson Addison Wesley.
- World Health Organization. (2020a). Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Retrieved January 5, 2022, from [https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
- World Health Organization. (2020b). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Retrieved January 5, 2022, from <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>



# PERANCANGAN PROTOTIPE WEB DISEMINASI SENSUS PERTANIAN 2023 DENGAN *RESPONSIVE WEB DESIGN*

Faturrokhman<sup>1</sup>, Farid Ridho<sup>2</sup>

<sup>1</sup> Badan Pusat Statistik Kota Banjarmasin

<sup>2</sup> Politeknik Statistika STIS

e-mail: <sup>1</sup>faturrokhman@bps.go.id

*Received: November 25, 2022; Revised: February 17, 2023; Accepted: June 22, 2023; Published: June 30, 2023*

## Abstrak

Setelah pelaksanaan cacah lengkap dalam kegiatan Sensus Pertanian 2023, BPS (Badan Pusat Statistik) berkewajiban untuk menyajikan data yang telah diperoleh melalui diseminasi data, salah satunya menggunakan situs web. Namun, web diseminasi sensus sebelumnya, ST2013, hanya mendapatkan hasil *usability testing* sebesar 66,19% dan pengguna perangkat seluler kesulitan dalam melihat informasi di dalam web. Maka dari itu, diperlukan web diseminasi ST2023 yang memiliki desain responsif di semua ukuran layar. Untuk mengatasi masalah tersebut, peneliti melakukan penelitian untuk merancang Prototipe Web Diseminasi Sensus Pertanian 2023 yang responsif menggunakan teknik *Responsive Web Design*. Prototipe kemudian dievaluasi menggunakan *usability testing* dan mendapatkan hasil sebesar 81,64%. Uji coba responsif dan kompatibilitas pada berbagai browser juga telah dilaksanakan dan mendapatkan hasil bahwa prototipe sudah responsif di semua ukuran layar dan tampil dengan baik di sebagian besar browser di perangkat yang berbeda.

**Kata kunci:** Prototipe, Diseminasi, Sensus Pertanian 2023, *Responsive Web Design*

## Abstract

*After carrying out the complete enumeration in the 2023 Agricultural Census, BPS is obliged to present the data that has been obtained through data dissemination, one of which uses a website. However, the previous census web dissemination, ST2013, only got usability testing results of 66.19% and mobile device users had difficulty viewing information on the web. Therefore, we need a ST2023 web dissemination that has a responsive design on all screen sizes. To overcome this problem, researchers conducted a study to design a responsive 2023 Agricultural Census Dissemination Web Prototype using Responsive Web Design techniques. The prototype was then evaluated using usability testing and got a result of 81.64%. Responsiveness and compatibility tests on various browsers have also been carried out and got the results that the prototype is responsive on all screen sizes and performs well in most browsers on different devices.*

**Keywords:** Prototype, Dissemination, 2023 Agricultural Census, *Responsive Web Design*

## PENDAHULUAN

Sensus Pertanian atau ST, yang dilakukan oleh BPS setiap sepuluh tahun sekali, bertujuan untuk mendapatkan data pertanian yang lengkap, aktual, dan akurat. Data pertanian yang dihasilkan akan dimanfaatkan menjadi bahan perencanaan dan evaluasi hasil pembangunan di sektor pertanian (Badan Pusat Statistik, 2013). Pada tahun 2023, BPS telah mencanangkan akan melaksanakan Sensus Pertanian yang ketujuh, yaitu Sensus Pertanian 2023 (ST2023). Sensus Pertanian merupakan rangkaian panjang kegiatan yang dilakukan oleh BPS mulai dari persiapan hingga pelaksanaannya. Puncak kegiatan ST2013, pencacahan lengkap usaha pertanian, dilaksanakan pada bulan Mei 2013. Setelah itu, data yang didapat dari pencacahan lengkap akan melalui proses pembersihan (*cleaning*), kemudian ditabulasi, dianalisis hingga akhirnya dipublikasi melalui diseminasi data.

Diseminasi merupakan kegiatan penting yang dilakukan oleh BPS dalam sensus atau survei yang telah diselenggarakan. Diseminasi bertujuan untuk memberikan informasi hasil sensus atau survei ke sekelompok orang sehingga data tersebut dapat dimanfaatkan dengan maksimal (Simbolon, 2019). Hasil diseminasi data dapat berupa berbagai bentuk, salah satunya melalui situs web yang dapat diakses oleh semua orang. Penggunaan web dalam diseminasi data terbukti cukup dalam menyajikan data yang diperlukan pengguna data. Sebagaimana terlihat dalam hasil Survei Kebutuhan Data 2021 yang menyatakan bahwa sebanyak 43,56% konsumen data di PST BPS Pusat mendapatkan data melalui web BPS, [www.bps.go.id](http://www.bps.go.id) (Badan Pusat Statistik, 2021).

Web diseminasi data hasil Sensus Pertanian 2013 dapat dilihat melalui situs <https://st2013.bps.go.id/>. Di halaman utama, ditampilkan berbagai informasi umum mengenai kondisi pertanian di Indonesia. Selain itu, terdapat pula fitur pencarian data yang disajikan dalam bentuk tabel, perbandingan antar wilayah, galeri, serta publikasi ST2013. Web diseminasi

ST2013 masih memiliki tampilan yang tidak responsif dengan ukuran layar pengguna. Konten dalam web tidak dapat menyesuaikan dengan ukuran layar pengguna sehingga pengguna perangkat seluler harus memperbesar konten untuk dapat melihat lebih jelas. Selain itu, beberapa komponen di web akan tertutup dengan komponen lain jika ukuran layar berubah.

Tampilan antarmuka web yang dapat tampil dengan baik di layar yang kecil merupakan sebuah keharusan di web masa kini. StatCounter melaporkan bahwa lalu lintas web seluler secara global telah melampaui pengguna layar desktop dan menjadi mayoritas dengan persentase sebesar 56,05% pada Februari 2022 (StatCounter, 2022). Sebagai perbandingan, pengguna yang menggunakan perangkat seluler untuk mengakses internet di Indonesia mencapai 68,91% pada Februari 2022 (StatCounter, 2022). Ketika lebih dari 60% pengunjung potensial web menggunakan perangkat seluler untuk menjelajahi internet, developer tidak bisa hanya melayani mereka dengan halaman yang dirancang untuk layar desktop (Duò, 2021). Web tersebut akan sulit untuk dibaca dan digunakan oleh pengguna perangkat seluler serta akan menyebabkan pengalaman pengguna (*user experience*) yang buruk.

Tujuan dari penelitian ini adalah untuk mengevaluasi web diseminasi ST2013 untuk menemukan hal yang perlu diperbaiki dan ditingkatkan serta mengetahui respons pengguna terhadap web tersebut. Tujuan kedua adalah merancang dan membangun prototipe web diseminasi ST2023 dengan antarmuka yang lebih responsif di berbagai macam ukuran layar. Kemudian, tujuan terakhir adalah mengevaluasi respons pengguna terhadap prototipe web diseminasi ST2023.

Penelitian yang telah ada sebelumnya dan berkaitan dengan topik penelitian ini pernah dilakukan oleh Nawir dkk. Penelitian tersebut mencoba membandingkan tingkat *usability* desain responsif web terhadap persepsi pengguna di beberapa web perguruan tinggi di

Indonesia (Nawir, Syarief, & Irfansyah, 2018). Didapatkan hasil bahwa tingkat *usability* web memengaruhi persepsi pengguna terhadap tampilan web sebesar 67,1%. Hasil dari penelitian Nawir ini menunjukkan bahwa terdapat pengaruh yang signifikan antara tingkat *usability* web dengan persepsi pengguna terhadap tampilan web. Sehingga untuk meningkatkan persepsi pengguna terhadap tampilan web, tampilan antarmuka web perlu dibuat sereresponsif mungkin guna meningkatkan tingkat kegunaan web tersebut.

Phunkaew dkk membuat sebuah penelitian yang bertujuan untuk mendesain dan mengevaluasi prototipe aplikasi *mobile Interactive Learning Story*. Prototipe didesain dengan menggunakan *Mobile Responsive Design* dan *Flat Design* (Phunkaew, Phandan, & Wongwatkit, 2019). Setelah prototipe selesai dibuat, Phunkaew melakukan evaluasi dengan mengamati langsung beberapa pengguna dalam menggunakan prototipe. Respons dan tanggapan mereka direkam untuk dianalisis lebih lanjut. Fokus utama dari evaluasi tersebut adalah mengenai *UI Design*, transisi di dalam prototipe, warna yang digunakan, komponen dan elemen dalam prototipe, navigasi, serta interaksi yang dilakukan oleh prototipe.

Banyaknya pengguna dengan ukuran perangkat yang sangat beragam membuat web tidak lagi cukup memiliki desain web statis yang hanya terlihat bagus di layar desktop. Web harus bisa menampilkan antarmuka yang responsif di semua ukuran layar pengguna. Selain itu, BPS telah mencanangkan pelaksanaan Sensus Pertanian 2023 (ST2023) dan membutuhkan web diseminasi. Dengan adanya kedua peluang tersebut, peneliti akan memberikan sebuah tampilan antarmuka baru untuk web diseminasi ST2023 dalam bentuk sebuah prototipe web. Prototipe tersebut diharapkan dapat meningkatkan tingkat kepuasan pengguna dan dapat tampil secara responsif di berbagai ukuran layar pengguna, terutama untuk pengguna layar kecil.

## METODE PENELITIAN

### 1. *Design Thinking*

Metode yang digunakan dalam analisis hingga perancangan dan evaluasi sistem pada penelitian ini adalah *Design Thinking*. *Design Thinking* berfokus pada inovasi dan penciptaan yang merupakan cara berpikir untuk dapat mengembangkan produk dan solusi baru untuk suatu masalah (Razi, Mutiaz, & Setiawan, 2018). Gambar 1 menunjukkan tahapan metode *design thinking*.

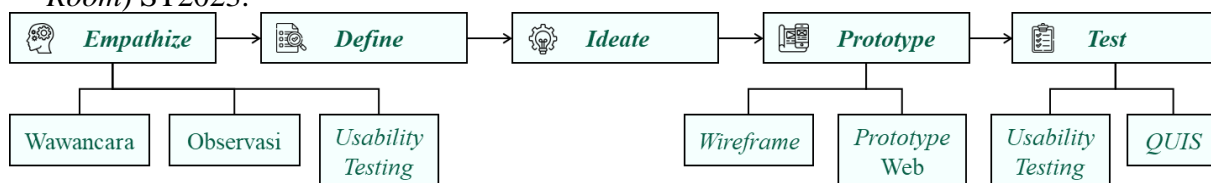
Metode *Design Thinking* memiliki lima tahapan, yaitu :

- a. Tahap pertama adalah tahapan *emphatize* (empati). Tahap ini berupaya dalam memahami kebutuhan pengguna melalui pengumpulan data yang dibagi menjadi tiga, wawancara, observasi, dan kuesioner. Wawancara dilakukan untuk mendapatkan saran terkait perancangan prototipe web kepada Tim Web Diseminasi BPS dan Tim AGRO ST2023. Observasi berupa pengamatan terhadap aplikasi yang telah dikembangkan sebelumnya yaitu web diseminasi Sensus Pertanian 2013 dengan tujuan menemukan peluang dalam pengembangan web dan melihat struktur dasar dari web. Kemudian, kuesioner *usability testing* diberikan kepada pengguna web untuk mengetahui tingkat kegunaan antarmuka web berjalan.
- b. *Define* atau mendefinisikan permasalahan yang muncul dari pengumpulan data di tahap empati sehingga dapat menjadi *input* dalam menentukan strategi di tahap selanjutnya.
- c. *Ideate* merupakan tahap menentukan solusi dari permasalahan yang sudah dianalisis sebelumnya. Tahapan ini berfungsi menghasilkan gagasan yang akan digunakan sebagai dasar dalam pengembangan web selanjutnya. Strategi yang digunakan dalam mengembangkan prototipe adalah *Responsive Web Design*, yaitu strategi untuk menangani ukuran layar yang tidak diketahui sehingga konten atau isi di halaman web

dapat menyesuaikan dengan ukuran perangkat yang mengaksesnya (Shandi, 2017).

- d. Tahapan *prototype* menghasilkan rancangan web dalam bentuk *wireframe*, dan akhirnya diimplementasikan ke dalam bentuk prototipe web sehingga dapat diakses oleh pengguna. Desain yang dikembangkan menerapkan konsep *responsive web design* dan *mobile-first design* sehingga pembuatan desain akan dimulai dari rancangan untuk perangkat seluler, kemudian perangkat tablet, dan terakhir perangkat desktop. Data yang digunakan dalam prototipe yang dirancang berasal dari Web Diseminasi ST2013 dan Tim AGRO (Agriculture Room) ST2023.

- e. Tahap *test* atau uji coba, yaitu untuk menguji coba dan mengevaluasi prototipe yang telah dibuat. Testing yang dilakukan pada tahap ini adalah pengujian pada beberapa *browser* melalui perangkat yang sering digunakan oleh pengguna. Kemudian peneliti melakukan *usability testing* untuk mengukur tingkat kegunaan prototipe web dan membandingkannya dengan hasil *usability testing* web diseminasi ST2013. Testing terakhir yaitu melakukan pengujian *QUIS* (*Questionnaire for User Interface Satisfaction*) untuk mengukur kepuasan pengguna terhadap prototipe. (Razi, Mutiaz, & Setiawan, 2018)



Gambar 5. Tahapan Metode *Design Thinking*

## 2. Usability Testing

*Usability testing* atau tingkat kegunaan merupakan pengujian yang dilakukan untuk mengevaluasi proses pengembangan suatu sistem. Salah satu metode *usability* adalah evaluasi heuristik yang digunakan untuk mengevaluasi desain antar muka pengguna (*user interface*). Heuristik merupakan aturan yang menguraikan prinsip-prinsip dalam *user interface* yang efektif. Terdapat banyak *guideline* yang digunakan dalam evaluasi heuristik, salah satunya adalah 10 prinsip heuristik menurut Jacob Nielsen. Nielsen (1994) menyebutkan bahwa evaluasi ini membutuhkan tiga hingga lima evaluator untuk mendapatkan kualitas terbaik.

Turner (2011) mendefinisikan 10 prinsip heuristik Nielsen menjadi lebih transparan dan disesuaikan dengan penerapan dalam aplikasi menjadi 10 parameter dengan 45 poin atau pertanyaan. Setiap pertanyaan memiliki tingkat kepentingan yang akan digunakan untuk menghitung skor total evaluasi heuristik. Tiap parameter memiliki bobot yang

digunakan untuk mengetahui kategori desain antarmuka pengguna yang dinilai. Pertanyaan menggunakan *severity rating* (tingkat keparahan) dari 1 (sangat buruk) sampai 5 (sangat baik). Rating tersebut kemudian dikali dengan persentase bobot dari pertanyaan menghasilkan total rating. Kemudian skor rating dihitung dengan mengalikan total rating dengan total bobot dibagi jumlah evaluator dan dikali 100%. Skor rating kemudian diinterpretasikan menjadi beberapa kategori mulai dari Sangat Buruk (Skor < 29) hingga Luar Biasa (Skor > 89) (Turner, 2011).

## HASIL DAN PEMBAHASAN

### 1. Analisis Permasalahan Sistem Berjalan

Berdasarkan hasil observasi terhadap web diseminasi Sensus Pertanian 2013, dapat ditemukan beberapa hal yang masih dapat diperbaiki, yaitu sebagai berikut :

- a. Tampilan web diseminasi tidak responsif dengan ukuran layar. Tampilan antarmuka dalam web diseminasi hanya ditujukan untuk pengguna layar besar, mulai dari laptop

hingga desktop. Ketika dibuka menggunakan perangkat dengan ukuran layar yang kecil hingga sedang, pengguna perlu memperkecil dan memperbesar konten, serta menggeser ke kanan dan kiri untuk melihat lebih jelas (Gambar 2(a)). Tim Web Diseminasi BPS menjelaskan bahwa web diseminasi ST2013 dibangun dalam era yang belum mengandalkan tampilan antarmuka yang responsif. Sehingga dalam pembangunannya pun, web diseminasi hanya menggunakan satu tampilan untuk semua perangkat, atau bisa dibayangkan tampilannya masih bersifat statis.

- b. Terdapat banyak ruang kosong di kanan kiri web, namun tulisan di dalamnya masih terlalu mepet dan kecil (Gambar 2(b)).
- c. Fitur subsektor yang merupakan salah satu data utama dalam Sensus Pertanian hanya ditampilkan kecil di atas yang menunjukkan bahwa hierarki informasi dalam web kurang jelas.
- d. Menu navigasi masih dapat diperbaiki dengan mengelompokkannya sesuai fungsinya.
- e. Tampilan dalam satu halaman web bisa digabungkan dengan halaman lain sehingga dapat memudahkan pengguna dalam mengeksplorasi dan menemukan bagian yang dicari.



Gambar 6. Tampilan Web ST2013 ketika diakses melalui (a) perangkat seluler dan (b) laptop

## 2. Hasil Usability Testing Web Diseminasi ST2013

Pada penelitian ini, peneliti melibatkan lima orang evaluator, yaitu mahasiswa dan pegawai BPS, untuk mengevaluasi web diseminasi Sensus Pertanian 2013. Evaluator yang dilibatkan pada evaluasi ini

merupakan calon pengguna web dan *subject matter* dari web itu sendiri, sehingga diharapkan hasilnya dapat valid. Hasil dari evaluasi ini adalah tingkat kegunaan antarmuka web diseminasi ST2013 memiliki persentase sebesar 66,19% yang tergolong dalam kategori Sedang ( $49 \leq \text{Skor} < 69$ ). Skor tersebut menunjukkan bahwa pengguna masih dapat menggunakan sistem, tetapi pengalaman pengguna harus ditingkatkan lagi secara signifikan. Tabel hasil evaluasi heuristik per parameter dapat dilihat pada Tabel 1 berikut.

Tabel 1. Hasil Evaluasi Heuristik Web ST2013

Parameter	Skor	Kategori
Fitur dan Fungsionalitas	80,40%	Bagus
Beranda	71,20%	Bagus
Navigasi	71,38%	Bagus
Pencarian	51,43%	Sedang
Kontrol dan Umpan Balik	66,50%	Sedang
Formulir	58,15%	Sedang
Kesalahan	67,08%	Sedang
Konten dan Teks	67,29%	Sedang
Bantuan	57,00%	Sedang
Performa	58,91%	Sedang
<b>Total</b>	<b>66,19%</b>	<b>Sedang</b>

Dari 10 parameter dalam evaluasi heuristik, hanya parameter Fitur dan Fungsionalitas, Beranda, dan Navigasi yang termasuk dalam kategori Bagus, sedangkan parameter lainnya masih dalam kategori Sedang. Selain hasil evaluasi di atas, terdapat beberapa komentar dan saran dari evaluator mengenai web diseminasi ST2013, antara lain adalah :

- a. Segi pemilihan warna, ukuran *font*, serta warna *highlight* web ST2013 yang masih dapat dieksplorasi lebih baik lagi. Web belum terlihat *clean* dan masih terkesan kaku serta kurang responsif ketika diakses melalui perangkat seluler.
- b. Saran yang diajukan antara lain adalah agar web diseminasi lebih responsif ketika diakses melalui perangkat seluler.

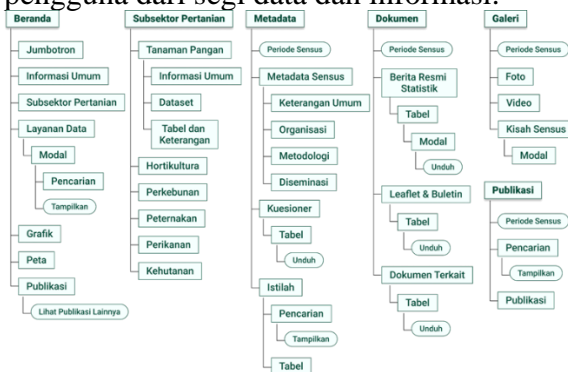


Berfokus kepada keterbacaan informasi dan kemudahan fungsi bagi pengguna seperti menampilkan kotak pencarian. Kemudian hierarki informasi, alur, serta bahasa yang digunakan dapat dipahami oleh masyarakat pada umumnya.

Berdasarkan tahap empati ini, dapat didefinisikan bahwa masalah yang terjadi adalah tingkat kegunaan web dalam kategori biasa dan pengguna perangkat seluler mengalami kesulitan dalam melihat konten dalam web.

### 3. Rancangan Peta Situs

Tahap *Ideate* pada penelitian ini adalah dengan mengumpulkan ide sebanyak mungkin mengenai tampilan dalam web yang responsif. Ide dikumpulkan dari berbagai sumber penelitian, panduan membuat aplikasi, hingga kumpulan desain UI/UX pada web. Solusi yang didapatkan adalah menggunakan *responsive web design* dan memakai teknik *mobile-first design*. Sebagai bentuk dari solusi ini, peneliti merancang peta situs yang berupa arsitektur informasi yang akan digunakan dalam web diseminasi ST2023 dan ditampilkan di Gambar 3. Rancangan ini dibuat tidak berdasarkan *user persona* (karena penelitian ini tidak berfokus pada pengguna), melainkan hasil dari observasi web diseminasi ST2013 dan wawancara dengan *subject matter*. Tim Web Diseminasi BPS selaku *subject matter* menyebutkan bahwa walaupun tampilan web ST2013 belum responsif, tetapi fitur dan fungsi yang terdapat di dalam web sudah dapat memenuhi kebutuhan pengguna dari segi data dan informasi.



Gambar 3. Rancangan Peta Situs Prototipe Web Diseminasi ST2023

Sehingga, rancangan ini dibuat berpedoman pada web tersebut, dengan tujuan tidak menghilangkan fungsi dan fitur utama dari web diseminasi. Beberapa informasi lain seperti latar belakang, tujuan, data hasil, dan daftar FAQ ST2023 merupakan konten baru yang ditambahkan sesuai dengan permintaan dari Tim AGRO ST2023, sekretariat ST2023.

### 4. Wireframe

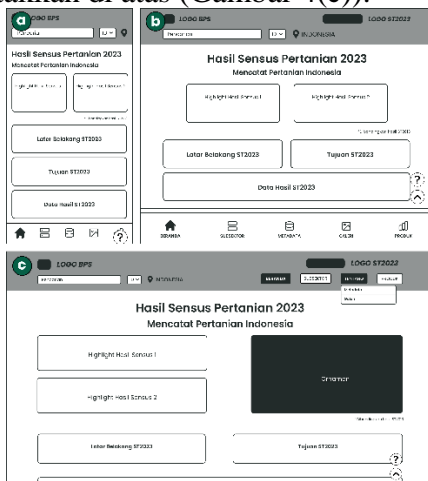
Dalam *responsive web design*, terdapat beberapa teknik yang dapat digunakan untuk merancang web. Salah satunya adalah *mobile-first design*, yaitu dengan memfokuskan desain untuk layar kecil terlebih dahulu, baru kemudian beralih secara berangsur-angsur ke layar yang lebih besar (Phunkaew, Phandan, & Wongwatkit, 2019). Dengan menggunakan *mobile-first design*, peneliti dapat menerapkan *content first*, yaitu cara memprioritaskan konten dan informasi yang penting untuk ditampilkan pada layar yang terbatas. Salah satu dampaknya adalah beberapa konten dihapus dari desain seluler, tetapi tetap ditampilkan dalam desain desktop untuk memenuhi ruang yang lebih lebar. Perancangan *wireframe* bertujuan untuk menggambarkan tata letak konten dan navigasi halaman.

*Wireframe* halaman Beranda berisikan beberapa fitur dan fungsi unggulan di dalam web diseminasi ST2023. Konten yang akan menarik pengguna pertama kali adalah *jumbotron* yang menampilkan data hasil sensus pertanian. Di bawah judul terdapat *tagline* dari Sensus Pertanian 2023, yaitu Mencatat Pertanian Indonesia. Data hasil sensus pertanian yang ditampilkan merupakan persentase perubahan (dibandingkan dengan data sensus pertanian sebelumnya) dari data Sensus Pertanian, yaitu Jumlah Rumah Tangga Usaha Pertanian dan Jumlah Perusahaan Pertanian. Kemudian di samping data hasil sensus pertanian terdapat ornamen yang hanya muncul di ukuran layar besar.



Bagian kedua dari halaman beranda berisi informasi umum Sensus Pertanian 2023 yang terdiri dari Latar Belakang, Tujuan, dan Data Hasil ST2023. Di bawahnya terdapat daftar subsektor yang terdiri dari Tanaman Pangan, Hortikultura, Perkebunan, Peternakan, Perikanan, Kehutanan, dan Jasa Pertanian. Jika kotak subsektor diklik, maka pengguna akan diarahkan ke halaman subsektor dan menampilkan tab subsektor yang sesuai. Bagian keempat dari halaman ini adalah kartu-kartu layanan data. Terdapat tiga layanan data yang menjadi fitur dalam web diseminasi, yaitu layanan *dataset*, layanan perbandingan antar wilayah, dan layanan perbandingan antar tahun.

Selanjutnya merupakan grafik perbandingan antar sensus pertanian. Data didapatkan dari Badan Pusat Statistik melalui web diseminasi ST2013. Di bawahnya terdapat peta persebaran rumah tangga usaha pertanian. Kemudian, bagian terakhir di halaman beranda adalah daftar publikasi dan tombol Lihat Publikasi Lainnya yang akan membawa pengguna ke halaman publikasi. Hasil perancangan *wireframe* untuk halaman Beranda dapat dilihat di Gambar 4. Perbedaan paling mendasar dari desain di ukuran layar kecil (Gambar 4(a)) dengan ukuran layar di atasnya adalah tidak ada logo ST2023 di atas halaman dan ukuran beberapa elemen yang menyesuaikan dengan lebar layar. Menu navigasi pada ukuran layar kecil dan sedang (Gambar 4(b)) diletakkan di bawah sedangkan pada ukuran layar besar diletakkan di atas (Gambar 4(c)).

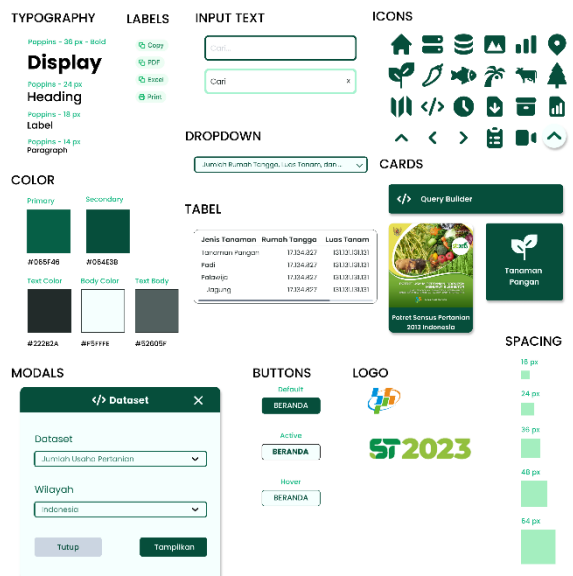


Gambar 7. *Wireframe* Halaman Beranda Prototipe Web ST2023 di (a) layar kecil, (b) layar sedang, dan (c) layar besar

## 5. Sistem Desain

Dalam mengembangkan sebuah aplikasi, diperlukan sebuah sistem desain yang akan digunakan sebagai patokan dalam membuat aplikasi tersebut. Pada penelitian ini, peneliti juga menerapkan sistem desain dengan berpanduan pada prinsip-prinsip desain dari *guideline* Microsoft. Panduan tersebut berisikan bagaimana cara menggunakan komponen web dalam desain web yang baik (Hartmut Hoehle, 2016). Komponen web yang ditampilkan sudah lengkap beserta contoh dalam bentuk desain yang dapat diedit secara langsung. Komponen yang akan digunakan dalam prototipe ini antara lain adalah *font*, *color*, *button*, *modal*, *spacing*, *input text*, tabel, *icon*, logo, *dropdown*, dan label. Sistem desain yang digunakan dalam web diseminasi ST2023 ditampilkan dalam Gambar 5.

### Design System



Gambar 8. Sistem Desain Prototipe Web Diseminasi ST2023

## 6. Hasil Prototipe Web Diseminasi ST2023

Dengan menggunakan sistem desain yang telah dibuat sebelumnya, maka rancangan *wireframe* dapat diubah ke bentuk *mockup* aplikasi kemudian

mengubahnya ke bentuk prototipe halaman web. Rancangan *wireframe* langsung bisa diubah ke dalam bentuk *mockup* tanpa pengujian terhadap pengguna terlebih dahulu. Hal ini dikarenakan perancangan desain tampilan antarmuka pada penelitian ini tidak berpusat pada pengguna. Pengguna hanya perlu menguji prototipe yang sudah selesai dan telah diuji oleh evaluator.

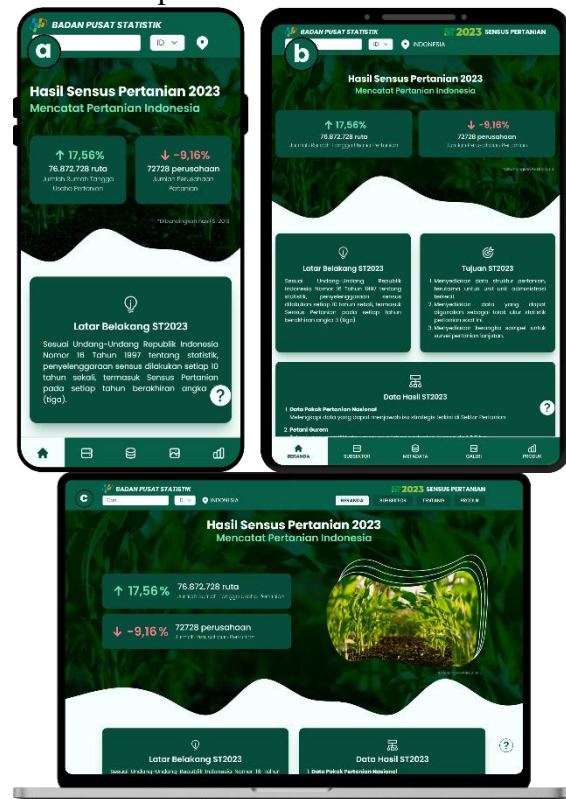
Dalam merancang halaman web, penelitian ini menggunakan *breakpoint* dari *framework* Tailwind CSS sebagai dasar dalam menerapkan *responsive design*. Adapun *viewport meta tag* digunakan untuk memberi tahu *browser* bahwa mereka harus menyetel lebar area web ke lebar sebenarnya dari perangkat sehingga *breakpoint* yang telah ditetapkan sebelumnya dapat berfungsi seperti seharusnya.

Pendekatan *mobile-first design* yang digunakan adalah dengan menerapkan satu kolom dan ukuran *font* yang lebih kecil sebagai bentuk *default* dari web. Beberapa komponen dalam web menerapkan aksi sentuh ke samping (*horizontal scroll*) pada perangkat seluler hingga tablet karena keterbatasan ukuran layar. Setiap komponen dalam prototipe ini menggunakan ukuran relatif yang dapat menyesuaikan dengan ukuran layar dengan ukuran *default* kecil.

*Layout* yang digunakan adalah *flexbox* dan *grid*, yang berfungsi untuk mengatur posisi dan ukuran elemen secara efisien dalam halaman web. *Flexbox* dapat memperluas elemen di dalamnya untuk mengisi ruang kosong yang tersedia atau mengecilkannya untuk mencegah melebihi ruang tersebut. Sedangkan *Grid* dapat mengatur lebih banyak elemen sekaligus dan lebih fleksibel penggunaannya jika dibandingkan dengan *flexbox*.

Gambar 6(a) menunjukkan tampilan prototipe web diseminasi ST2023 halaman beranda ketika dibuka menggunakan perangkat seluler atau layar kecil. Jika dibuka melalui perangkat tablet, maka

halaman beranda akan tampak seperti Gambar 6(b). Kemudian Gambar 6(c) menampilkan halaman beranda prototipe web diseminasi ST2023 di perangkat laptop atau desktop.



Gambar 9. Hasil Prototipe Halaman Beranda di perangkat (a) Smartphone, (b) Tablet, dan (c) Laptop

## 7. Hasil Usability Testing Prototipe Web Diseminasi ST2023

Evaluasi pertama terhadap prototipe adalah *usability testing* dengan heuristik yang bertujuan untuk mengetahui tingkat kegunaan prototipe dan membandingkannya dengan hasil dari evaluasi heuristik pada sistem berjalan. Hasilnya adalah tingkat kegunaan antarmuka prototipe mengalami peningkatan menjadi 81,64% yang tergolong dalam kategori Bagus ( $69 \leq \text{Skor} < 89$ ). Skor tersebut menunjukkan bahwa pengguna dapat menggunakan prototipe web dengan relatif mudah dan dapat menyelesaikan sebagian besar tugas penting.

Tabel 2. Perbandingan Hasil Evaluasi Web ST2013 dan Prototipe Web ST2023

N o	Parameter	Hasil Web ST2013	Hasil Prototipe Web ST2023
1	Fitur dan Fungsionalitas	80,40%	87,60%
2	Beranda	71,20%	89,20%
3	Navigasi	71,38%	84,92%
4	Pencarian	51,43%	73,14%
5	Kontrol dan Umpan Balik	66,50%	75,00%
6	Formulir	58,15%	79,38%
7	Kesalahan	67,08%	72,00%
8	Konten dan Teks	67,29%	88,47%
9	Bantuan	57,00%	74,67%
10	Performa	58,91%	82,91%
<b>Total</b>		<b>66,19%</b>	<b>81,64%</b>

Berdasarkan Tabel 2, parameter dengan kenaikan skor hingga 21% adalah parameter Pencarian, Formulir, dan Konten dan Teks. Ketiga parameter tersebut mengalami kenaikan yang cukup tinggi karena masih berhubungan dengan antarmuka pengguna. Fitur pencarian di Web ST2013 tidak terlihat dengan jelas, sedangkan di Prototipe Web ST2023 sudah terlihat dengan jelas dan dapat diakses dengan mudah oleh pengguna. Antarmuka formulir juga terlihat lebih rapi jika dibandingkan dengan web sebelumnya. Kemudian, konten dan teks dalam Prototipe Web ST2023 sudah mempertimbangkan masukan dari para evaluator sehingga hasilnya sudah dapat memenuhi ekspektasi dari evaluator.

Parameter dengan kenaikan skor di bawah 10% adalah parameter *Error*, Fitur dan Fungsionalitas, serta Kontrol dan Umpan Balik. Pertanyaan dari ketiga parameter tersebut melibatkan banyak fungsi dari *Back-End*, yang dalam penelitian ini tidak tercakup. Seperti pesan kesalahan ketika pengguna melakukan kesalahan, fitur jalan pintas untuk pengguna ahli, dan umpan balik untuk membatalkan aksi atau tindakan dari pengguna. Tim Web Diseminasi BPS juga telah mengatakan bahwa meskipun tampilan dari web

diseminasi ST2013 belum responsif, tetapi fitur yang tersedia sudah dapat memenuhi kebutuhan pengguna. Oleh karena itu, peneliti tetap mempertahankan fitur-fitur yang sudah ada di web sebelumnya dan menambah beberapa fitur lain yang masih belum ada.

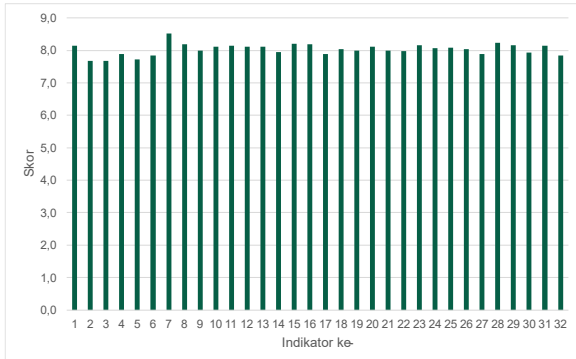
Selain hasil evaluasi pada Tabel 2, terdapat beberapa pertanyaan mengenai perangkat yang digunakan evaluator untuk mengakses prototipe. Dari total lima evaluator, sebanyak tiga evaluator menggunakan perangkat Laptop dan *Smartphone* untuk mengakses prototipe, satu evaluator hanya menggunakan perangkat *Smartphone*, dan satu evaluator lagi menggunakan perangkat Laptop saja. Ketika evaluator membuka prototipe di perangkat tersebut, 100% evaluator menjawab bahwa tampilan prototipe sudah responsif dan menampilkan antarmuka yang sesuai dengan perangkat yang digunakan.

## 8. Hasil QUIS Prototipe Web Diseminasi ST2023

Evaluasi kedua adalah menggunakan kuesioner QUIS yang bertujuan untuk mengukur tingkat kepuasan pengguna terhadap prototipe. Kuesioner ini ditujukan kepada calon pengguna web diseminasi untuk mengetahui penilaian dan pendapat calon pengguna mengenai antarmuka prototipe yang selesai dikembangkan (Chin, Diehl, & Norman, 1988). Kuesioner QUIS yang dipakai merupakan QUIS versi 7.0 yang terdiri dari enam kategori atau aspek antarmuka dengan total pertanyaan mencapai 32 pertanyaan. Setiap pertanyaan berkaitan dengan antarmuka pengguna yang memiliki skala dari 0 (buruk) hingga 9 (sangat baik sekali).

Responden QUIS berjumlah 44 orang dengan rentang usia dari 18 hingga 30 tahun. Persebaran responden QUIS bervariasi dari DKI Jakarta hingga Bali. Perangkat yang paling banyak digunakan untuk mengakses prototipe adalah Android sebanyak 56%. Kemudian, 36% responden

menggunakan Laptop, dan sisanya sebanyak 8% responden mengakses melalui iPhone. Di perangkat yang mereka gunakan, 100% responden menjawab bahwa tampilan antarmuka prototipe sudah responsif dengan ukuran layar mereka.



Gambar 10. Hasil Skor Indikator QUIS

Skor untuk setiap indikator QUIS ditampilkan dalam Gambar 7. indikator dengan skor rata-rata tertinggi adalah indikator ketujuh, yaitu Tulisan atau Karakter Pada Layar dengan skor mencapai 8,52. Skor tersebut menunjukkan bahwa Tulisan atau Karakter pada Layar cenderung Mudah Dibaca. Kemudian disusul oleh indikator ke-28 (Penggunaan Warna) dengan skor 8,23, yang artinya penggunaan warna pada prototipe cenderung Bagus.

### 9. Uji Coba Kompatibilitas *Browser* dan Validasi Halaman

Dalam mencapai desain web yang responsif untuk semua perangkat, masalah kompatibilitas dengan HTML5 dan CSS3 tidak dapat dihindari (Jiang, Zhang, Zhou, Jiang, & Zhang, 2014). Setiap *browser* dapat menafsirkan kode-kode HTML dan CSS dengan cara yang sedikit berbeda. Dengan kata lain, tampilan halaman web bisa saja berbeda bagi pengguna yang mengakses web menggunakan *browser* yang berbeda. Uji coba ini berfokus dalam melihat kompatibilitas pada sejumlah *browser* di beberapa perangkat yang sering digunakan oleh pengguna.

Tabel 3. Kompatibilitas Prototipe Web ST2023

	Chrome	Safari	Firefox	Edge	Opera
Android	Versi terbaru	Tidak Tersedia	Versi terbaru	Versi terbaru	Versi terbaru
iOS	Versi terbaru	Versi terbaru	Versi terbaru	Versi terbaru	Versi terbaru
Windows	Versi 90 ke atas	Tidak Tersedia	Versi 88 ke atas	Versi 87 ke atas	Versi 70 ke atas
macOS	Versi 90 ke atas	Versi 12.1 ke atas	Versi 88 ke atas	Versi 87 ke atas	Versi 70 ke atas

Aplikasi *browser* di perangkat Android dan iOS umumnya adalah versi terbaru, karena pengguna selalu memperbarui versi aplikasi mereka secara berkala. Untuk *browser* di Windows dan macOS terdapat beberapa versi berbeda yang masih digunakan oleh pengguna hingga sekarang. Secara keseluruhan berdasarkan Tabel 3, prototipe web diseminasi dapat ditampilkan dengan jelas dan berfungsi dengan normal di sebagian besar *browser* yang dapat diakses melalui perangkat Android, iOS, Windows maupun macOS.

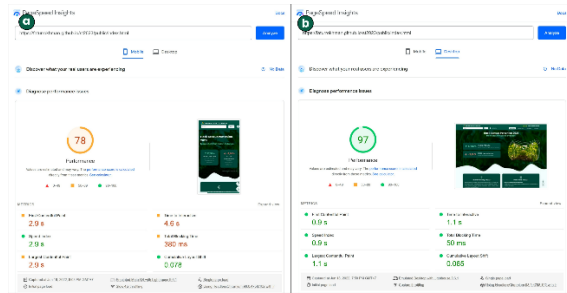
Prototipe kemudian divalidasi menggunakan validator dari W3C. Validator ini memungkinkan untuk memverifikasi setiap halaman dokumen web telah mengikuti aturan bahasa HTML. Berdasarkan hasil validator pada setiap halaman prototipe web menyatakan bahwa tidak menunjukkan adanya *error* atau *warning* di semua halaman tersebut. Artinya, prototipe web ST2023 telah mengikuti aturan bahasa HTML.

### 10. Uji Responsif Prototipe Web Diseminasi ST2023

Untuk menguji responsivitas prototipe web ST2023, prototipe diuji menggunakan *tools* Mobile-Friendly Test dan PageSpeed Index (PSI) dari Google. *Tool* tersebut bertujuan untuk menguji web sudah ramah terhadap pengguna *mobile* dan mengetahui kinerja web tersebut di perangkat seluler maupun desktop. Dari pengujian pertama menggunakan Mobile-Friendly Test, didapatkan hasil bahwa prototipe web ST2023 sudah *Mobile-Friendly* atau



responsif di perangkat seluler. Pengujian kedua menggunakan PSI, yaitu salah satu alat bantuan dari Google yang ditujukan untuk mengetahui tingkat responsivitas dari situs web. PSI melaporkan kinerja dari halaman web pada perangkat seluler dan desktop, serta memberikan saran peningkatan untuk halaman web tersebut.



Gambar 11. Tangkapan Layar Hasil PSI Halaman Beranda di perangkat (a) *Mobile* dan (b) *Desktop*

Berdasarkan Gambar 8, didapatkan hasil skor performa halaman Beranda Prototype Web ST2023 untuk perangkat *mobile* adalah 68, sedangkan untuk perangkat desktop adalah 97. Skor 78 masuk ke dalam rentang warna oranye yang artinya masih bisa ditingkatkan lagi. Skor 97 sudah termasuk ke dalam rentang warna hijau yang berarti sudah baik. Performa di perangkat *mobile* lebih rendah dibandingkan dengan performa di perangkat desktop. Hal ini dikarenakan di perangkat *mobile*, komponen web tetap dimuat semuanya walaupun beberapa komponen tidak ditampilkan. Hal ini memang menjadi salah satu kekurangan dalam menerapkan *Responsive Web Design*.

## KESIMPULAN DAN SARAN

Penelitian ini sudah mencapai tahap akhir dan telah dapat ditarik kesimpulan dari hasil penelitian dengan menjawab tujuan penelitian di awal. Pertama, evaluasi heuristik telah dilakukan untuk mengevaluasi tingkat kegunaan dari web diseminasi ST2013 dan didapatkan hasil sebesar 66,19% yang tergolong dalam kategori Sedang. Kedua, telah dirancang dan dibangun sebuah Prototype Web Diseminasi ST2023 dengan tampilan

antarmuka yang lebih modern dan responsif, tanpa membangun *back-end* dan *database*. Pembangunan prototipe menggunakan teknik *Responsive Web Design* dan *Mobile First Design* untuk memastikan prototipe dapat diakses dengan nyaman oleh pengguna perangkat seluler. Untuk menguji responsivitas dari prototipe web, dilakukan dua uji menggunakan *tools* dari Google dan mendapatkan hasil bahwa prototipe web sudah responsif di berbagai macam ukuran layar dan perangkat, terutama di perangkat seluler atau layar kecil.

Ketiga, Prototype Web ST2023 telah melalui serangkaian evaluasi dan uji coba. Evaluasi heuristik untuk mengetahui tingkat kegunaan prototipe mendapatkan hasil sebesar 81,64% dan tergolong dalam kategori Bagus, meningkat jika dibandingkan dengan hasil evaluasi web ST2013. QUIS untuk mengetahui tingkat kepuasan pengguna terhadap prototipe mendapatkan hasil sebesar 8,03 atau pengguna merasa sudah cukup puas dengan tampilan antarmuka prototipe. Uji coba kompatibilitas browser menunjukkan bahwa Prototype Web ST2023 dapat tampil dengan sesuai pada sebagian besar browser di perangkat yang berbeda.

Saran yang dapat diberikan untuk penelitian selanjutnya adalah performa prototipe web ST2023 menurut PSI di perangkat *mobile* masih dalam kategori warna oranye, yang artinya masih bisa ditingkatkan lagi. Kemudian, saran selanjutnya adalah membangun *back-end* dan memasukkan *database* ke bagian *front-end* yang sudah dibuat.

## DAFTAR PUSTAKA

- Badan Pusat Statistik. (2013). *Sensus Pertanian 2013*. Dipetik November 4, 2021, dari <https://st2013.bps.go.id/dev2/index.php/metadata/index>
- Badan Pusat Statistik. (2021). *Analisis Hasil Survei Kebutuhan Data 2021*. Jakarta: BPS RI.
- Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user

- satisfaction of the human-computer interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88)*, 213-218. doi:<https://doi.org/10.1145/57167.57203>
- Duò, M. (2021, Agustus 30). *The Beginner's Guide to Responsive Web Design (Code Samples & Layout Examples)*. Dipetik Januari 25, 2022, dari Kinsta: <https://kinsta.com/blog/responsive-web-design/>
- Hartmut Hoehle, R. A. (2016). Leveraging Microsoft's mobile usability guidelines: Conceptualizing and developing scales for mobile application usability. *International Journal of Human-Computer Studies*, 35-53.
- Jiang, W., Zhang, M., Zhou, B., Jiang, Y., & Zhang, Y. (2014). Responsive Web Design Mode and Application. *IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*, 1303-1306. doi:10.1109/WARTIA.2014.6976522
- Nawir, F., Syarief, A., & Irfansyah. (2018). Pengaruh Tingkat Usability Desain Responsif Web Mobile Perguruan Tinggi Terhadap Persepsi Pengguna. *Jurnal Visualita*, 7(1), 1-10.
- Nielsen, J. (1994, April 24). *10 Usability Heuristics for User Interface Design*. Diambil kembali dari Nielsen Norman Group: <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Othman, A., & Gloaguen, R. (2013). River Courses Affected by Landslides and Implications for Hazard Assessment: A High Resolution Remote Sensing Case Study in NE Iraq–W Iran. *Remote Sensing*, 5(3), 1024–1044.
- Phunkaew, T., Phandan, C., & Wongwatkit, C. (2019). Design and Evaluation of Interactive Learning Story and User Interface Prototyping for Mobile Responsive Learning Application. *4th International Conference on Information Technology (InCIT)*, 132-137. doi:10.1109/INCIT.2019.8912103
- Razi, A. A., Mutiaz, I. R., & Setiawan, P. (2018). Penerapan Metode Design Thinking Pada Model Perancangan UI/UX Aplikasi Penanganan Laporan Kehilangan dan Temuan Barang Tercecer. *Demandia*, 3(2), 75-93.
- Shandi, Y. J. (2017). Menentukan Desain Antarmuka yang Tepat Untuk Sebuah Website yang Modern. *Media Informatika*, 16(1), 41-46.
- Sianturi, R. (t.thn.). *Riyanthi Sianturi*. Dipetik 2022, dari <https://riyanthisianturi.com/heuristic-evaluation/>
- Simbolon, V. T. (2019). *Prototipe Aplikasi Diseminasi Sensus Penduduk 2020 Berbasis Web*. Jakarta: Skripsi Politeknik Statistika STIS.
- StatCounter. (2022). *Desktop vs Mobile vs Tablet Market Share in Indonesia - February 2022*. Diambil kembali dari StatCounter - GlobalStats: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/indonesia/#monthly-201112-202112>
- StatCounter. (2022). *Desktop vs Mobile vs Tablet Market Share Worldwide - February 2022*. Diambil kembali dari StatCounter - GlobalStats: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide/#monthly-201112-202002>
- Turner, N. (2011, Februari 11). *A guide to carrying out usability reviews*. Diambil kembali dari UXM: <https://www.uxforthemasses.com/usability-reviews/>

# PERAMALAN KASUS HARIAN MONKEYPOX DUNIA DENGAN PENDEKATAN *SUPPORT VECTOR REGRESSION*

Marcel Laverda Subiyanto<sup>1</sup>, Yulia Amanda<sup>2</sup>, Muhammad Nadhil Fachrian<sup>3</sup>, Afriani<sup>4</sup>, Achmad Yazid Busthomi Rohim<sup>5</sup>, Nur Chamidah<sup>6\*</sup>

<sup>1,2,3,4,5,6</sup>Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga  
e-mail: <sup>1</sup>marcel.laverda.subiyanto-2019@fst.unair.ac.id, <sup>2</sup>yulia.amanda-2019@fst.unair.ac.id,  
<sup>3</sup>muhammad.nadhil.fachrian-2019@fst.unair.ac.id, <sup>4</sup>afriani-2019@fst.unair.ac.id,  
<sup>5</sup>achmad.yazid.bustomi-2019@fst.unair.ac.id, <sup>6</sup>nur-c@fst.unair.ac.id

\*Penulis Korespondensi

Received: December 21, 2022; Revised: June 6, 2023; Accepted: June 20, 2023; Published: June 30, 2023

## ABSTRAK

*Monkeypox* saat ini menjadi perhatian masyarakat global. Maka, penting untuk mengetahui perkembangan jumlah kasus *monkeypox* kedepannya. Pada penelitian ini dilakukan peramalan kasus harian *monkeypox* menggunakan metode *Support Vector Regression* (SVR) dengan Fungsi *Kernel Radial Basis Function* (RBF). Data yang digunakan adalah data sekunder berupa deret waktu harian mulai 29 Mei sampai 20 Oktober 2022. Untuk memperoleh parameter optimal pada model SVR, peneliti menggunakan algoritma *grid search* untuk memprediksi data *testing* secara akurat. Berdasarkan hasil pencarian parameter optimal menggunakan algoritma *grid search* diperoleh parameter optimalnya adalah  $C = 2^{0,6}$ ,  $\gamma = 2^{0,2}$ , dan  $\varepsilon = 0,09$ . Nilai RMSE untuk hasil peramalan menggunakan model SVR dengan parameter optimal pada data *training* dan *testing* sebesar 352,3 dan 809,7.

**Kata kunci :** Monkeypox, Peramalan, SVR, RBF, *Grid Search*

## ABSTRACT

*Monkeypox* is a disease that becomes public health emergency of international concern. It is important to know the development of the number of *monkeypox* cases in the future. In this study, forecasting *monkeypox* daily cases was carried out using the *Support Vector Regression* (SVR) method with the *Kernel Radial Basis Function* (RBF). The data used is secondary data from May 29<sup>th</sup> to October 20<sup>th</sup> 2022. To obtain optimal parameters in the SVR model, researcher uses a *grid search* algorithm to accurately predict the testing data. Based on the search results for optimal parameters using the *grid search* algorithm, the optimal parameters are  $C = 2^{0,6}$ ,  $\gamma = 2^{0,2}$ , and  $\varepsilon = 0.09$ . RMSE values for forecasting using the SVR model with optimal parameters on training and testing data are 352.3 and 809.7.

**Keywords:** *Monkeypox*, Forecasting, SVR, RBF, *Grid Search*

## PENDAHULUAN

Salah satu penyakit endemi yang kini menjadi perhatian masyarakat global adalah penyakit *monkeypox* atau cacar monyet. Awal mula Virus *monkeypox* menginfeksi manusia adalah pada tahun 1970 di Republik Demokratik Kongo dimana virus tersebut dianggap sebagai penyakit endemik (Cabanillas *et al.*, 2022). Belum ditemukan alasan spesifik terjadinya penyebaran penularan penyakit *monkeypox* di daerah endemik dan non-endemik, akan tetapi hingga saat ini sebagian besar penularan terjadi melalui manusia ke manusia setelah melakukan perjalanan ke negara-negara Afrika. Penularan antar manusia terjadi melalui tetesan pernapasan, kontak dengan cairan tubuh, lingkungan atau barang-barang pasien yang terkontaminasi, lesi kulit orang yang terinfeksi dan dengan penularan seksual dari individu yang terinfeksi dengan lesi selangkangan dan genital (Jamil *et al.*, 2022).

Sejak awal Mei 2022, lebih dari 3000 infeksi virus cacar monyet telah dilaporkan di lebih dari 50 negara dari 5 benua mendorong *World Health Organization* (WHO) menyatakan cacar monyet sebagai keadaan darurat kesehatan global (Thornhill *et al.*, 2022). Pada tanggal 20 Juli 2022, dilaporkan bahwa kasus cacar monyet mencapai 14 ribu kasus di dunia oleh *World Health Organization* (WHO). Pada awal agustus 2022, kasus *monkeypox* kembali meningkat dimana lebih dari 26.500 kasus yang tersebar di 80 negara. Hingga Kamis, 18 Agustus 2022, dilaporkan bahwa kasus cacar monyet telah menyebar lebih dari 90 negara pada negara non-endemik dimana terdapat 38.800 kasus jumlah pasien yang telah terkonfirmasi *monkeypox* di dunia. (Farid, 2022).

Kasus pertama cacar monyet pertama kali muncul di Indonesia dilaporkan pada tanggal 20 Agustus 2022 (Wijaya, 2022). Pasien merupakan seorang laki-laki berusia 27 tahun asal Jakarta. Pasien tersebut pernah tercatat melakukan perjalanan ke beberapa negara pada 8 Agustus 2022. Pasien menunjukkan gejala demam, limpa membesar, dan ruam di wajah, telapak

tangan, kaki, dan bagian sekitar alat kelamin. Merespons munculnya kasus pertama cacar monyet di Indonesia, Ikatan Dokter Indonesia (IDI) telah membentuk gugus tugas yang didedikasikan untuk memerangi virus. Gugus tugas baru yang dibentuk mencakup hingga 25 dokter dari setidaknya enam organisasi medis yang berbeda, termasuk Perhimpunan Dokter Spesialis Penyakit Dalam Indonesia (PAPDI) dan Perhimpunan Dokter Spesialis Kulit dan Kelamin Indonesia (Perdoski) (Suhenda, 2022).

Berdasarkan penelitian yang dilakukan oleh John P. Thornhill *et al.*, (2022), melaporkan bahwa terdapat 528 infeksi yang didiagnosis antara 27 April dan 24 Juni 2022, di 43 lokasi di 16 negara. Secara keseluruhan, 98% orang yang terinfeksi adalah pria *gay* atau biseksual, 75% berkulit putih, dan 41% memiliki infeksi virus *human immunodeficiency*; dengan usia rata-rata adalah 38 tahun. Asumsi peneliti menyebutkan bahwa penularan kasus *monkeypox* terjadi melalui aktivitas seksual pada 95% orang yang terinfeksi. Infeksi menular seksual bersamaan dilaporkan pada 109 dari 377 orang (29%) yang diuji. Dalam rangkaian kasus ini, *monkeypox* dimanifestasikan dengan berbagai temuan klinis dermatologis dan sistemik. Identifikasi simultan kasus di luar daerah di mana *monkeypox* secara tradisional menjadi endemik menyoroti perlunya identifikasi dan diagnosis kasus yang cepat untuk mencegah penyebaran lebih lanjut.

Penelitian lain terkait metode yang akan peneliti gunakan dilakukan oleh Lestari *et al.*, (2021) mengenai peramalan pertambahan pasien Covid-19 menggunakan *Support Vector Regression* (SVR). Hasil pemodelannya menunjukkan hasil yang baik dengan nilai MSE sebesar 89,6 dan nilai MAPE sebesar 13,9% yang menjadikan model tersebut memiliki kemampuan peramalan yang baik. Mengacu pada penelitian yang telah dilakukan sebelumnya, dalam penelitian ini akan dilakukan peramalan kasus harian *monkeypox* menggunakan *machine learning* dengan metode SVR. SVR merupakan hasil pengembangan dari



*Support Vector Machine* (SVM) dimana hasil proses prediksi diperoleh dari fungsi regresi  $f(x)$  yang digunakan sebagai garis pemisah atau *hyperplane*. Kelebihan dari metode SVR adalah tingkat *error* yang rendah dalam menginterpretasi hasil dan dapat menghasilkan prediksi yang baik. Akan tetapi, metode SVR memiliki kelemahan pada saat *training* dimana apabila dilakukan dengan kumpulan data yang besar akan mengakibatkan proses *training* akan menjadi lambat (Hendayanti *et al.*, 2019).

Menanggapi kasus cacar monyet di seluruh dunia yang terus bertambah dengan cepat, penting untuk mengetahui dan menganalisis *trend* kasus cacar monyet di kemudian hari. Hasil analisis tersebut nantinya dapat digunakan sebagai acuan dalam merumuskan kebijakan yang tepat untuk pengendalian dan pencegahan kasus cacar monyet oleh pemerintah Indonesia dan dunia terutama WHO. Selain itu penelitian ini juga merupakan bentuk implementasi dari *Sustainable Development Goals* (SDGs) poin ketiga yaitu kehidupan sehat dan sejahtera (*Good health and well-being*).

## METODOLOGI

### 1. Tinjauan Pustaka

#### *Monkeypox*

*Monkeypox* adalah *zoonosis* baru yang disebabkan oleh virus *monkeypox* (anggota genus *Orthopoxvirus* dari keluarga *Poxviridae*). Penyakit ini memiliki gejala yang sangat mirip dengan cacar yang telah diberantas pada tahun 1980. Meski gejalanya lebih ringan dari cacar, cacar monyet terjadi secara sporadis di beberapa wilayah Afrika, terutama Afrika bagian tengah dan barat. Penularan pada manusia terjadi melalui kontak langsung dengan orang atau hewan yang terinfeksi, atau melalui benda yang terkontaminasi virus. Penyakit ini dapat ringan dan gejalanya berlangsung 2-4 minggu, tetapi dapat berkembang menjadi penyakit yang berat bahkan kematian (kematian 3-6%) (Kementerian Kesehatan RI, 2022).

#### Uji Terasvirta

Uji Terasvirta adalah salah satu pengujian nonlinear yang dikembangkan berdasarkan model *neural network*. Uji ini dikembangkan melalui pendekatan ekspansi deret *Taylor* dan masuk ke dalam kelompok uji *Lagrangian Multiplier* (LM). Prosedur dalam pengujian Terasvirta dijabarkan sebagai berikut (Terasvirta *et al.*, 1993):

1. Meregresikan  $X_t$  pada  $1, X_{t-1}, \dots, X_{t-p}$  dan sehingga diperoleh model linier

$$X_t = f_t + \hat{e}_t$$

Dimana  $f_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}$

Kemudian menghitung nilai jumlah kuadrat nilai residual  $SSR_0 = \sum \hat{e}_t^2$

2. Meregresikan  $\hat{e}_t$  pada  $1, X_{t-1}, X_{t-2}, \dots, X_{t-p}$  dan  $m$  prediktor tambahan yang berasal dari nilai koefisien suku kuadrat dan kubik yang dihasilkan dari pendekatan ekspansi deret *Taylor*. Selanjutnya, hitung jumlah kuadrat dari residualnya  $SSR_1 = \sum \hat{v}_t^2$

3. Menghitung nilai statistik uji

$$F_{hit} = \frac{\frac{SSR_0 - SSR_1}{m}}{\frac{SSR_1}{(n - p - 1 - m)}}$$

dengan  $n$  adalah jumlah pengamatan.

Hipotesis yang digunakan adalah sebagai berikut

$H_0 : f(x)$  adalah fungsi linier dalam  $x$  (model linier)

$H_1 : f(x)$  adalah fungsi nonlinier dalam  $x$  (model nonlinier)

Keputusan tolak  $H_0$  jika  $F_{hit} > F_{(n-p-1-m)}$  atau  $p$ -value kurang dari taraf nyata 0,05.

#### Fungsi Kernel

Kernel merupakan suatu fungsi yang memetakan fitur data dari dimensi awal (rendah) ke dalam fitur baru dengan dimensi yang lebih tinggi (Agustina *et al.*, 2017). Berikut ini merupakan beberapa fungsi kernel (Hsu, Chang, & Lin, 2016):

- a. Linear

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

- b. Polinomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + r)^d$$

c. *Radial Basic Function* (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$$

d. Sigmoid

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$$

Pemilihan fungsi kernel yang sesuai penting untuk diperhatikan karena fungsi kernel menentukan fitur baru di lokasi *hyperplane* yang nantinya akan dicari. Fungsi kernel digunakan untuk mengimplementasikan model pada ruang berdimensi lebih tinggi (*feature space*) tanpa mendefinisikan fungsi pemetaan dari *input space* ke *feature space*, yang pada kasus nonlinear diasumsikan linear pada *feature space*.

### Support Vector Regression (SVR)

*Support Vector Regression* (SVR) adalah adaptasi *machine learning* untuk klasifikasi berdasarkan model regresi *Support Vector Machine* (SVM) (Liu, Lee, & Chiu, 2009). SVM yang pada dasarnya suatu teknik klasifikasi dimana variabel responnya merupakan variabel ordinal, dikembangkan sebagai teknik regresi dimana variabel responnya adalah variabel numerik berupa bilangan *real* dan kontinu. SVR merupakan salah satu metode yang dapat mengatasi *overfitting* sehingga dapat bekerja dengan baik dan memiliki performa yang bagus (Suprayogi & Pardede, 2022).

Adapun detail penjelasan dari SVR menurut Cao & Wu (2016) adalah sebagai berikut: Diketahui data awal adalah  $\mathbf{x}$  dipetakan secara nonlinier ke ruang fitur dimensi yang lebih tinggi. Data *training* dinyatakan sebagai  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ , dimana  $\mathbf{x}_i = \{x_1, x_2, \dots, x_p\}^T \in R^n$ ,  $\mathbf{x}_i$  adalah vektor *input*,  $y_i$  adalah nilai luaran,  $n$  adalah jumlah *training set*. Rumus SVR dapat direpresentasikan seperti berikut:

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b$$

dimana  $\mathbf{w}$  adalah vektor bobot,  $\varphi(x)$  adalah fungsi yang memetakan  $\mathbf{x}$  dalam suatu dimensi,  $b$  adalah konstanta.

Untuk mengestimasi koefisien  $\mathbf{w}$  dan  $b$  pada persamaan di atas, dilakukan dengan meminimumkan fungsi resiko (*risk function*) dari persamaan berikut (Haykin, 2008).

$$R(f(\mathbf{x})) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_\varepsilon(y_i - f(\mathbf{x}_i))$$

atau dapat ditulis

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_\varepsilon(y_i - f(\mathbf{x}_i))$$

di mana  $\|\mathbf{w}\|$  adalah regularisasi, yaitu fungsi yang diminimalkan agar fungsinya sehalus mungkin (*flat*). Konstanta  $C > 0$  adalah kompromi (*trade off*) antara ketipisan fungsi dan batas atas deviasi, yang lebih dari yang dapat ditoleransi (Smola & Schölkopf, 2004). Fungsi kerugian  $\varepsilon$ -*insensitive loss function*  $E_\varepsilon$  didefinisikan sebagai berikut

$$E_\varepsilon(y_i - f(\mathbf{x}_i)) = \begin{cases} 0, & |y_i - f(\mathbf{x}_i)| \leq \varepsilon \\ |y_i - f(\mathbf{x}_i)| - \varepsilon, & \text{lainnya} \end{cases}$$

Fungsi regresi  $f$  adalah fungsi yang dapat mengaproksimasi semua titik  $(\mathbf{x}_i, y_i)$  dengan ketelitian  $\varepsilon$ . Semua titik yang terletak pada interval  $f \pm \varepsilon$  disebut dapat titik yang dapat diterima (*feasible*). Titik di luar rentang  $f \pm \varepsilon$  disebut titik yang tidak layak diterima (*infeasible*). Titik *infeasible* ini diperluas dengan variabel *slack* positif  $\xi_i$  dan  $\xi_i^*$  untuk mengatasi batasan tidak layak (*infeasible constrain*) dari masalah optimisasi. Sehingga persamaan fungsi resiko (*risk function*) berubah menjadi

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

dengan kendala

$$\begin{cases} y_i \leq \mathbf{w}^T \varphi(\mathbf{x}_i) + \varepsilon + \xi_i \\ y_i \geq \mathbf{w}^T \varphi(\mathbf{x}_i) - \varepsilon - \xi_i^* \\ \xi_i, \xi_i^* > 0 \end{cases}$$

Menurut Smola dan Schölkopf (2004), solusi dari persamaan di atas adalah dengan menggunakan koefisien *Lagrange* dengan koefisien *Lagrange* untuk setiap kendala yang ditunjukkan pada persamaan berikut.

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}_i) - y_i) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}_i) + y_i)$$

dengan  $\alpha_i, \alpha_i^*, \xi_i, \xi_i^* > 0$  adalah koefisien *lagrange*. *Lagrangian* menghasilkan solusi optimal parameter  $\mathbf{w}$  dalam bentuk

koefisien *lagrange*  $\alpha_i$  dan  $\alpha_i^*$  sebagai berikut

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varphi(\mathbf{x}_i)$$

Sedangkan estimasi akhir dari  $b$  adalah  $b = y_i - \mathbf{w}\varphi(\mathbf{x}_i) - \varepsilon$  untuk  $0 \leq \alpha_i \leq C$  dan  $b = y_i - \mathbf{w}\varphi(\mathbf{x}_i) + \varepsilon$  untuk  $0 \leq \alpha_i^* \leq C$ . Sehingga fungsi regresi SVR dapat dituliskan sebagai

$$f(\mathbf{x}) = \sum_{i=1}^n (a_i - a_i^*) \varphi(\mathbf{x}_i) \varphi(\mathbf{x}) + b$$

Menurut Vapnik (1995), perkalian skalar (*dot product*)  $\varphi(\mathbf{x}_i) \varphi(\mathbf{x})$  dapat diganti dengan fungsi kernel. Ini karena pemetaan data  $\mathbf{x}$  dari ruang input ke ruang fitur berdimensi lebih tinggi dengan fungsi  $\varphi$  biasanya tidak diketahui dan sulit dipahami. Dengan demikian, fungsi  $f(x)$  ditulis sebagai berikut.

$$f(\mathbf{x}) = \sum_{i=1}^n (a_i - a_i^*) K(\mathbf{x}_i, \mathbf{x}_j) + b$$

di mana  $a_i$   $a_i^*$  dan adalah pengali *Lagrange*,  $K(\mathbf{x}_i, \mathbf{x}_j)$  adalah fungsi kernel. Fungsi kernel memungkinkan model untuk diimplementasikan dalam ruang dimensi yang lebih tinggi (ruang fitur) tanpa harus mendefinisikan fungsi pemetaan dari ruang *input* ke ruang fitur. Meskipun ada beberapa pilihan fungsi kernel, fungsi kernel yang paling umum digunakan adalah fungsi *Radial Basis Function* (RBF) karena parameter yang harus ditentukan lebih sedikit dan memiliki kemampuan pemetaan nonlinear dari data pelatihan ke ruang dimensi tak terbatas (Cao & Wu, 2016). Fungsi kernel RBF didefinisikan sebagai berikut.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

dimana  $\gamma$  menunjukkan *bandwidth* RBF. Dengan demikian, RBF diterapkan dalam penelitian ini sebagai fungsi kernel.

#### **Algoritma Grid Search**

*Grid search* merupakan pencarian parameter optimal dari SVR yang berdasarkan subset dari ruang *hyperparameter* yang berada dalam rentang nilai minimum (*lower bound*) dan nilai maksimum (*upper bound*) yang telah ditentukan (Syarif et al., 2016). *Grid search* membagi rentang parameter yang akan

dioptimalkan ke dalam *grid* dan iterasi melalui semua titik untuk menemukan parameter optimal. Algoritma *grid search* mencari parameter optimal SVR *cross validation*. Tujuannya adalah untuk mengidentifikasi kombinasi *hyperparameter* yang optimal untuk memprediksi data *testing* secara akurat. Oleh karena itu disarankan untuk mencoba beberapa variasi pasangan parameter pada *hyperplane* SVR (Yasin et al., 2014).

Butuh waktu lama untuk menyelesaikan pencarian *hyperparameter* yang optimal melalui metode *grid search*. Oleh karena itu, disarankan untuk melakukan pencarian *hyperparameter* menggunakan metode *grid search* dalam dua langkah, yaitu *loose grid* dan *finer grid* (Purnama & Hendarsin, 2020). *Loose grid* adalah langkah parameter optimal yang dipilih dari pangkat bilangan bulat. *Finer grid* adalah langkah selanjutnya dari *loose grid*, di mana pencarian parameter optimal didasarkan pada nilai persekitaran parameter yang dihasilkan pada langkah *loose grid*. Berdasarkan hal tersebut, penelitian ini menggunakan pencarian *hyperparameter* optimal dengan metode *grid search* dalam dua langkah, yaitu. *loose grid* dan *finer grid*. Parameter optimal adalah parameter yang memiliki keakuratan terbaik dengan nilai *error* terkecil.

#### **Root Mean Square Error (RMSE)**

Salah satu ukuran untuk mengukur tingkat akurasi peramalan adalah dengan menggunakan *Root Mean Square Error* (RMSE). RMSE adalah standar deviasi dari residual (kesalahan prediksi). RMSE adalah ukuran seberapa tersebar dari residual. Dengan kata lain, RMSE menyatakan seberapa terkonsentrasi data di sekitar garis prediksi. RMSE umumnya digunakan dalam klimatologi, peramalan, dan analisis regresi untuk memverifikasi hasil eksperimen. RMSE dapat dihitung dengan rumus berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

Dimana  $y_t$  merupakan data aktual pada periode ke- $t$ ,  $\hat{y}_t$  merupakan data prediksi

pada periode ke- $t$ , dan  $n$  merupakan banyak periode waktu.

## 2. Data dan Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder diambil dari *website Our World In Data (OWID)* yang merupakan penerbitan daring yang menyajikan kumpulan data dengan lisensi terbuka. Data yang diambil merupakan data kuantitatif jumlah kasus terjangkit virus *monkeypox* dari seluruh dunia periode 29 Mei 2022 sampai dengan 20 Oktober 2022.

## 3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah jumlah kasus harian *monkeypox*. Variabel pada penelitian tersaji pada Tabel 1 berikut:

Tabel 1. Variabel Penelitian

Variabel	Keterangan
$Y_t$	Jumlah Kasus Harian <i>Monkeypox</i>

Data yang digunakan pada penelitian ini dibagi menjadi dua *dataset* yaitu data *training* untuk membentuk model SVR dan data *testing* sebagai evaluasi model SVR yang terbentuk dari data *training* yang nantinya akan dipakai untuk melihat akurasi peramalan. Pembagian *dataset* didasarkan pada kriteria *Woschnagg & Cipan (2004)* dimana data *training* dan data *testing* memiliki perbandingan rasio 8:2 dari total data.

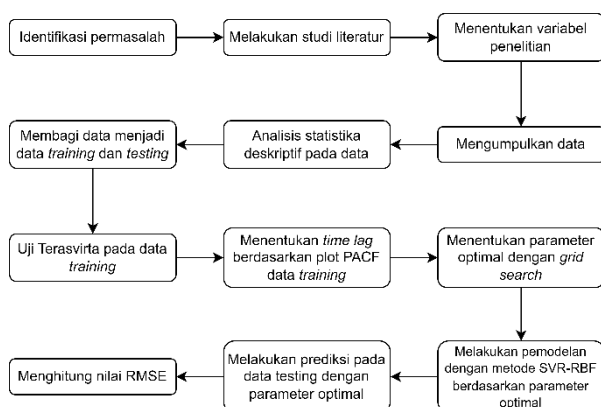
## 4. Prosedur Analisis

Langkah-langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mengidentifikasi masalah, dalam penelitian ini adalah bagaimana memprediksi secara akurat kasus harian *monkeypox* di seluruh dunia.
2. Melakukan studi literatur atau kepustakaan. Studi literatur terkait konsep peramalan dan prediksi menggunakan SVR.
3. Menentukan variabel penelitian, yaitu data jumlah kasus harian penyakit *monkeypox* di seluruh dunia.
4. Mengumpulkan data jumlah kasus harian penyakit *monkeypox* di seluruh dunia periode 29 Mei 2022 sampai dengan 20 Oktober 2022.

5. Melakukan analisis statistika deskriptif pada data.
6. Membagi data menjadi 2 *dataset* yaitu data *training* dan data *testing* dengan rasio 8:2.
7. Melakukan pengujian linearitas menggunakan uji Terasvirta pada data *training*.
8. Melakukan pemodelan pada data *training* dengan metode SVR menggunakan *software R*. Langkah-langkah pemodelan data dengan metode SVR adalah sebagai berikut:
  - a. Penentuan variabel *input* pada model SVR berdasarkan plot PACF data *training*.
  - b. Penentuan model terbaik dengan mencari parameter optimal dengan algoritma *grid search*.
  - c. Melakukan pemodelan data dengan metode SVR fungsi *Kernel Radial Basis Function (RBF)* berdasarkan parameter optimal.
  - d. Memprediksi data *testing* menggunakan model terbaik pada langkah b.
  - e. Menghitung nilai RMSE pada hasil pemodelan dan hasil prediksi pada model SVR.

Langkah-langkah analisis di atas dapat digambarkan dalam diagram alir penelitian pada Gambar 1 berikut.



Gambar 1. Diagram Alir Penelitian

## HASIL DAN PEMBAHASAN

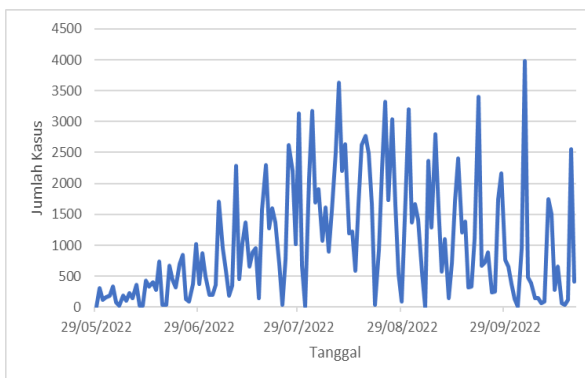
### Statistika Deskriptif

Berikut ini ditampilkan nilai statistika deskriptif dari data yang digunakan.

Tabel 2. Statistika Deskriptif Jumlah Kasus Harian Virus *Monkeypox*

Variabel	Mean	StDev	Max	Min	n
$Y_t$	1034,6	951,1	3986		2

Berdasarkan Tabel 2, dengan menggunakan jumlah kasus harian virus *monkeypox* di seluruh dunia pada tanggal 29 Mei 2022 hingga 20 Oktober 2022, diperoleh rata-rata kasus harian *monkeypox* sebesar 1034,6 atau 1034 kasus. Nilai simpangan baku hasil statistika deskriptif menunjukkan angka yang cukup besar yaitu 951,1 dengan nilai maksimum kasus harian *monkeypox* sebesar 3986.



Gambar 2. Jumlah Kasus Harian *Monkeypox* Dunia

Berdasarkan Gambar 2 terlihat mulai terjadi kenaikan secara signifikan jumlah kasus terjangkit *monkeypox* di seluruh dunia mulai dari akhir bulan Mei. Menurut informasi tersebut, berdasarkan laporan yang diterbitkan oleh *Reuters* tanggal 25 Mei 2022, lebih dari 20 negara non-endemik melaporkan kasus *monkeypox* dan telah mengonfirmasi lebih dari 100 kasus. Kemunculan kasus baru tersebut paling banyak terjadi di Eropa. Kemudian, tanggal 20 Juli 2022, peningkatan kasus sebanyak lebih dari 14.000 kasus di seluruh dunia dilaporkan *World Health Organization* (WHO). Sejak peningkatan ini, pada 23 Juli 2022, WHO menyatakan cacar monyet atau *monkeypox* sebagai darurat kesehatan global.

Dalam pemodelan dan prediksi yang melibatkan data *time series*, hal yang perlu diperhatikan adalah pola datanya. Dalam

Gambar 2 juga dapat ditunjukkan bahwa nilai simpangan baku yang cukup besar digambarkan melalui garis yang fluktuatif dengan *trend* yang cenderung naik tiap harinya.

### Pengujian Linearitas Data dengan Uji Terasvirta

Sebelum dilakukan analisis lebih lanjut, data jumlah kasus harian *monkeypox* akan diuji linearitasnya, untuk melihat apakah data cocok dimodelkan menggunakan metode nonlinear seperti SVR. Hasil pengujian linearitas pada data menggunakan uji Terasvirta disajikan pada Tabel 3.

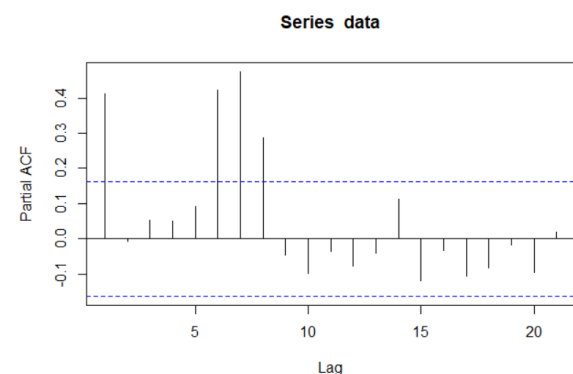
Tabel 3. Uji Terasvirta pada Data Jumlah Kasus Harian *Monkeypox*

Variabel	<i>p-value</i>
$Y_t$	0,004618

Berdasarkan Tabel 3, ditunjukkan bahwa nilai *p-value* hasil uji terasvirta sebesar 0,004618 lebih kecil dari taraf signifikan (0,05) sehingga dapat disimpulkan bahwa data jumlah kasus harian *monkeypox* mengandung pola nonlinear, sehingga data perlu dianalisis menggunakan model nonlinear untuk mendapatkan prediksi model yang lebih baik.

### Penentuan Variabel Input Berdasarkan Plot PACF

Langkah awal sebelum melakukan pemodelan menggunakan metode SVR adalah menentukan *lag* yang signifikan dari data dimana nantinya dipakai sebagai variabel *input* untuk model. *Lag* signifikan ditentukan berdasarkan plot PACF dari data. Plot PACF dari data disajikan pada Gambar 3.



Gambar 3. Plot PACF Data Jumlah Kasus Harian *Monkeypox* Dunia

Berdasarkan Gambar 3 tampak bahwa *lag* yang signifikan adalah *lag* 1, *lag* 6, *lag* 7, dan *lag* 8. Dengan demikian, variabel input pada model SVR pada penelitian ini adalah *lag* 1, *lag* 6, *lag* 7, dan *lag* 8.

**Pemodelan Data Kasus Harian *Monkeypox* Dunia Menggunakan SVR**

Saat pembentukan model SVR, perlu ditentukan parameter optimal agar model mampu memprediksi secara akurat. Pencarian parameter optimal dalam penelitian ini menggunakan algoritma *grid search* dengan parameter yang dicari adalah *C*,  $\gamma$  dan  $\epsilon$ . Penentuan nilai optimal dari parameter pada menggunakan *grid search* pada penelitian ini dilakukan melalui dua langkah yaitu langkah *loose grid* dan langkah *finer grid*.

Pada tahap awal proses pencarian parameter optimal pada algoritma *grid search* perlu ditentukan *range* nilai parameternya. Pada langkah *loose grid*, digunakan nilai pangkat bilangan bulat untuk parameter *C* dan  $\gamma$ . Sedangkan nilai  $\epsilon$  memiliki rentang nilai dari 0 hingga 1. Tabel 4 menunjukkan rentang nilai parameter yang diimplementasikan dalam pencarian parameter optimal pada langkah *loose grid*.

Tabel 4. Rentang Nilai Parameter Tahapan *Loose Grid* Metode *Grid Search*

Parameter	Rentang Nilai
<i>C</i>	$2^{-3}, 2^{-2}, \dots, 2^5, 2^6$
$\gamma$	$2, 2^2, \dots, 2^9, 2^{10}$
$\epsilon$	0,01, 0,02, ..., 0,1

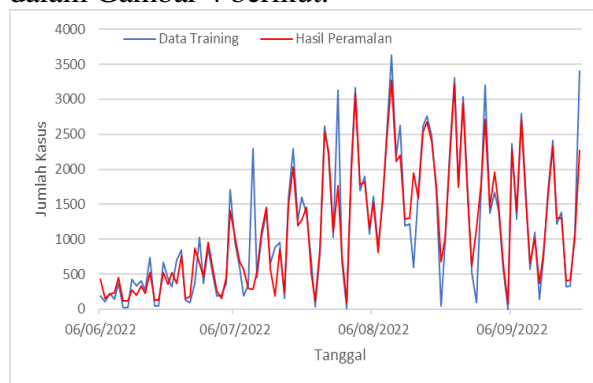
Berdasarkan Tabel 4, diperoleh nilai parameter optimal dari langkah *loose grid* untuk masing-masing parameter adalah *C* = 2,  $\gamma$  = 2, dan  $\epsilon$  = 0,09. Pada langkah selanjutnya yaitu tahap *finer grid* akan digunakan nilai persekitaran *C* dan  $\gamma$  pada langkah *loose grid* untuk mencari parameter optimalnya. Rentang nilai parameter yang diimplementasikan pada langkah *finer grid* ditunjukkan pada Tabel 5.

Tabel 5. Rentang Nilai Parameter Tahapan *Finer Grid* Metode *Grid Search*

Parameter	Rentang Nilai
<i>C</i>	$2^{0,2}, 2^{0,4}, \dots, 2^{1,6}, 2^{1,8}$

$\gamma$	$2^{0,2}, 2^{0,4}, \dots, 2^{1,6}, 2^{1,8}$
$\epsilon$	0,09

Berdasarkan hasil pencarian parameter optimal pada langkah *finer grid* diperoleh parameter optimalnya dari masing-masing parameter adalah *C* =  $2^{0,6}$ ,  $\gamma$  =  $2^{0,2}$ , dan  $\epsilon$  = 0,09. Tahapan berikutnya setelah mendapatkan model SVR terbaik adalah melihat seberapa akurat model SVR untuk peramalan data kasus harian *monkeypox* melalui nilai RMSE dan plot *time series* antara data *training* dengan hasil peramalan. Plot *time series* antara data *training* dengan hasil peramalan disajikan dalam Gambar 4 berikut.



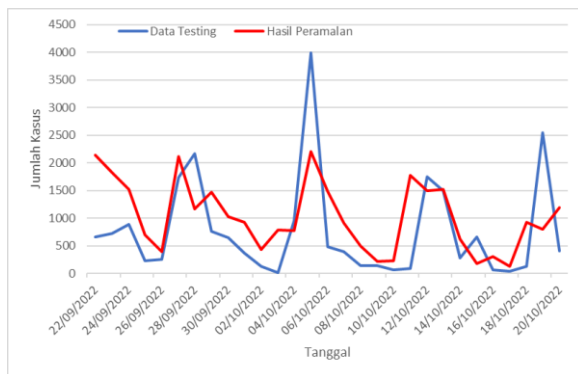
Gambar 4. Plot *Time Series* Hasil Peramalan dengan Data *Training*

Berdasarkan Gambar 4, terlihat bahwa model mampu memprediksi pada data *training* dengan cukup akurat. Hal tersebut terlihat dari hasil peramalan yang nilainya mendekati data aktualnya dan kesesuaian *trend* antara hasil peramalan dengan data aktualnya. Selain itu, diperoleh nilai RMSE pada data *training* sebesar 352,3. Sehingga, jika melihat hasil plot *time series* dan nilai RMSE pada *training*, dapat disimpulkan bahwa model SVR dengan parameter optimal mampu memprediksi data *training* dengan cukup baik.

**Peramalan Data Kasus Harian *Monkeypox* Dunia Menggunakan SVR**

Model SVR yang telah diperoleh kemudian diimplementasikan dalam meramalkan pada data *testing* dengan hasil plot *time series*-nya disajikan pada Gambar 5 berikut.





Gambar 5. Plot *Time Series* Hasil Peramalan dengan Data *Testing*

Berdasarkan Gambar 5 hasil peramalan nilainya mendekati nilai data *testing*-nya dengan nilai RMSE pada hasil peramalan data *testing*-nya diperoleh sebesar 809,7. Selain itu, hasil peramalan jika dilihat secara *trend*, memiliki *trend* yang sama dengan data *testing*-nya. Sehingga berdasarkan kedekatan nilai, kesamaan *trend*, dan nilai RMSE, maka model SVR dengan parameter optimal bisa dikatakan cukup baik dalam melakukan peramalan pada data *testing*.

## KESIMPULAN

Berdasarkan hasil analisis diperoleh kesimpulan bahwa hasil peramalan kasus harian *monkeypox* dunia menggunakan metode SVR, baik pada data *training* maupun *testing*-nya, nilai prediksinya mendekati nilai data aktualnya. Selain itu, hasil peramalan jika dilihat secara *trend*, memiliki *trend* yang sama dengan data aktualnya. Sehingga berdasarkan kedekatan nilai dan *trend* maka model SVR bisa dikatakan cukup baik dalam melakukan peramalan. Selain itu, diperoleh nilai RMSE pada data *training* dan *testing* masing-masing sebesar 352,3 dan 809,7. Saran yang bisa diberikan untuk penelitian selanjutnya adalah dimungkinkan untuk melakukan berbagai penyesuaian dalam pemilihan parameter SVR yang optimal, atau menggunakan algoritma pencarian parameter optimal yang berbeda. Selain itu, penelitian selanjutnya disarankan untuk membandingkan juga dengan metode *machine learning* lain sehingga dapat membandingkan prediksi yang lebih akurat.

## DAFTAR PUSTAKA

- Agustina, S., Agoestanto, A., dan Hendikawati P. 2017. Klasifikasi Tingkat Kesejahteraan Keluarga Jawa Tengah Tahun 2015 Menggunakan Metode Regresi Logistik Ordinal dan Support Vector Machine (SVM). *UNNES Journal of Mathematics*. Vol. 6(1), 59-69.
- Cabanillas, B., Valdelvira, R., dan Akdis, C. A. (2022). Monkeypox outbreak in Europe, UK, North America, and Australia: A changing trend of a zoonotic disease. *Allergy*, Vol. 77(8), 2284–2286. <https://doi.org/10.1111/all.15393>
- Cao, G., dan Wu, L. 2016. Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy*. Vol. 115, 734-745.
- Farid, N. (2022, Agustus 19). *Kesehatan*. Tirta.id.
- Haykin, S. 2008. *Neural Networks and Learning Machines 3rd Edition*. New York: Pearson.
- Hendayanti, N. P., Suniantara, I. K., dan Nurhidayati, M. 2019. Penerapan Support Vector Regression (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik ke Bali. *Jurnal Varian*. Vol. 3(1), 43-50.
- Hsu, C.-W., Chang, C.-C., dan Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*. Diambil kembali dari <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- Jamil, H., Tariq, W., Tahir, M. J., Mahfooz, R. S., Asghar, M. S., dan Ahmed, A. 2022. Human monkeypox expansion from the endemic to non-endemic regions: Control measures. *Annals of Medicine and Surgery*. Vol. 79. 104048.
- Kementerian Kesehatan RI. (2022). *Pedoman Pencegahan dan Pengendalian Penyakit Monkeypox*. Jakarta: Kementerian Kesehatan RI.
- Lestari, M. P., Winarsyah, D. J., dan Hamami, F. 2021. Peramalan

- Pertambahan Pasien Covid-19 Menggunakan Support Vector Regression. *eProceedings of Engineering*. Vol. 8(5), 9497.
- Liu, C.-J., Lee, T.-S., dan Chiu, C.-C. 2009. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*. Vol. 47(2), 115-125.
- Purnama, D. I., Hendarsin, O. P. 2020. Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jambura Journal of Mathematics*. Vol. 2(2), 49-59.
- Smola, A. J., dan Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and Computing*. Vol. 14(3), 199–222.
- Suhenda, D. (2022, Agustus 3). *IDI forms task force to prevent monkeypox outbreak in Indonesia*. The Jakarta Post.
- Suprayogi, D., dan Pardede, H. F. 2022. Support Vector Regression Dalam Prediksi Penurunan Jumlah Kasus Penderita Covid-19. *Journal of Information Technology and Computer Science*. Vol. 7(2), 63-70.
- Syarif, I., Prugel-Bennett, A., dan Wills, G. 2016. SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA*. Vol. 14(4), 1502-1509.
- Terasvirta, T., Linc, F., dan Granger, C., W., J. 1993. Power of The Neural Networks Linearity Test. *Journal of Time Series Analysis*. Vol. 14(2), 159-171.
- Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., Palich, R., Nori, A., Reeves, I., Habibi, M. S., Apea, V., Boesecke, C., Vandekerckhove, L., Yakubovsky, M., Sendagorta, E., Blanco, J. L., Florence, E., Moschese, D., Maltez, F. M., Goorhuis, A., Pourcher, V., Migaud, P., Noe, S., Pintado, C., Maggi, F., Hansen, A.-B. E., Hoffmann, C., Lezama, J. I., Mussini, C., Cattelan, A. M., Makofane, K., Tan, D., Nozza, S., Nemeth, J., Klein, M. B., dan Orkin, C. M. 2022. Monkeypox Virus Infection in Humans across 16 Countries - April-June 2022. *The New England Journal of Medicine*. Vol. 387(8), 679-691.
- Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.
- Wijaya, L. D. (2022, Agustus 22). *Indonesia's First Monkeypox Patient Visited Western Europe in Early August*. Tempo.co.
- Woschnagg, E., Cipan, J. 2004. *Evaluating Forecast Accuracy*. Austria: University of Vienna.
- Yasin, H., Prahutama, A., dan Utami, T. W. 2014. Prediksi Harga Saham Menggunakan Support Vector Regression dengan Algoritma Grid Search. *Media Statistika*. Vol. 7(1), 29-35.



# PENERAPAN *MODEL-BASED CLUSTERING* PADA PENGELOMPOKAN SAHAM BERDASARKAN RASIO KEUANGAN

Irena Sekar Dwi Hasnida<sup>1</sup>, Rosita Kusumawati<sup>2</sup>

<sup>1,2</sup>Statistika, Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta  
e-mail: <sup>1</sup>irenasekardwihasnida@gmail.com

*Received: January 2, 2023; Revised: June 27, 2023; Accepted: June 27, 2023; Published: June 30, 2023*

## Abstrak

Untuk meminimalkan kerugian dengan tingkat keuntungan tertentu, investor perlu memilih saham potensial agar keuntungan yang diperoleh optimal. Penelitian ini bertujuan untuk mengetahui penerapan *Model-Based Clustering* (MBC) dalam mengelompokkan perusahaan berdasarkan kinerja keuangan saham. Indikator keuangan yang digunakan yaitu data rasio likuiditas, profitabilitas, dan solvabilitas tahun 2020 untuk perusahaan yang terdaftar pada indeks LQ45. Pemilihan model dan jumlah komponen akan dilakukan berdasarkan nilai *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL). Dari proses *clustering*, terbentuk 6 *cluster* dengan nilai BIC dan ICL berturut-turut sebesar 709,3757 dan -709,376, dan model optimal terpilih VEV (*Variable volume, Equal shape, Variable orientation*). Berdasarkan nilai rata-rata setiap rasio, *cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profitabilitas terbaik serta rasio solvabilitas terendah. *Cluster* 6 memiliki kemampuan yang tinggi dibandingkan perusahaan *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban keuangannya jangka pendek.

**Kata kunci:** *Model-Based Clustering*, pengelompokan saham, rasio keuangan.

## Abstract

*To minimize losses at a certain profit level, investors need to choose potential stocks so that the profits are optimal. This study aims to determine the application of model-based clustering (MBC) in classifying companies based on stock performance. The financial indicators used are data on liquidity ratios, profitability, and solvency for 2020 for companies listed on the LQ45 index. The selection of the model and the number of components will be based on the Bayesian Information Criterion (BIC) and Completed Likelihood Criterion (ICL) values. From the clustering process, six clusters were formed with BIC and ICL values of 709.3757 and -709.376, respectively, and the optimal model was selected as VEV (variable volume, equal shape, variable orientation). Based on the average value of each ratio, cluster 6 is the best cluster because it has the majority of the best liquidity and profitability ratios and the lowest solvency ratio. Cluster 6 has a high ability, compared to other cluster companies, to finance the company's operational activities and fulfill its short-term financial obligations.*

**Keywords:** *Model-Based Clustering*, stock clustering, financial ratios.

## PENDAHULUAN

Pengambilan keputusan investasi oleh investor perlu dibuat secara rasional untuk memaksimalkan keuntungan. Dalam hal ini analisis fundamental perusahaan dapat dimanfaatkan oleh investor sebagai pertimbangan dalam keputusan investasinya. Analisis fundamental bertujuan untuk memperkirakan harga saham di masa mendatang dengan memperkirakan nilai faktor fundamental yang akan mempengaruhi harga saham di masa mendatang, seperti perkiraan laba per saham, perputaran pendapatan perusahaan, risiko pendapatan masa depan perusahaan, penggunaan hutang oleh manajemen dan kebijakan deviden (Weston dan Brigham, 1993). Analisis fundamental melihat kondisi internal perusahaan salah satunya melalui analisis laporan keuangan menggunakan rasio-rasio keuangan yang menginterpretasikan kondisi keuangan dan hasil operasi suatu perusahaan (Suad, 2004).

Untuk menilai keuntungan perusahaan dapat dianalisis melalui data yang tercermin dalam laporan keuangan perusahaan, dengan menganalisis laporan keuangan perusahaan sendiri (Andy dan Megawati, 2019). Analisis rasio keuangan adalah analisis yang menggabungkan evaluasi laporan keuangan dan laporan laba rugi terhadap satu dengan lainnya, yang memberikan gambaran tentang sejarah perusahaan dan penilaian kesehatan suatu perusahaan tertentu. Analisis rasio dapat membimbing investor membuat keputusan atau pertimbangan tentang apa yang akan dicapai oleh perusahaan dan atau bagaimana prospek yang akan dihadapi dimasa yang akan datang (Andayani, 2016). Masing-masing rasio keuangan mencerminkan aspek tertentu. Saat menganalisis rasio keuangan, pemberi pinjaman mencari informasi tentang kemampuan perusahaan untuk membayar kembali pinjamannya tepat waktu, sementara investor lebih tertarik pada kemampuan perusahaan untuk menghasilkan keuntungan.

Beberapa penelitian telah dilakukan pada portofolio dan strategi pemilihan aset/saham. Salah satunya dengan melakukan pengelompokan terhadap saham menggunakan teknik *clustering* seperti yang dilakukan oleh Nanda et.al (2010) yang memanfaatkan beberapa teknik clustering

pada saham yaitu *K-means*, *SOM (Self-Organization Map)*, dan *Fuzzy C-means* yang menggunakan data pasar Bombay Stock Exchange. Subekti dkk (2017) juga melakukan *clustering* saham menggunakan *K-Means* dan *Average Linkage* untuk menyusun portofolio yang optimal dalam rangka strategi diversifikasi yang makin beragam. Pada penelitian lainnya, Subekti et.al (2018) menggunakan *Ant colony algorithm* untuk *clustering* dapat memberikan portofolio dengan performa yang lebih baik berdasarkan nilai *Sharpe index*.

Dalam perspektif *data mining*, masalah pemilihan saham bertujuan untuk mengidentifikasi saham potensial atau saham yang memberikan keuntungan tinggi. *Clustering* saham bisa digunakan untuk mengelompokkan saham berdasarkan beberapa indikator kinerja perusahaan. Seorang investor bisa memilih saham hasil analisis karakteristik saham dari *clustering* yang terbentuk. Terdapat beberapa jenis clustering yang digunakan, diantaranya adalah *Centroid-based Clustering*, *Density-based Clustering*, *Distribution-based Clustering*, dan *Hierarchical Clustering*. Pada artikel ini penulis akan menggunakan *Model/Distribution-based Clustering* sebagai metode *clustering* saham. Jenis *clustering* ini mengelompokkan data berdasarkan fungsi peluang campuran atau distribusi campuran (*mixture distribution*), dimana setiap kelompok dimodelkan oleh fungsi peluangnya sendiri. *Expectation-maximization* merupakan algoritma yang digunakan untuk mengimplementasikan metode *Model-Based Clustering*.

*Model-Based Clustering (MBC)*, menganggap sebaran data mengikuti distribusi campuran (*mixture distribution*). *Mixture distribution* adalah gabungan dari beberapa komponen distribusi nilai keanggotaan data pada suatu cluster ditentukan oleh nilai peluang terbesar dari komponen distribusi. Tidak seperti metode *k-means*, *MBC* menggunakan *soft assignment*, di mana setiap titik data memiliki probabilitas untuk menjadi milik setiap kelompok. Qona'ah et. al. (2020) melakukan pengelompokan pada 100 laboratorium di Institut Teknologi Sepuluh Nopember (ITS) berdasarkan produktivitas

laboratorium dalam menjalankan fungsinya. Penelitian ini membandingkan metode *K-Means*, *K-Medoids*, and *Model-Based Clustering* (MBC) dalam proses pengelompokan. Hasil penelitian menunjukkan bahwa MBC memberikan performa yang lebih baik dari *K-Means* dan *K-Medoids*.

Penelitian ini bertujuan untuk mengetahui penerapan algoritma MBC untuk mengelompokkan saham berdasarkan rasio keuangan sebagai indikator kinerja perusahaan dengan MBC berdasarkan distribusi multivariate mixture gaussian. Eksplorasi *cluster* akan dilakukan serta rata-rata *return* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk.

## METODOLOGI

Penelitian ini menerapkan *Model-Based Clustering* (MBC) untuk membentuk *cluster* perusahaan terlebih dahulu berdasarkan nilai rasio keuangan. Selanjutnya dilakukan eksplorasi data terhadap masing-masing kelompok yang terbentuk dari proses *clustering*. Rata-rata *return* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk. Data yang digunakan untuk proses *clustering* merupakan data rasio keuangan saham pada indeks LQ45 yang tersedia pada [website www.yahoofinance.com](http://www.yahoofinance.com) periode tahun 2020. Jumlah saham yang digunakan hanya sebanyak 36 saham, saham ini dipilih dengan pertimbangan kelengkapan data yang tersedia.

### Rasio Keuangan

Rasio keuangan dihitung dengan menggunakan nilai numerik yang diambil dari laporan keuangan untuk mendapatkan informasi yang berarti tentang perusahaan. Ini memungkinkan untuk mengikuti kinerja perusahaan dari waktu ke waktu dan menemukan tanda-tanda masalah. Sehingga dalam penelitian ini rasio keuangan akan digunakan sebagai variabel penelitian.

Menurut Munawir (2002) ada 4 kelompok rasio keuangan yaitu rasio likuiditas, rasio aktivitas, rasio profitabilitas dan rasio solvabilitas. Artikel ini menggunakan 3 aspek rasio keuangan, diantaranya adalah Rasio

Likuiditas, Rasio Probabilitas, dan Rasio Solvabilitas. Secara lebih terperinci, variabel dari masing-masing aspek rasio keuangan yang digunakan akan dijelaskan pada poin-poin berikut:

a. Rasio likuiditas adalah rasio yang menentukan kemampuan perusahaan untuk membiayai operasi dan memenuhi kewajiban keuangannya jangka pendek. Rasio ini biasanya digunakan terutama untuk menilai kemampuan perusahaan dalam memenuhi kewajibannya. Adapun untuk jenis rasio likuiditas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

#### 1) *Cash Ratio*

Persamaan untuk menghitung nilai *cash ratio* adalah sebagai berikut:

$$\text{Cash Ratio} = \frac{\text{Cash or Cash Equipment}}{\text{Current Liabilities}}$$

#### 2) *Current Ratio*

Persamaan untuk menghitung nilai *current ratio* adalah sebagai berikut:

$$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$$

#### 3) *Quick Ratio*

Persamaan untuk menghitung nilai *quick ratio* adalah sebagai berikut:

$$\text{Quick Ratio} = \frac{\text{Current Assets} - \text{Liabilities}}{\text{Current Liabilities}}$$

b. Rasio profitabilitas merupakan rasio yang menentukan kemampuan perusahaan untuk menghasilkan laba dari berbagai kebijakan dan keputusan yang telah diambil. Rasio profitabilitas menilai kemampuan perusahaan untuk menghasilkan pendapatan, aset, dan keuntungan modal, dan mengukur pengembalian yang diperoleh dari modal perusahaan. Adapun untuk jenis rasio profitabilitas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

#### 1) *Return on Assets* (ROA)

Persamaan untuk menghitung nilai *return on assets* adalah sebagai berikut:

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$$

#### 2) *Return on Equity* (ROE)

Persamaan untuk menghitung nilai return on equity adalah sebagai berikut:

$$ROE = \frac{Net\ Income}{Equity}$$

### 3) Gross Profit Margin

Persamaan untuk menghitung nilai gross profit margin adalah sebagai berikut:

$$Gross\ Profit\ Margin = \frac{Net\ Sales - Cost\ of\ Goods\ Sold}{Sales}$$

### 4) Net Profit Margin

Persamaan untuk menghitung nilai net profit margin adalah sebagai berikut:

$$Net\ Profit\ Margin = \frac{Net\ Income}{Sales}$$

- c. Rasio solvabilitas adalah rasio untuk mengukur seberapa jauh aset perusahaan dibiayai oleh hutang. Rasio ini mengukur perbandingan dana yang disediakan oleh pemiliknya dengan dana yang dipinjam dari kreditur perusahaan tersebut. Semakin tinggi rasio ini, semakin sedikit ekuitas dibandingkan dengan hutang. Bagi perusahaan, sebaiknya besarnya hutang tidak boleh melebihi modal sendiri agar beban tetapnya tidak terlalu tinggi. Adapun untuk jenis rasio solvabilitas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

#### 1) Debt to Equity Ratio

Persamaan untuk menghitung nilai debt to equity ratio adalah sebagai berikut:

$$Debt\ to\ Equity\ Ratio = \frac{Total\ Debt}{Equity}$$

#### 2) Long-term Debt Ratio

Persamaan untuk menghitung nilai long-term debt ratio adalah sebagai berikut:

$$Long - term\ Debt\ Ratio = \frac{Long - term\ Debt}{Total\ Assets}$$

### Model-Based Clustering (MBC)

MBC merupakan algoritma *clustering* yang dikembangkan berdasarkan model probabilistik *finite mixture* untuk *probability densities*. Kata “model” dalam metode tersebut biasa digunakan untuk menunjukkan jenis kendala dan sifat geometris dari matriks kovarians (Guojun Gan et. al., 2007). Dalam pendekatan MBC, data dipandang berasal dari

campuran distribusi probabilitas, data berdistribusi *finite mixture* atau berdistribusi campuran. Distribusi mixture merupakan distribusi yang tersusun dari beberapa komponen distribusi.

Algoritma *clustering* tradisional seperti *K-means* dan pengelompokan hierarki lainnya adalah algoritma berbasis heuristik yang menurunkan kluster secara langsung berdasarkan data daripada mempertimbangkan ukuran probabilitas dari masing-masing data (Kassambra, 2017). Tidak seperti metode tersebut, MBC mencoba menggunakan *soft assignment*, di mana setiap titik data memiliki probabilitas untuk menjadi milik setiap kelompok. MBC menggabungkan korelasi antar variabel ke dalam matriks jaraknya, sehingga lebih fleksibel terhadap bentuk dan ukuran *cluster* (Kessler, 2019).

Menurut Fraley dan Raftery (2002), untuk menemukan jumlah *cluster* dan model *clustering* terbaik, dibutuhkan penyelesaian dari masalah pemilihan model statistik, dimana model dengan jumlah dan jenis distribusi komponen yang berbeda dibandingkan. MBC mengimplementasikan teori ini dengan menggunakan *Expectation Maximization* (EM) yang diinisialisasi dan menggunakan kriteria *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL) untuk membandingkan model. Oleh karena itu, MBC memberikan dasar statistika yang lebih kuat untuk melakukan inferensi dan interpretasi karena berbasis distribusi.

MBC adalah jenis pemodelan *finite mixture*, dengan asumsi bahwa data berasal dari campuran subpopulasi yang berbeda mengikuti distribusi yang diberikan, biasanya menggunakan multivariat normal (Gergely dan Vargha, 2021). *Mixture distribution* adalah distribusi yang terbentuk dari beberapa komponen penyusun distribusi. Persamaan *mixture* yang diadaptasi dari Bouveyron et. al. (2019) adalah sebagai berikut:

$$p(\mathbf{x}_i) = \sum_{k=1}^K \tau_k f_k(\mathbf{x}_i | \boldsymbol{\theta}_k) \quad (1)$$

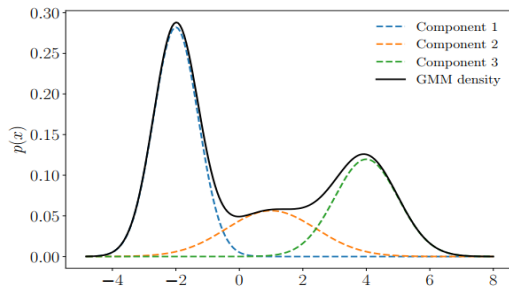
Dalam persamaan (1),  $\pi_k$  merupakan bobot probabilitas bahwa suatu pengamatan dihasilkan oleh komponen ke- $k$ , dengan batasan  $\pi_k \geq 0, k = 1, \dots, K$  dan  $\sum_{k=1}^K \pi_k = 1$ . Sedangkan  $f_k(\mathbf{x}_n | \boldsymbol{\theta}_k)$  adalah kepadatan komponen ke- $k$  dengan parameter  $\boldsymbol{\theta}_k$ .

a. *Gaussian Mixture Model (GMM)*

Distribusi probabilitas data yang digunakan pada artikel ini adalah distribusi gaussian, sehingga dari model *finite mixture* sebelumnya akan diarahkan pada *Gaussian Mixture Models (GMM)*. GMM adalah model densitas yang terdiri dari sejumlah distribusi  $K$  Gaussian  $N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ , sehingga dari persamaan (1) GMM bisa dituliskan (Bishop, 2006):

$$p(\mathbf{x}_i|\boldsymbol{\theta}_k) = \sum_{k=1}^K \tau_k N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (2)$$

GMM memberikan pendekatan klasik dan kuat untuk analisis *clustering* (Banfield dan Raftery, 1993), GMM juga berguna untuk memahami dan menyarankan kriteria pengelompokan yang kuat (Gan, Ma, dan Wu 2007). Ilustrasi GMM bisa dilihat pada Gambar 1.



Gambar 1. Ilustrasi *Gaussian Mixture Model*.

(Deisenroth, Faisal, & Ong, 2020)

gaussian dan lebih ekspresif dari komponen individu. Garis putus-putus mewakili komponen gaussian tertimbang. Fungsi densitas peluang dari distribusi multivariate GMM adalah sebagai berikut:

$$N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (3)$$

b. *Algoritma Expectation-Maximization (EM) untuk GMM*

Seperti metode *k-means* yang melakukan iterasi parameter untuk mendapatkan jarak yang konvergen dari pusat *cluster*, sehingga didapatkan hasil *cluster* berdasarkan jarak terdekat. *Model-Based Clustering (MBC)* juga memerlukan iterasi untuk update parameter. Proses iterasi untuk mendapatkan estimasi parameter *clustering* dilakukan dengan menggunakan algoritma EM untuk menyederhanakan perolehan estimasi parameter. Algoritma EM adalah

algoritma optimasi iterative untuk memaksimalkan fungsi likelihood dari model probabilistik dengan data yang hilang (*missing data*). Pada artikel ini algoritma EM digunakan untuk memaksimalkan fungsi ekspektasi likelihood dengan data yang tidak lengkap, yaitu data berasal dari komponen distribusi tertentu.

Metode ini merupakan pendekatan paling umum digunakan untuk mencari kemungkinan maksimum data yang terdiri dari  $n$  pengamatan multivariate  $(x_n, z_n)$ , di mana  $x_n$  merupakan variabel teramati dan  $z_n$  tidak teramati atau variabel laten (label kelompok belum diketahui). Jika label ini diketahui, maka akan didapatkan estimasi parameter di setiap distribusi komponen dengan membagi observasi ke dalam kelompok masing-masing. Tujuan utama dari proses ini adalah untuk mengetahui parameter dari distribusi yaitu  $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ . Berikut ini adalah algoritma EM untuk mencari parameter distribusi *mixture* pada kasus distribusi gaussian:

1. Inisialisasi parameter awal  $\tau^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)}$
2. Langkah Ekspektasi (*E-step*)

Evaluasi probabilitas posterior  $v_{i,k}^{(s+1)}$  untuk setiap data point menggunakan parameter saat ini. Dimana probabilitas posterior diberikan oleh  $v_{i,k}^{(s+1)} = \frac{\tau_k^{(s)} N_k(x_i|\boldsymbol{\theta}_k^{(s)})}{\sum_{j=1}^K \tau_j^{(s)} N_j(x_i|\boldsymbol{\theta}_j^{(s)})}$ . Oleh karena penelitian ini menggunakan pendekatan distribusi *multivariate gaussian*,  $v_{i,k}^{(s+1)}$  bisa dituliskan sebagai berikut:

$$v_{i,k}^{(s)} = \frac{\tau_k^{(s)} |\boldsymbol{\Sigma}_k^{(s)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \boldsymbol{\mu}_k^{(s)})^T [\boldsymbol{\Sigma}_k^{(s)}]^{-1} (x_i - \boldsymbol{\mu}_k^{(s)})\right)}{\sum_l \tau_l^{(s)} |\boldsymbol{\Sigma}_l^{(s)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \boldsymbol{\mu}_l^{(s)})^T [\boldsymbol{\Sigma}_l^{(s)}]^{-1} (x_i - \boldsymbol{\mu}_l^{(s)})\right)} \quad (4)$$

3. Langkah Maksimisasi (*M-step*)

Langka Maksimisasi merupakan estimasi ulang parameter  $\tau^{(s)}, \boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}$  menggunakan  $v_{i,k}^{(s+1)}$  yang diperoleh dari *E-step*. Langkah ini dilakukan dengan persamaan yang diperoleh dari turunan ekspektasi log likelihood dengan parameter yang akan dicari. Menurut Deisenroth, Faisal, & Ong (2020),

turunan untuk mendapatkan nilai estimasi parameter dapat dilakukan dengan langkah-langkah berikut:

a.) Mendapatkan bobot ( $\tau_k$ )

Untuk menemukan turunan parsial log likelihood dari bobor parameter  $\tau_k, k = 1, \dots, K$ , dengan kendala  $\sum_{k=1}^K \tau_k = 1$  dapat digunakan *Lagrange multipliers* sebagai berikut:

$$\mathcal{L} = \sum_{i=1}^n \log \sum_{k=1}^K \tau_k N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + \lambda (\sum_{k=1}^K \tau_k - 1) \quad (5)$$

Maka untuk menentukan turunan partial terhadap  $\pi_k$  adalah sebagai berikut:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \pi_k} &= \sum_{i=1}^n \frac{N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j N(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda \\ &= \frac{1}{\tau_k} \sum_{i=1}^n v_{i,k}^{(s+1)} + \lambda \end{aligned}$$

Dan turunan parsial terhadap  $\lambda$  adalah sebagai berikut:

$$\frac{\partial \mathcal{L}}{\lambda} = \sum_{k=1}^K \tau_k - 1$$

Dengan menyamakan kedua persamaan diatas menjadi 0, diperoleh

$$\tau_k = - \left( \sum_{i=1}^n v_{i,k}^{(s+1)} \right) / \lambda \quad (6)$$

$$\sum_{k=1}^K \tau_k = 1 \quad (7)$$

Dari persamaan diatas  $\pi_k^{(s+1)}$  diperoleh:

$$\begin{aligned} \sum_{k=1}^K \tau_k &= 1 \\ - \sum_{k=1}^K \left( \sum_{i=1}^n v_{i,k}^{(s+1)} \right) / \lambda &= 1 \\ - \sum_{k=1}^K \left( \sum_{i=1}^n v_{i,k}^{(s+1)} \right) &= \lambda \end{aligned}$$

Substitusikan  $\lambda$  ke persamaan (6), sehingga didapatkan:

$$\tau_k^{(s+1)} = \frac{\sum_{i=1}^n v_{i,k}^{(s+1)}}{\sum_{k=1}^K \sum_{i=1}^n v_{i,k}^{(s+1)}} \quad (8)$$

b.) Mendapatkan  $\boldsymbol{\mu}$

$$\begin{aligned} \frac{\partial \log p(\mathbf{x}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \frac{\partial p(\mathbf{x}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} \\ &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \sum_{j=1}^K \tau_j \frac{\partial N(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\partial \boldsymbol{\mu}_k} \right) \\ &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \tau_k \frac{\partial N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\mu}_k} \right) \\ &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \tau_k (\mathbf{x}_i - \boldsymbol{\mu}_k) \boldsymbol{\Sigma}^{-1} N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\ &= \sum_{i=1}^n \frac{\tau_k N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \tau_j N(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} (\mathbf{x}_i - \boldsymbol{\mu}_k) \boldsymbol{\Sigma}^{-1} \end{aligned}$$

$$= \sum_{i=1}^n v_{i,k}^{(s+1)} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}$$

Sehingga dengan menyamakan persamaan di atas dengan 0, diperoleh:

$$\begin{aligned} \sum_{i=1}^n v_{i,k}^{(s+1)} \mathbf{x}_i &= \sum_{i=1}^n v_{i,k}^{(s+1)} \boldsymbol{\mu}_k^{(s+1)} \\ \boldsymbol{\mu}_k^{(s+1)} &= \frac{\sum_{i=1}^n v_{i,k}^{(s+1)} \mathbf{x}_i}{\sum_{i=1}^n v_{i,k}^{(s+1)}} \quad (9) \end{aligned}$$

c.) Mendapatkan  $\boldsymbol{\Sigma}$

$$\begin{aligned} \frac{\partial \log p(\mathbf{x}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \frac{\partial p(\mathbf{x}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} \\ &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \tau_k (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \right) \right) \\ &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \tau_k (2\pi)^{-\frac{d}{2}} \left[ \frac{\partial}{\partial \boldsymbol{\Sigma}_k} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \right) + |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp \left( -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \right) \right] \right) \end{aligned}$$

Dengan menggunakan identitas untuk menghitung gradien berdasarkan Deisenroth, Faisal, dan Ong (2020),

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} &\text{ dan } \frac{\partial}{\partial \boldsymbol{\Sigma}_k} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \\ &\text{ dapat diselesaikan sebagai berikut:} \\ \frac{\partial}{\partial \boldsymbol{\Sigma}_k} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} &= -\frac{1}{2} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \boldsymbol{\Sigma}_k^{-1} \\ \frac{\partial}{\partial \boldsymbol{\Sigma}_k} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) &= \\ &= -\boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} \end{aligned}$$

Sehingga,

$$\begin{aligned} \frac{\partial \log p(\mathbf{x}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} &= \sum_{i=1}^n \frac{1}{p(\mathbf{x}_i | \boldsymbol{\theta})} \left( \tau_k N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \left[ -\frac{1}{2} (\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}) \right] \right) \\ &= \sum_{i=1}^n \frac{\tau_k N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \tau_j N(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \left[ -\frac{1}{2} (\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}) \right] \\ &= -\frac{1}{2} \sum_{i=1}^n v_{i,k}^{(s+1)} (\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}) \\ &= -\frac{1}{2} \boldsymbol{\Sigma}_k^{-1} \sum_{i=1}^n v_{i,k}^{(s+1)} \\ &\quad + \frac{1}{2} \boldsymbol{\Sigma}_k^{-1} \left( \sum_{i=1}^n v_{i,k}^{(s+1)} (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \right) \boldsymbol{\Sigma}_k^{-1} \end{aligned}$$

Oleh karena  $\sum_{i=1}^n v_{i,k}^{(s+1)} = N_k$  dan dengan menyamakan persamaan di atas dengan 0, maka:

$$N_k \Sigma_k^{-1} = \Sigma_k^{-1} \left( \sum_{i=1}^n v_{i,k}^{(s+1)} (x_i - \mu_k)(x_i - \mu_k)^T \right) \Sigma_k^{-1}$$

$$N_k I = \Sigma_k^{-1} \left( \sum_{i=1}^n v_{i,k}^{(s+1)} (x_n - \mu_k)(x_n - \mu_k)^T \right)$$

$$\frac{I}{\Sigma_k^{-1}} = \frac{1}{N_k} \left( \sum_{i=1}^n v_{i,k}^{(s+1)} (x_i - \mu_k)(x_i - \mu_k)^T \right)$$

Sehingga  $\Sigma_k^{(s+1)}$  dapat dituliskan:

$$\Sigma_k^{(s+1)} = \frac{1}{\sum_{i=1}^n v_{i,k}^{(s+1)}} \left( \sum_{i=1}^n v_{i,k}^{(s+1)} (x_i - \mu_k)(x_i - \mu_k)^T \right) \quad (10)$$

4. Ulangi terus-menerus langkah tersebut sampai mendapatkan hasil iterasi yang konvergen jika iterasi belum konvergen kembali ke langkah b).
5. Batas Geometris *Multivariate Gaussian Mixture Model*

Parameter yang perlu diduga untuk  $\Sigma_k$  dipersamaan (7) terlalu banyak sehingga dilakukan pendekatan baru untuk menduga  $\Sigma_k$ . Banfield dan Raftery (1993) mengusulkan kerangka umum untuk batasan *cross-cluster* geometris dalam *multivariate mixture model* dengan parameterisasi matriks kovarians melalui dekomposisi nilai eigen, hal tersebut bisa dilakukan dengan persamaan berikut (Guojun Gan et.al, 2007):

$$\Sigma_k = \lambda_k D_k A_k D_k^T \quad (11)$$

Menurut Bouveyron et. al. (2019) terdapat 14 kemungkinan model dari matriks kovarians  $\Sigma_k$  dalam MBC (lihat Tabel 2.1 pada buku Bouveyron et. al., 2019), 14 model tersebut berbeda untuk data, bentuk (bola atau ellipsoidal) dan volume. Dalam kasus model ellipsoidal, penyesuaian sumbu dan perbedaan bentuk ellipsoidal yang dipasang ditentukan. Ini dikenal sebagai dekomposisi *Volume-Shape-Orientation* (VSO). Untuk model tertentu, volume, bentuk, dan orientasi dapat dibatasi ke varian yang sama, dilambangkan dengan 'E'. Jika varians bebas berubah, model dilambangkan 'V'. Selain itu, orientasi klaster relatif terhadap satu sama lain dapat dibatasi ke *Equal* (sama) atau *Varying* (bervariasi), atau model dapat memiliki keselarasan terbatas pada sumbu koordinat, dan diberi label 'T'. Sebagai contoh (Kassambara, 2017):

1. EVI menunjukkan model di mana volume semua *cluster* adalah sama (E), bentuk *cluster* dapat bervariasi (V), dan

orientasinya adalah identitas (I) atau "sumbu koordinat".

2. EEE berarti *cluster* memiliki volume, bentuk, dan orientasi yang sama dalam ruang dimensi-d.
3. VEI berarti *cluster* memiliki volume variabel, bentuk dan orientasi yang sama dengan sumbu koordinat.

MBC menerapkan 14 model dan mengidentifikasi salah satu yang model yang paling mendefinisikan data. Untuk memilih model yang optimal penelitian ini akan menggunakan *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL), kemudian hasil clustering terpilih akan dibandingkan dengan metode *K-means* menggunakan indeks *Davies-Bouldin* (DB).

## HASIL DAN PEMBAHASAN

Penelitian ini mendeskripsikan algoritma *Mode-Based Clustering* (MBC) untuk membentuk *cluster* perusahaan terlebih dahulu berdasarkan nilai rasio keuangan. Selanjutnya Eksplorasi *cluster* akan dilakukan serta rata-rata *return* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk

Proses pengelompokan saham berdasarkan nilai rasio keuangan dilakukan dengan menggunakan fungsi *Mclust* pada *package mclust* (Fraley, Raftery, dan Scrucca, 2016) pada program R. Jumlah model yang mampu diidentifikasi menggunakan *package* tersebut ada sebanyak 14 model dengan jumlah kelompok maksimal 9 kelompok apabila tidak ada spesifikasi jumlah kelompok tertentu dari peneliti. Estimasi parameter MBC menggunakan *gaussian mixture model* dan dilakukan dengan algoritma *Expectation-Maximization* (EM). Pemilihan model terbaik didasarkan pada kriteria *Bayesian information criterion* (BIC) dan *Completed Likelihood Criterion* (ICL). Model dengan nilai BIC dan ICL tertinggi diantara semua iterasi yang dilakukan akan dipilih menjadi model terbaik.

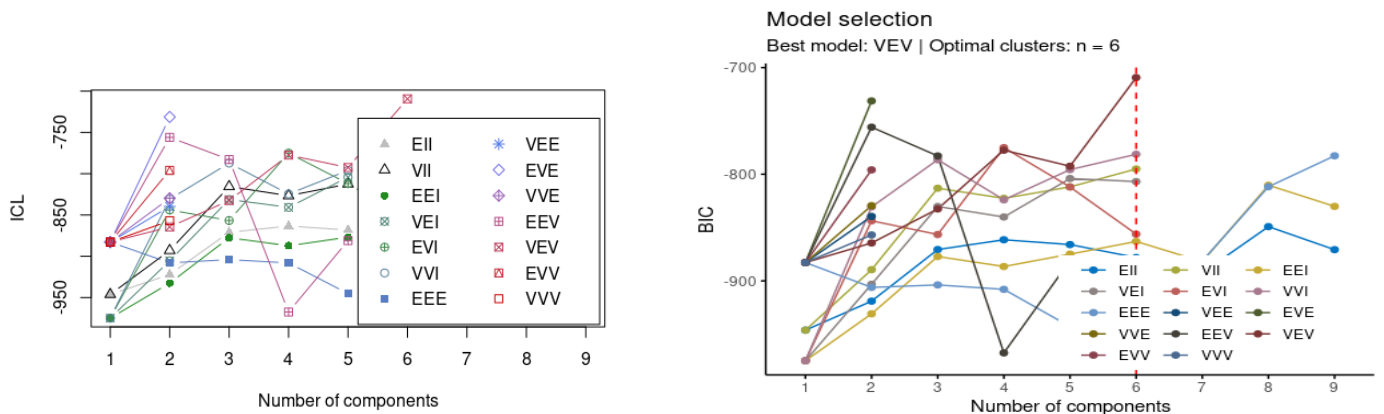
Jumlah komponen yang diperoleh ekuivalen dengan jumlah *cluster* optimal. Proses perhitungan MBC menggunakan program R, mengidentifikasi model optimal dan jumlah *cluster* untuk data. Hasil



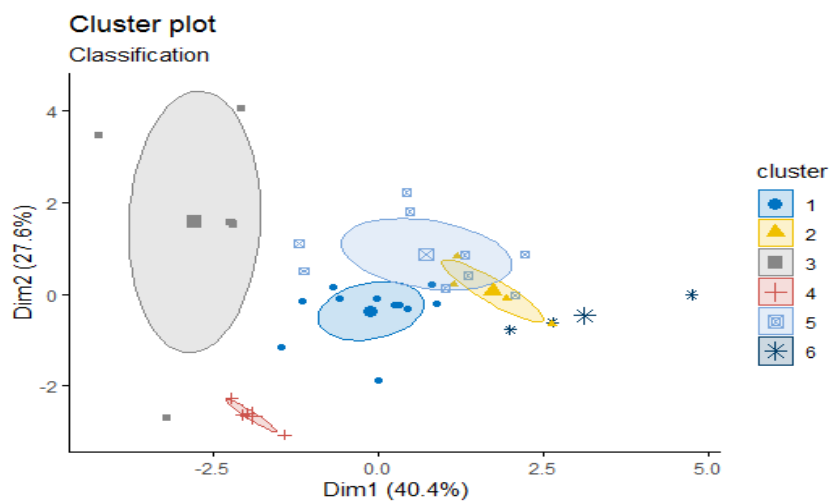
perhitungan masing-masing kriteria dengan nilai BIC dan ICL terbaik pada masing-masing model dapat dilihat pada Tabel 1. Dari Tabel 1

Tabel 1. BIC dan ICL Masing-masing Model

Model	Bayesian Information Criterion (BIC)	Jumlah Cluster	Completed Likelihood Criterion (ICL)	Jumlah Cluster
EII	-849,104	8	-851,725	8
VII	-795,065	6	-795,53	6
EEI	-810,312	8	-810,342	8
VEI	-803,994	5	-804,218	5
EVI	-775,037	4	-775,043	4
VVI	-781,276	6	-781,443	6
EEE	-782,71	9	-782,724	9
VEE	-839,531	2	-839,667	2
EVE	-731,269	2	-731,296	2
VVE	-829,456	2	-829,693	2
EEV	-755,865	2	-755,94	2
VEV	-709,3757	6	-709,376	6
EVV	-795,893	2	-795,894	2
VVV	-856,887	2	-856,891	2



Gambar 2. Pemilihan Model Terbaik Berdasarkan Nilai BIC dan ICL



Gambar 3. Visualisasi Hasil Clustering Model-Based Clustering

diketahui bahwa dari masing-masing kriteria model, nilai BIC dan ICL tertinggi sama-sama pada model VEV dengan nilai BIC dan ICL berturut-turut sebesar -709,3757 dan -709,376, dan jumlah komponen sebanyak 6 komponen. Visualisasi pemilihan model terbaik dapat dilihat pada Gambar 2. Dalam model VEV (*Variable volume, Equal shape, Variable orientation*), V berarti bahwa *cluster* memiliki volume yang bervariasi yang ditunjukkan dengan jumlah anggota *cluster* yang berbeda, E berarti setiap dimensi memiliki variansi yang kurang lebih sama, dan V berarti memiliki orientasi yang bervariasi ini artinya *cluster* tidak mengikuti garis sumbu (bisa dilihat pada Gambar 2, tidak seperti model dengan huruf ketiga I, model dengan huruf ketiga V memiliki arah yang berbeda). Visualisasi setiap kelompok bisa dilihat pada gambar 3. Model dan jumlah *cluster* yang terpilih memberikan hasil *clustering* yang lebih baik dibandingkan menggunakan metode *K-means* dengan jumlah *cluster* optimal sebanyak 3 *cluster* berdasarkan indeks *Davies-Bouldin* (DB). Hasil perbandingan tersebut dapat dilihat pada Tabel 4 berikut :

Tabel 4. Perbandingan Indeks *Davies-Bouldin* (DB) MBC dengan *K-means*

Metode	Indeks DB
<i>Model-Based Clustering</i>	1,253362
<i>K-Means</i>	1,374315

Seperti yang dapat dilihat pada Tabel 4, dengan menggunakan indeks DB, metode MBC memiliki nilai yang lebih kecil dibandingkan dengan metode *K-means*. Sehingga berdasarkan kriteria penilaian indeks DB, metode MBC memberikan hasil yang lebih baik (Charrad et al. 2010).

Estimasi model *finite mixture* seperti pada persamaan (2) yang diperoleh dari output fungsi *mclust* pada program R adalah sebagai berikut:

$$\begin{aligned}
 p(\mathbf{x}_n) = & 0,30555556 N_1(\mathbf{x}_n | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\
 & + 0,11111111 N_2(\mathbf{x}_n | \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \\
 & + 0,13888889 N_3(\mathbf{x}_n | \boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3) \\
 & + 0,11111111 N_4(\mathbf{x}_n | \boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4) \\
 & + 0,25000000 N_5(\mathbf{x}_n | \boldsymbol{\mu}_5, \boldsymbol{\Sigma}_5) \\
 & + 0,08333333 N_6(\mathbf{x}_n | \boldsymbol{\mu}_6, \boldsymbol{\Sigma}_6)
 \end{aligned}$$

Dengan,

$$N(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) \right)$$

untuk  $k = 1, 2, 3, 4, 5, 6$

Hasil estimasi untuk masing-masing komponen adalah sebagai berikut:

a. *Means* ( $\boldsymbol{\mu}_k$ )

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0,2016216 \\ 0,2869895 \\ -0,4182307 \\ 0,2075350 \\ -0,2597670 \\ -0,2265134 \\ -0,3204897 \\ -0,3771211 \\ -0,3684507 \end{bmatrix} \quad \dots \quad \boldsymbol{\mu}_6 = \begin{bmatrix} 1,6730409 \\ 0,6308741 \\ 0,1272041 \\ 0,4878447 \\ -0,8976315 \\ -0,7618192 \\ 1,6083048 \\ 1,9652379 \\ 0,8360117 \end{bmatrix}$$

b. Matriks kovarians ( $\boldsymbol{\Sigma}_k$ )

$$\begin{aligned}
 \boldsymbol{\Sigma}_1 &= \begin{bmatrix} 44844685 & 27248523 & \dots & 26557783,9 \\ 27248523 & 17454993 & \dots & 8509959,8 \\ \vdots & \vdots & \ddots & \vdots \\ 26557784 & 8509960 & \dots & 101467242,0 \end{bmatrix} \\
 \boldsymbol{\Sigma}_6 &= \begin{bmatrix} 0,026483 & 0,006483 & \dots & 0,035579 \\ 0,006483 & 0,002005 & \dots & 0,007546 \\ \vdots & \vdots & \ddots & \vdots \\ 0,035578 & 0,00755 & \dots & 0,648961 \end{bmatrix}
 \end{aligned}$$

Anggota bagi masing-masing *cluster* dipilih berdasarkan nilai peluang menggunakan persamaan (4). Sampel dengan peluang tertinggi dari keenam *cluster* akan masuk dalam *cluster* tersebut. Besaran peluang dapat dilihat pada bagian tabel  $z$  dari output fungsi *Mclust*. Peluang sampel pada masing-masing *cluster* dapat dilihat pada Tabel 2.

Sehingga dari Tabel 2 diperoleh rincian hasil pengelompokan rasio keuangan pada masing-masing saham sebagai berikut:

- Cluster 1* : ADHI, AKRA, ANTM, ASII, ELSA, EXCL, INDF, PTPP, SMGR, TLKM, dan WIKA
- Cluster 2* : ADRO, PTBA, TPIA, dan UNTR
- Cluster 3* : ASRI, BBTN, BKSL, LPKR, dan WSKT
- Cluster 4* : BBKA, BBNI, BBRI, dan BMRI

- e. *Cluster 5* : BSDE, ICBP, INDY, INKP, JSMR, MEDC, PGAS, SRIL, dan SSMS
- f. *Cluster 6* : CPIN, KLBF, dan MNCN

Setelah *cluster* terbentuk, bisa diketahui karakteristik data pada masing-masing kelompok berdasarkan nilai rata-rata masing-masing rasio, yang bisa dilihat pada Tabel 3. Dari Tabel 3 bisa diketahui karakteristik dari masing-masing kelompok yang terbentuk berdasarkan nilai rasio, dapat dijelaskan sebagai berikut:

a. *Cluster 1*

*Cluster 1* bisa dibilang memiliki rata-rata nilai rasio yang stabil dibandingkan *cluster* lainnya. Tidak ada nilai yang ekstrim, seperti nilai rata-rata tertinggi atau yang terendah. Akan tetapi, apabila dibandingkan dengan rata-rata rasio secara keseluruhan, *cluster 1* memiliki rasio profitabilitas yang lebih tinggi. Ini artinya kemampuan *cluster 1* dalam menghasilkan profit masih di atas rata-rata.

Untuk ukuran risiko perusahaan, *cluster* ini memiliki risiko yang relative rendah, mengingat nilai rasio solvabilitas *cluster* ini lebih kecil dibandingkan rata-rata rasio solvabilitas secara keseluruhan. Akan tetapi, kemampuan dalam membiayai kegiatan operasional perusahaan lebih rendah dari rata-rata rasio likuiditas secara keseluruhan.

b. *Cluster 2*

*Cluster 2* memiliki *cash ratio* yang paling tinggi dibandingkan *cluster* yang lain, sehingga *cluster* ini memiliki kemampuan yang lebih tinggi dibandingkan *cluster* lain dalam membiayai kegiatan operasional perusahaan. Akan tetapi, hal ini tidak sejalan dengan kemampuan *cluster* ini dalam menghasilkan profit, salah rasio profitabilitas *cluster* ini terendah dibandingkan *cluster* yang lain. Untuk ukuran risiko perusahaan, *cluster* ini memiliki risiko yang relative rendah dibandingkan *cluster* lain, mengingat nilai rasio solvabilitas *cluster* ini dekat dengan nilai rasio solvabilitas terendah.

c. *Cluster 3*

*Cluster 3* memiliki nilai rasio likuiditas yang lebih rendah dibandingkan rata-rata rasio likuiditas secara keseluruhan, selain itu rasio likuiditas *cluster 3* juga relatif dekat dengan rasio likuiditas terendah, sehingga *cluster* ini

memiliki kemampuan dalam membiayai kegiatan operasional perusahaan lebih rendah dari rata-rata. *Cluster 3* juga memiliki nilai rata-rata RoA, RoE, dan *net profit margin* terendah diandingkan dengan *cluster* lainnya, artinya secara garis besar kemampuan perusahaan *cluster 3* dalam menghasilkan profit untuk perusahaan lebih rendah dibandingkan dengan *cluster* lainnya. Risiko perusahaan pada *cluster 3* juga cukup besar, hal ini dikarenakan salah satu rasio solvabilitas pada *cluster* ini yaitu *debt to equity ratio* paling rendah diantara *cluster* lainnya.

d. *Cluster 4*

*Cluster 4* memiliki nilai rata-rata *cash ratio* dan *quick ratio* terendah diandingkan dengan *cluster* lainnya, sehingga *cluster* ini memiliki kemampuan yang lebih rendah dalam membiayai kegiatan operasional perusahaan. Meskipun demikian, kemampuan perusahaan pada *cluster* ini dalam menghasilkan profit sangat baik, hal ini ditunjukkan dengan rasio profitabilitas *cluster 4* yang lebih tinggi dari *cluster* lainnya yaitu *gross profit margin* dan *net profit margin*. Untuk ukuran risiko perusahaan, *cluster* ini memiliki risiko yang relatif rendah dibandingkan *cluster* lain, mengingat salah satu nilai rasio solvabilitas *cluster* ini adalah yang terendah.

e. *Cluster 5*

*Cluster 5* memiliki kemampuan dalam membiayai kegiatan operasional perusahaan yang cukup baik, hal ini ditunjukkan dengan rasio likuiditas *cluster* ini dekat dengan nilai rasio likuiditas tertinggi. Akan tetapi, hal ini tidak sejalan dengan kemampuan *cluster* ini dalam menghasilkan profit, rasio profitabilitas *cluster* ini lebih rendah dari rata-rata rasio profitabilitas secara keseluruhan dan relatif dekat dengan yang terendah. Risiko perusahaan *cluster* ini juga cukup tinggi, mengingat salah satu nilai rasio solvabilitas *cluster* ini adalah yang tertinggi.

f. *Cluster 6*

*Cluster 6* memiliki nilai rata-rata *current ratio* dan *quick ratio* tertinggi diandingkan dengan *cluster* lainnya, sehingga *cluster* ini memiliki kemampuan yang lebih tinggi dibandingkan *cluster* lain dalam membiayai kegiatan operasional perusahaan. Dalam menghasilkan profit, *cluster 6* juga memiliki

Tabel 2. Peluang pada Maing-Masing Cluster

Kode Saham	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
ADHI	1	0	0	0	6.96E-69	0
ADRO	2.57E-72	1	0	0	5.60E-130	0
AKRA	1	0	0	0	4.70E-146	0
ANTM	1	0	0	0	5.68E-188	0
ASII	1	0	0	0	2.77E-98	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
UNTR	7.40E-130	1	0	0	7.67E-200	0
WIKA	1	0	0	0	6.49E-66	0
WSKT	0	0	1	0	0	0

Tabel 3. Rata-rata Masing-Masing Rasio di Setiap Cluster

Rasio	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Rata-rata Rasio
Cash Ratio	0,3538	1,0472	0,1261	0,0912	0,7380	0,8582	0,5081
Curent Ratio	1,1919	1,8804	1,2208	0,2600	2,0005	3,4979	1,5632
Quick Ratio	0,9194	1,6314	0,4419	0,2586	1,5953	2,2276	1,1367
RoA	0,03072	0,0492	-0,0637	0,0132	0,0109	0,1123	0,0195
RoE	0,0751	0,0775	-0,3072	0,0895	0,0111	0,1484	0,0139
Gross Profit Margin	0,2606	0,2024	0,3566	0,9798	0,3629	0,4250	0,3867
Net Profit Margin	0,05161	0,0809	-0,6217	0,2529	0,0228	0,1427	-0,0159
Long-Term Debt Ratio	0,1528	0,1020	0,2232	0,0350	0,3674	0,0616	0,1899
Debt to Equity Ratio	2,1719	0,6684	6,5092	6,2590	2,4089	0,2987	2,9645

kemampuan yang sangat baik, hal ini ditunjukkan dengan RoA dan RoE tertinggi dibandingkan dengan cluster lainnya. Untuk ukuran risiko perusahaan, cluster ini memiliki risiko yang rendah dibandingkan cluster lain, mengingat *debt to equity ratio* cluster ini adalah yang terendah dan nilai *long term debt ratio* cluster ini juga mendekati nilai terendah.

Berdasarkan karakteristik masing-masing cluster, diketahui bahwa mayoritas nilai rasio pada cluster 6 memiliki nilai yang paling baik. Perusahaan pada cluster ini memiliki kemampuan yang tinggi dibandingkan perusahaan dari cluster lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban jangka pendek. Sedangkan untuk rasio solvabilitas, cluster ini memiliki nilai *debt to equity ratio* cluster yang terendah dan nilai *long term debt*

*ratio cluster* yang juga mendekati nilai terendah Hal ini berarti, perusahaan pada cluster ini memiliki risiko relatif lebih rendah dibandingkan perusahaan pada cluster yang lain. Cluster 6 merupakan cluster terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terendah.

## KESIMPULAN DAN SARAN

### Kesimpulan

Dari proses *clustering*, terbentuk 6 cluster dengan nilai nilai *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL) berturut-turut sebesar -709,3757 dan -709,376, dengan model optimal yang dipilih adalah model VEV.

Berdasarkan nilai rata-rata setiap rasio, *cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terendah. *Cluster* 6 memiliki kemampuan yang tinggi dibandingkan perusahaan *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban keuangannya jangka pendek.

### Saran

Penelitian ini menggunakan distribusi *Gaussian* dalam *Model-Based Clustering* (MBC), penelitian selanjutnya bisa digunakan distribusi lain seperti distribusi *t multivariate* untuk mengetahui pengelompokan yang lebih baik terutama untuk data dengan pencilan.

### DAFTAR PUSTAKA

- Andy, and Melly Megawati. 2019. "Analysis of Liquidity , Profitability and Solvency Ratios to Assess the Financial Performance of Companies in Cigarette Industries Listed on the Indonesia Stock Exchange." *ECo-Fin* 1 (1): 22–34. <https://doi.org/https://doi.org/10.32877/e.f.v1i1.54>.
- Andayani, Mery. 2016. "Analisis Rasio Likuiditas Dan Rasio Profitabilitas Terhadap Perubahan Laba." *Jurnal Ilmu Dan Riset Akuntansi* 5 (7): 1–19.
- Banfield, Jeffrey D., and Adrian E. Raftery. 1993. "Model-Based Clustering and Non-Gaussian Clustering." *Biometrics* 49 (3): 803–21.
- Bennett, Mark J., and Dirk L. Hugen. 2016. *Financial Analytics with R: Building a Laptop Laboratory for Data Science*. United Kingdom: Cambridge University Press.
- Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer
- Bouveyron, Charles, Gilles Celeux, T. Brenden Murphy, and Adrian E. Raftery. 2019. *Model-Based Clustering and Classification for Data Science: With Application in R*. United Kingdom: Cambridge University Press.
- Charrad, Malika, Yves Lechevallier, Mohamed Ben Ahmed, and Gilbert Saporta. 2010. "On the Number of Clusters in Block Clustering Algorithms." In *23rd International FLAIRS Conference*, 392–97. Florida, United States.
- Deisenroth, Marc Peter, A Aldo Faisal, and Cheng Soon Ong. 2020. *Mathematics For Machine Learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108679930>
- Fraley, Chris, and Adrian E Raftery. 2002. "Model-Based Clustering , Discriminant Analysis , and Density Estimation." *Journal of the American Statistical Association* 97 (458): 611–31.
- Gan, Guojun, Chaoqun Ma, and Jianhong Wu. 2007. *Data Clustering: Theory, Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability)*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics American.
- Gergely, Bence, and András Vargha. 2021. "How to Use Model-Based Cluster Analysis Efficiently in Person-Oriented Research." *Journal for Person-Oriented Research* 7 (1): 22–35. <https://doi.org/10.17505/jpor.2021.23449>.
- Husnan, Suad, Enny Pudjiastuti, 2004. *Dasar-Dasar Manajemen Keuangan*. Edisi Keempat, Yogyakarta, UPP AMP YKPN.
- Kassambara, Alboukadel. 2017. *Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis Book 1)*. (<http://www.sthda.com>): STHDA.
- Kessler, Dave. 2019. "Introducing the MBC Procedure for Model-Based Clustering." In *SAS Proceeding*, 1–21.
- McNicholas, Paul D. 2016. "Model-Based Clustering." *Journal of Classification* 373 (November): 331–73. <https://doi.org/10.1007/s0035>.
- Munawir, S, 2002. *Akuntansi Keuangan dan Manajemen*, Edisi Pertama, Penerbit BPFE, Yogyakarta.
- Qona'ah, Niswatul, Alvita Rachma Devi, and I Made Gde Meranggi Dana. 2020. "Laboratory Clustering Using K-Means, K-Medoids, and Model-Based

- Clustering.*” Indonesian Journal of Applied Statistics 3 (1): 64–77.
- Rist, Michael, and Albert J. Pizzica. 2015. *Financial Ratios For Executives*. New York: Apress Berkeley, CA.
- Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) *mclust 5: clustering, classification and density estimation using Gaussian finite mixture models* *The R Journal* 8/1, pp. 289-317
- Shalev-shwartz, Shai, and Shai Ben-david. 2014. *Understanding Machine Learning Machine: From Theory to Algorithms*. New York: Cambridge University Press.
- S.R. Nanda, B. Mahanty, M.K. Tiwari (2010) *Clustering Indian Stock Market Data for Portfolio Management. Expert Systems with Applications* 37 (2010) 8793-8798
- Subekti, R, E R Sari, and R Kusumawati. 2018. “*Ant Colony Algorithm for Clustering in Portfolio Optimization.*” In *IOP Conf. Series: Journal of Physics: Conf. Series* 983.
- Subekti, Retno, Rosita Kusumawati, and Eminugroho Ratna Sari. 2017. “*K-Means Clustering dan Average Linkage dalam Pembentukan Portfolio Saham.*” In *Seminar Matematika Dan Pendidikan Matematika UNY T-31*, 219–24.
- Weston, J. Fred. dan Eugene F. Brigham. 1993. *Manajemen Keuangan, Edisi Kesembilan. Jilid 1 dan 2. Alih Bahas: Alfonsus Sirait*. Erlangga, Jakarta.





# PENGEMBANGAN LABORATORIUM STATISTIKA BERBASIS ANDROID

Yusrianti Hanike<sup>1</sup>, Siti Nurjannah<sup>2</sup>

<sup>1,2</sup>IAIN Ambon

e-mail : <sup>1</sup>[yusriantihanike@iainambon.ac.id](mailto:yusriantihanike@iainambon.ac.id)

Received: February 14, 2023; Revised: May 10, 2023; Accepted: June 5, 2023; Published: June 30, 2023

## Abstrak

Pentingnya praktikum sebagai wadah dalam penjelasan teoritis memiliki dukungan yang kuat dalam proses transfer pengetahuan, maka perlu diejawantahkan pada proses pembelajaran dalam bentuk yang mudah diakses oleh mahasiswa. Pengembangan laboratorium berbasis statistika adalah salah satu solusi untuk memudahkan mahasiswa dalam mengakses *software* statistika yang notabene sulit dan memiliki kapasitas yang besar. Tujuan pengembangan laboratorium statistika berbasis *Android* untuk mengimplementasikan keilmuan statistika dan didukung dengan uji kepraktisan dan keefektifan aplikasi pada mahasiswa dan dosen. Metode penelitian yang dilakukan melalui lima tahap yakni penilaian, perancangan, pengembangan, implementasi, dan evaluasi. Sampel penelitian diambil dari dua program studi (*prodi*) di dua perguruan tinggi di Ambon yakni *prodi* Statistika Universitas Pattimura dan *prodi* Ekonomi Syariah Institut Agama Islam Negeri (IAIN) Ambon. Hasil penelitian menunjukkan Laboratorium Statistika berbasis *Android* memiliki persentase yang cukup signifikan terhadap kelayakan, kepraktisan dan keefektifan praktikum statistika dengan penilaian kelayakan sebesar 71%, kepraktisan sebesar 79% dan 75% menunjukkan keefektifan penggunaan aplikasi.

**Kata kunci:** Praktikum, Laboratorium Statistika, Uji Kelayakan, Uji Kepraktisan, Uji Keefektifan.

## Abstract

*The importance of practicum as a forum for theoretical explanations has strong support in the knowledge transfer process, so it needs to be embodied in the learning process in a form that is easily accessible to students. The development of a statistics-based laboratory is one of the solutions to make it easier for students to access statistical software which is difficult and has a large capacity. The aim of developing an Android-based statistics laboratory is to implement statistical science and is supported by practicality and effectiveness tests for students and lecturers. The research method is carried out through five stages, namely assessment, design, development, implementation, and evaluation. The research samples were taken from two study programs (*prodi*) at two tertiary institutions in Ambon, namely the Statistics study program at Pattimura University and the Sharia Economics study program at the State Islamic Institute (IAIN) Ambon. The results showed that the Android-based Statistics Laboratory had a significant percentage of the feasibility, practicality and effectiveness of statistical practicum with a feasibility rating of 71%, practicality of 79% and 75% indicating the effectiveness of using the application.*

**Keywords:** *Practicum, Statistics Laboratory, Feasibility Test, Practicality Test, Effectiveness Test*

## PENDAHULUAN

Pendidikan merupakan perwujudan elemen penting dalam kehidupan manusia terutama dalam menghadapi perkembangan di era teknologi yang semakin canggih. Pendidikan mampu mengantar individu dalam pengembangan potensi diri dan persaingan dalam berinteraksi di kehidupan sosial yang maju seperti saat ini (Akbarini, N. R., Murtini, W., & Rahmanto 2018). Proses pendidikan yang demikian tidak hanya diperlukan pada pematapan teori saja melainkan juga pada pematapan prakteknya.

Praktek dalam dunia pendidikan formal terutama di bidang sains bisa diaplikasikan dengan kegiatan praktikum. Melalui praktikum suatu teori lebih mampu dipahami oleh pelajar. Menurut (Makransky, Thisgaard, and Gadegaard 2016) pelajar mampu menyelesaikan dan memahami suatu masalah tentang konsep secara mendetail pada proses sains yakni praktikum. Melalui Praktek, penjelasan teoritis memiliki dukungan yang kuat dalam proses transfer pengetahuan (Dinata 2018).

Pentingnya praktek, maka perlu diejawantahkan di jenjang pendidikan yang ada. Yang perlu dominan praktiknya adalah pendidikan vokasi. Pendidikan tinggi tidak semuanya pendidikan vokasi. Namun, permasalahan yang dialami di lapangan yakni fasilitas praktikum yang terbatas. Beberapa perguruan tinggi di Kota Ambon tidak memiliki fasilitas praktikum untuk digunakan dalam proses pembelajaran. Berdasarkan observasi awal, dari lima perguruan tinggi yang dikunjungi di Kota Ambon dua diantaranya tidak memiliki laboratorium komputer. Beberapa dosen pada perguruan tinggi tersebut mengaku kewalahan untuk mengajarkan aplikasi pada pelajaran yang diberikan. Namun, hal tersebut bisa diatasi dengan menggunakan laptop milik mahasiswa. Berdasarkan penuturan salah satu dosen Matematika di Universitas Pattimura Ambon bahwa saat ini dosen cenderung menyuruh mahasiswa untuk membawa laptop dibanding

menggiring mereka ke laboratorium komputer. Apatah lagi, laptop saat ini beralih menjadi kebutuhan primer bagi mahasiswa.

Munculnya wabah baru di tahun 2019 yakni COVID-19 menjadi tantangan baru bagi dunia pendidikan (Siti et al. 2021). Masa pandemi COVID-19 ini sebagian besar memaksa perguruan tinggi dalam proses belajar dan mengajar dilaksanakan secara *online*. Pelaksanaan praktikum tidak bisa dilakukan seperti biasanya disebabkan peningkatan kasus COVID-19 yang tak kunjung reda. Namun, proses praktek laboratorium tidak bisa serta merta dihilangkan, karena praktikum adalah bagian penting yang membantu proses pembelajaran. Pada prosesnya banyak hal yang menjadi kendala, terutama di mata kuliah Statistika yang membutuhkan instalasi *software* dengan kapasitas yang besar dan rumit. Pengiriman *software* kadang menjadi penghambat pada proses pembelajaran. Wawancara dengan salah satu dosen Statistika Unpatti, (Ahmad 2022), mengakui selama pandemi COVID-19 praktek yang diajarkan tidak maksimal. Senada dengan penuturan mahasiswa IAIN Ambon, (Oktavia, n.d.) menurutnya selama ini ia hanya belajar menggunakan modul praktikum saja dan melihat penggunaannya dari penjelasan dosen saja. Namun untuk mengaplikasikannya belum pernah karena kendala *software* yang sulit didapatkan dan sangat susah diinstal.

Kondisi tersebut yang menyebabkan pentingnya media yang mampu mengatasi segala kondisi seperti permasalahan diatas yakni fasilitas yang tidak ada, praktikum yang harus *online*, dan alat praktek yang rumit. Maka dari itu, penelitian merancang fasilitas laboratorium berbasis *Android*. Fasilitas laboratorium berbasis *Android* bisa menjadi solusi untuk permasalahan di atas, apalagi penggunaan telepon seluler *Android* yang kini mendominasi kalangan mahasiswa. Beberapa penelitian telah dilakukan, (Ferly and Hendra 2021) menghasilkan aplikasi pembelajaran aksara Lampung berbasis *Android*. Aplikasi

menggunakan *software unity 3D* dengan metode *extreme programming*. (Dara and Munirul 2021) merancang aplikasi media pembelajaran Asmaul Husna berbasis *Android*. Aplikasi ini dibangun dengan menggunakan *software Android Studio* dengan bahasa pemrograman *Java*. Sementara metode yang digunakan yakni sistem model *Unified Modeling Language* (UML) dengan bahasa pemodelan secara grafis guna membangun, memvisualisasikan, menspesifikasikan, serta mendokumentasikan semua artefak sistem *software*. Berbeda dengan penelitian yang akan dilakukan, aplikasi yang akan dibuat adalah aplikasi berbasis *Android* pengganti laboratorium statistika yang memuat berbagai macam analisis statistika pada praktikum.

Pengembangan aplikasi ini memuat modul, *software* dan testing pada mahasiswa dengan memanfaatkan aplikasi playstore sehingga mudah diakses oleh mahasiswa. Aplikasi ini diujicobakan pada mahasiswa di dua perguruan tinggi di Kota Ambon yakni IAIN Ambon pada prodi Ekonomi Syariah dan Universitas Pattimura di prodi Statistika yang mengampu mata kuliah Statistika. Adapun tujuan penelitian ini yakni untuk menguji kepraktisan dan keefektifan Laboratorium Statistika berbasis *Android*. Maka dari itu, fokus penelitian ini adalah pengembangan laboratorium statistika berbasis *Android*.

## **METODOLOGI**

### **Tinjauan Referensi**

Ada beberapa penelitian terdahulu yang membahas tentang rancangan dan pembuatan aplikasi berbasis *Android* diantaranya adalah: penelitian (Dara and Munirul 2021) dengan judul penelitian rancang bangun aplikasi media pembelajaran Asmaul Husna berbasis *Android* dengan menggunakan *software Android Studio* dengan bahasa pemrograman *Java*. Aplikasi ini menyediakan visual dan pelafalan Asmaul Husna untuk memudahkan penggunaannya dalam menghafal Asmaul Husna. Aplikasi

ini diperuntukkan bagi anak-anak sehingga pada penggunaannya sangat mudah diakses dan dioperasikan. Tujuan utama aplikasi ini yakni anak-anak bisa menggunakannya untuk mempelajari, memahami dan mengenal Asmaul Husna dengan baik dan benar. Namun, pada penelitian belum dipaparkan hasil testing atau uji coba terhadap aplikasi ini. Sehingga untuk melihat seberapa besar pemanfaatan aplikasi ini tidak dapat dibuktikan (Tri, Ita, and Ines 2019).

Penelitian selanjutnya berjudul rancang dan bangun aplikasi pembelajaran aksara Lampung berbasis *Android* yang dilakukan oleh (Ferly and Hendra 2021). Penelitian ini dilakukan berdasarkan metode pengembangan sistem yaitu *Extreme Programming (XP)*. Tahapan yang dilakukan terdiri dari empat tahap yakni perencanaan, desain sistem, pengkodean, dan testing. Aplikasi aksara lampung ini dapat dijalankan dalam *smartphone Android* dengan resolusi layar 18:9. Aplikasi disertai visual suara yang memudahkan penggunaannya dalam memahami cara pengucapan huruf-huruf aksara lampung. Namun, belum ada tes sebagai pengujian aksara untuk pengguna.

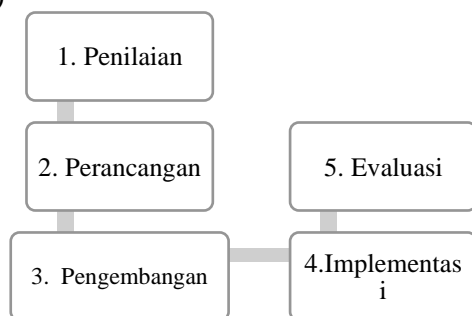
Penelitian yang memuat tema tentang perancangan aplikasi (Muchson, Betti, and Dwi 2019) yang melakukan penelitian dengan pengembangan virtual lab berbasis *Android* pada materi asam basa untuk siswa SMA. Aplikasi ini memiliki format APK yang diberi nama ABC Lab (*Acid Base Chemistry Laboratory*). Keunggulan aplikasi ini yakni adanya pretest untuk pemahaman awal siswa, aplikasi dilengkapi dengan visualisasi partikulat yang menunjang pemahaman konsep, terdapat menu praktikum yang dapat disimulasikan dan efisien. Kelemahannya produk ini hanya dapat dijalankan secara maksimal pada *smartphone* berbasis *Android* dengan versi 5.0 - 7.0 dan resolusinya tidak maksimal (Team 2017).

Selain itu, penelitian lain yang senada yakni penelitian (Saputra and Kurniawati 2021) tentang desain media pembelajaran

berbasis *Android* pada materi praktikum pengenalan alat laboratorium kimia sekolah menengah atas. Penelitian ini memuat media pembelajaran kimia berupa aplikasi *Android* dengan format file apk (*Android PacKage*) berukuran 13.46 *Megabyte*. Aplikasi ini dapat dioperasikan pada *Android* versi 2.2 dan versi di atasnya dengan tampilan terbaik pada ukuran layar minimal 4.5 inci. Penilaian validitas dan praktikalitas aplikasi ini sangat valid sebesar 96.92% dan hasil validator ahli media sebesar 85% sementara tanggapan penilaian guru kimia melalui uji praktikalitas sebesar 92%. Namun pada aplikasi masih ditampilkan dalam bentuk web yang notabene membutuhkan jaringan *online* yang stabil dalam pengaplikasiannya. Banyak penelitian yang menghasilkan produk berbasis *Android*. Namun, sebagian besar hanya berfokus pada tahapan pembuatan saja. Sementara kemanfaatan aplikasi menjadi poin penting dalam perancangan dan pembuatannya. Pada penelitian kali ini, peneliti fokus melakukan uji coba kepada pengguna untuk melihat bagaimana praktis dan efisien dari aplikasi yang digunakan.

### Metode Penelitian

Aplikasi laboratorium statistika berbasis *Android* ini diperuntukkan untuk mahasiswa perguruan tinggi yang mengampu mata kuliah Statistika. Metode yang digunakan menggunakan lima tahap. Tahapan terdiri dari tahap penilaian, perancangan, pengembangan, implementasi, dan evaluasi. Kelima tahap ini di adaptasi menurut (Lee and Owens, n.d.)



Gambar 1 Tahapan Penelitian

Tahapan pertama yakni penilaian atau analisis awal yang terdiri dua tahap yakni penilaian kebutuhan (*needs assessment*) dan analisis awal-akhir (*front-end analysis*). Pada tahap penilaian kebutuhan, dilakukan wawancara langsung dengan dosen Statistika di Perguruan Tinggi di Ambon antaranya IAIN Ambon dan Universitas Pattimura. Proses ini bertujuan untuk mengetahui dan membandingkan kurikulum dan praktikum yang selama ini dilakukan. Selanjutnya, tahapan front-end analisis berupa analisis mahasiswa yakni menganalisis kemampuan kognitif dan praktikum mahasiswa, analisis teknologi tentang meninjau *smartphone Android* yang biasa dipakai, analisis kejadian penting yakni penentuan materi sub bagian yang menjadi modul laboratorium Statistika, analisis tujuan untuk melihat pencapaian dari indikator yang ada pada praktikum, dan analisis media adalah analisis pemilihan aplikasi yang cocok untuk tampilan dari aplikasi yang digunakan.

Tahap selanjutnya atau di tahap kedua adalah perancangan, pada tahap ini merancang komposisi materi yang akan disajikan pada laboratorium statistika dan menyusun instrumen kelayakan, kepraktisan dan keefektifan. Tahap selanjutnya, yakni tahap ketiga yakni pengembangan. Penelitian ini akan melakukan pemrograman dan *storyboard*. Sementara di tahap implementasi, dilakukan penilaian kelayakan terhadap aplikasi dari sisi fungsi, kesesuaian dan isi konsep materi, dan keefektifan dan kepraktisan materi. Penilaian ini dilakukan oleh seorang pakar di bidang Statistika dan pemrograman. Selain itu, pada tahap ini dilaksanakan juga pengujian keefektifan dan kepraktisan terhadap mahasiswa. Uji Coba dilakukan dengan menguji aplikasi. Di akhir uji coba, setiap mahasiswa akan menyampaikan pendapatnya melalui kuesioner. Tahap terakhir yakni evaluasi, pada tahap ini dilakukan analisis berdasarkan hasil validasi dan uji coba pada tahap implementasi dan melakukan revisi

dari para ahli, dosen, mahasiswa dan pengguna sehingga dihasilkan aplikasi yang lebih baik.

Data primer meliputi data hasil observasi, wawancara angket/kuesioner, dan dokumentasi meliputi hasil validasi ahli pakar, dosen dan mahasiswa di Perguruan tinggi di kota Ambon. Data sekunder meliputi data yang diperoleh berbagai tulisan dari buku, jurnal, serta artikel mengenai konsep pemrograman laboratorium statistika.

Teknik dan instrumen pengumpulan data pada penelitian ini meliputi: observasi, wawancara dan dokumentasi. Pada saat pengumpulan data, penelitian ini diawali dengan observasi terkait gambaran kondisi lapangan. Teknik pengumpulan data adalah berupa instrumen penelitian. Instrumen ada yang dibagikan kepada para ahli, dosen dan kepada sampel dari populasi mahasiswa perguruan tinggi di Ambon yang mengampu mata kuliah Statistika di Tahun 2022, sampel diambil menggunakan random sampling. Lalu dilakukan dokumentasi kegiatan. Instrumen penelitian yang dilakukan yakni Instrumen kevalidan, kepraktisan dan keefektifan.

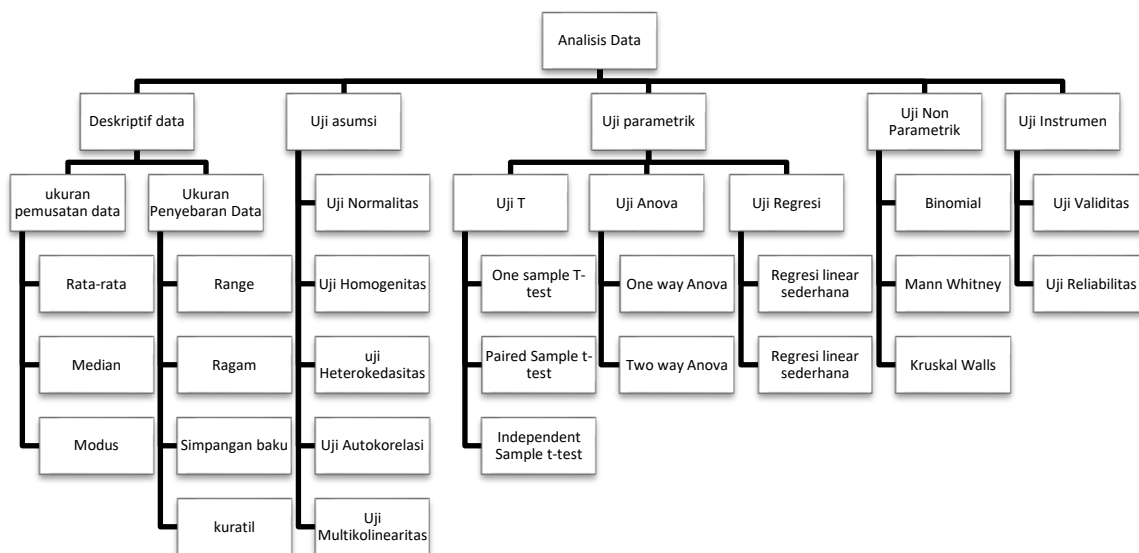
Analisis data terdiri dari : 1) Analisis Uji Validitas Laboratorium Statistika berbasis *Android* dianalisis dengan menggunakan analisis deskriptif berdasarkan indikator yang tertera pada lembar angket atau kuesioner. 2) Analisis Uji kepraktisan Laboratorium Statistika Berbasis *Android* menggunakan uji deskriptif dari hasil persentase indikator pada lembar instrument. 3) Analisis uji keefektifan Laboratorium Statistika Berbasis *Android* diperoleh dengan analisis deskriptif dari rasio yang dihasilkan pada lembar kuesioner yang diberikan pada pengguna.

## PEMBAHASAN

Kebutuhan praktikum statistika merujuk pada kurikulum yang berlaku pada setiap perguruan tinggi di Ambon. Pengujian statistika yang terdiri dari statistika deskriptif dan statistika inferensia adalah konsep awal kurikulum yang akan dibuat. Pada tahap awal penelitian yakni tahap penilaian kebutuhan yang dilakukan wawancara langsung dengan dosen Statistika di Perguruan Tinggi di Ambon antaranya IAIN Ambon dan Universitas Pattimura. Materi utama untuk kepentingan aplikasi meliputi Statistika deskriptif dan inferensi. Statistika deskriptif meliputi ukuran pemusatan data dan ukuran penyebaran data sementara statistika inferensi meliputi statistika parametrik dan nonparametrik. Tahapan analisis awal, pada analisis kejadian penting yakni penentuan materi sub bagian yang menjadi modul laboratorium Statistika. Berikut materi sub bagian modul yang dihasilkan. Analisis teknologi tentang meninjau *smartphone Android* yang biasa dipakai. Berdasarkan dua prodi IAIN Ambon 58 mahasiswa dan prodi Statistika sebanyak 63 mahasiswa bahwa sebagian besar memiliki tipe *Android* 10 sebanyak 31%. Berdasarkan hasil ini maka pada analisis media yang pada aplikasi *Android* yakni dengan format file apk (*Android PacKage*) berukuran maksimal 20 Megabyte dan Aplikasi ini dapat dioperasikan pada *Android* versi 5.0 dan versi di atasnya (Liu et al. 2016).

Table 1 tipe *Android* Telepon Seluler Mahasiswa

Tipe <i>Android</i>	Jumlah Mahasiswa	Persen
<i>Android</i> 8.0 – 8.1 (Oreo)	10	8%
<i>Android</i> 9 (Pie)	24	20%
<i>Android</i> 10 ( <i>Android</i> Q)	37	31%
<i>Android</i> 11 (Red Velvet Cake)	21	17%
<i>Android</i> 12 (Snow Cone)	19	16%
<i>Android</i> 13 (Tiramisu)	10	8%



Gambar 2 Sub materi Aplikasi Laboratorium Statistika

Analisis selanjutnya yakni menganalisis kemampuan kognitif praktikum mahasiswa sekaligus analisis tujuan yakni melihat pencapaian dari indikator yang ada pada praktikum. Berdasarkan hasil analisis pretest menunjukkan bahwa mayoritas mahasiswa memiliki kemampuan awal sebesar 39% di rentang nilai 61-80.

Table 2 Hasil Pretest kognitif Mahasiswa

Nilai	Jumlah	Persen
0-20	11	9%
21-40	19	16%
41-60	32	26%
61-80	47	39%
81-100	12	10%

Tahap selanjutnya adalah perancangan dan pengembangan. Perancangan laboratorium menggunakan aplikasi *Android studio*. Komposisi materi dan jenis analisis data pada aplikasi dijabarkan pada Gambar 2. Pada Gambar 3 disajikan tampilan halaman awal dan modul pada aplikasi. Tampilan awal memuat pilihan modul, analisis data, tabel, latihan soal dan panduan penggunaan aplikasi. Selain menampilkan modul Analisis data yang disajikan pada aplikasi

meliputi pengujian parametrik dan pengujian nonparametrik.

Pada tahapan perencanaan tingkat kelayakan laboratorium statistika diukur dari hasil segi fungsinya sebagai media praktikum berdasarkan hasil penilaian pakar yang terdiri dari dosen Statistika sebanyak 9 orang dan pengguna sebanyak 121 orang. Tahapan selanjutnya yakni pengembangan dengan melakukan pemrograman dan *storyboard*. Berikut salah satu *syntax* untuk pengujian deskriptif data:

```

override fun calculateDeskriptifData(
    analysis: String,
    edtDataName: String,
    edtDataValue: String,
    runCount: Int,
    tvSelectData: String
): DataAnalysisResult {
    val dataValue = edtDataValue.split(" ").map {
        it.toDouble() }.toDoubleArray()

    // Mean
    val mean = dataValue.average()

    // Median
    val sortedNumbers = dataValue.sorted()
    val middle = dataValue.size / 2
    val median = if (dataValue.size % 2 == 0) {
        (sortedNumbers[middle - 1] +
        sortedNumbers[middle]) / 2.0
    } else {
        sortedNumbers[middle]
    }
}

```

```

}

// Modus
val modus = dataValue.toTypedArray().groupBy
{ it }.maxByOrNull { it.value.size }?.key

// Menghitung range dari data
val range = dataValue.maxOrNull()!! -
dataValue.minOrNull()!!

// Menghitung deviation, variance/ragam, dan
standard deviation dari data
val deviation = sqrt(dataValue.sumOf { (it -
dataValue.average()).pow(2.0) } / (dataValue.size))
val variance = dataValue.sumOf { (it -
dataValue.average()).pow(2.0) } / (dataValue.size -
1)
val standardDeviation = sqrt(variance)

// Menghitung quartile 1, 2, dan 3 serta
interquartile range (IQR) dari data
val quartile1 = dataValue.sorted()[dataValue.size
/ 4]
val quartile2 = dataValue.sorted()[dataValue.size
/ 2]
val quartile3 = dataValue.sorted()[3 *
dataValue.size / 4]
val iqr = quartile3 - quartile1

return DataAnalysisResult(
    id = runCount,
    resultTitle = "Run #\$runCount -
\$tvSelectData",
    resultData = "Data yang digunakan:
\$edtDataName",
    descriptiveTitle = "\$edtDataName
descriptive:",
    descriptiveContent = "n:

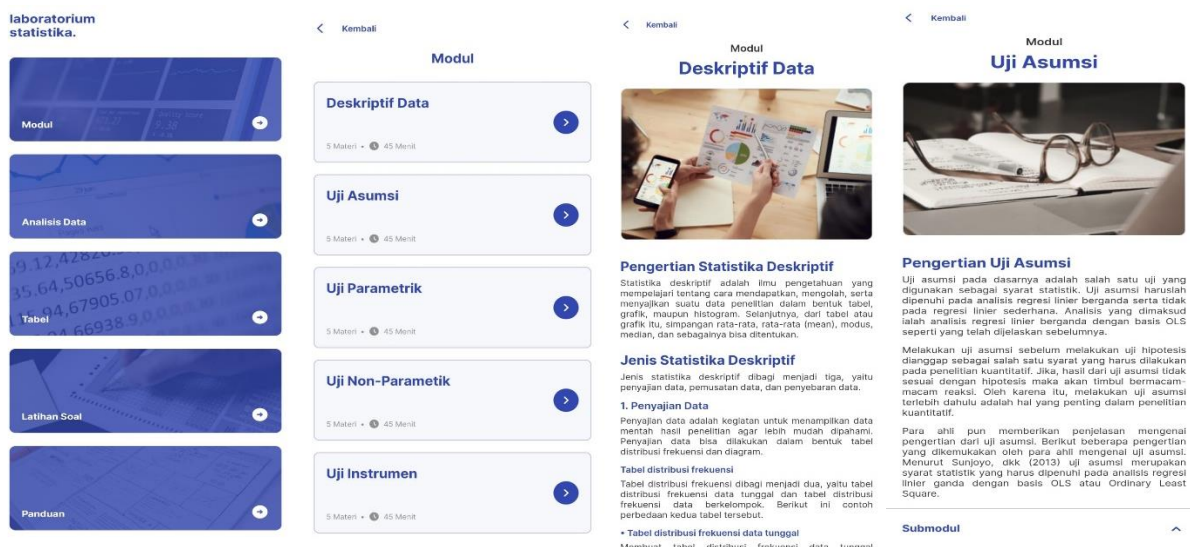
```

```

\${dataValue.size};\n" + when (analysis) {
    "Rata-rata" -> "Mean: \$mean"
    "Median" -> "Median: \$median"
    "Modus" -> "Modus: \$modus"
    "Range" -> "Range: \$range"
    "Ragam" -> "Ragam: \$variance"
    "Simpangan Baku" -> "Simpangan Baku
Populasi (Deviation): \$deviation\nSimpangan Baku
Sample (Standard Deviation): \$standardDeviation"
    "Kuartil" -> "Kuartil Bawah (Q1)\t\t\t\t\t:
\$quartile1\n" +
        "Kuartil Tengah (Q2)\t\t\t\t\t:
\$quartile2\n" +
        "Kuartil Atas (Q3)\t\t\t\t\t:
\$quartile3\n" +
        "Interkuartil Range (IQR): \$iqr"
    else -> ""
},
amountOfData = 1,
hideTestValues = true
)
}

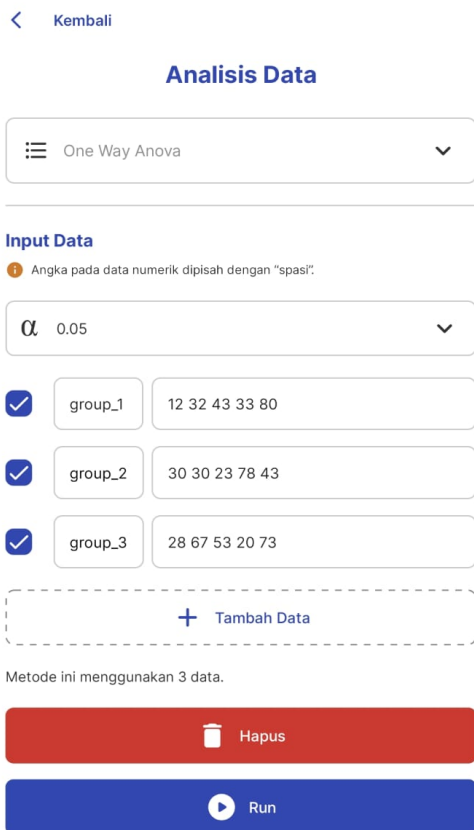
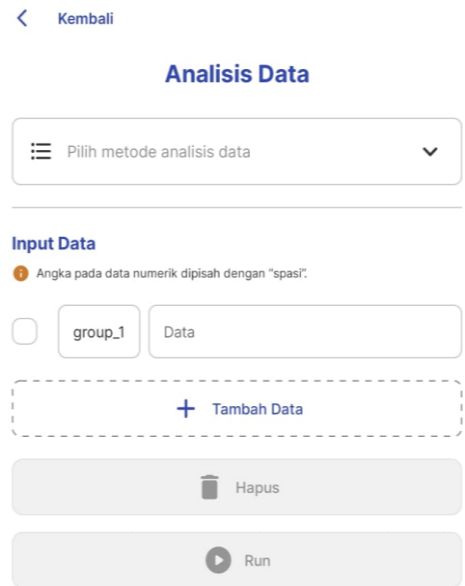
```

Aplikasi memuat jenis-jenis analisis statistika, baik parametrik maupun non parametrik. Pengujian analisis data ditampilkan pada Gambar 4 memuat penginputan data dan jenis analisis data. Pada tahap selanjutnya yakni tahap implementasi. Pada tahap ini pakar maupun pengguna diberikan aplikasi lalu melakukan uji coba terhadap analisis pengujian didalamnya beserta mengisi angket pada instrumen kelayakan, keefektifan dan kepraktisan aplikasi.



Gambar 3 Tampilan halaman awal dan modul laboratorium berbasis Android

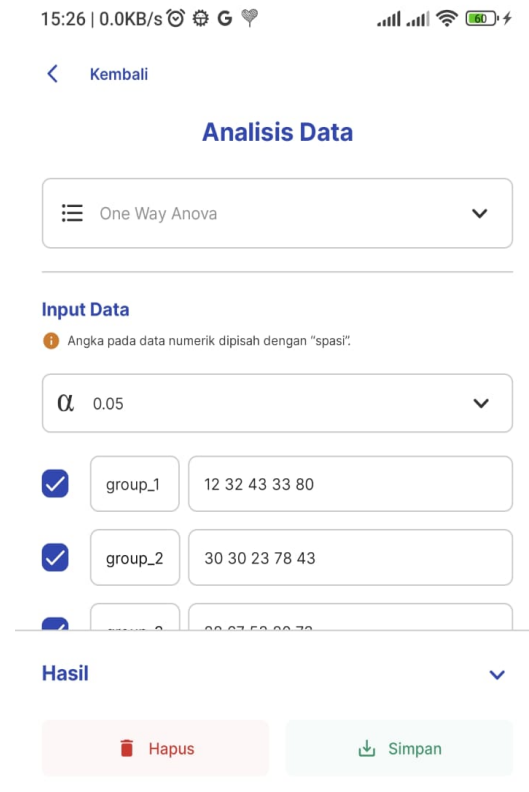




Gambar 4 Tampilan Analisis Data

Selain mengisi angket, pakar maupun pengguna menyampaikan pendapat melalui kuesioner yang dibagikan. Tahap terakhir yakni evaluasi, pada tahap ini dilakukan analisis berdasarkan hasil validasi dan uji coba pada tahap implementasi. Indikator instrumen kelayakan merujuk pada

Hasil analisis akan menampilkan pengujian deskriptif pada setiap variabel berupa jumlah data (N), rata-rata (mean), simpangan baku (SD) dan hasil analisis dari pengujian inferensi. Hasil output disajikan pada Gambar 5



**Run #2 - One Way Anova**  
 Data yang digunakan: group\_1, group\_2, group\_3

**group\_1 descriptive:**  
 n: 5; Mean: 40.0; SD: 25.029982021567655

**group\_2 descriptive:**  
 n: 5; Mean: 40.8; SD: 22.01590334281108

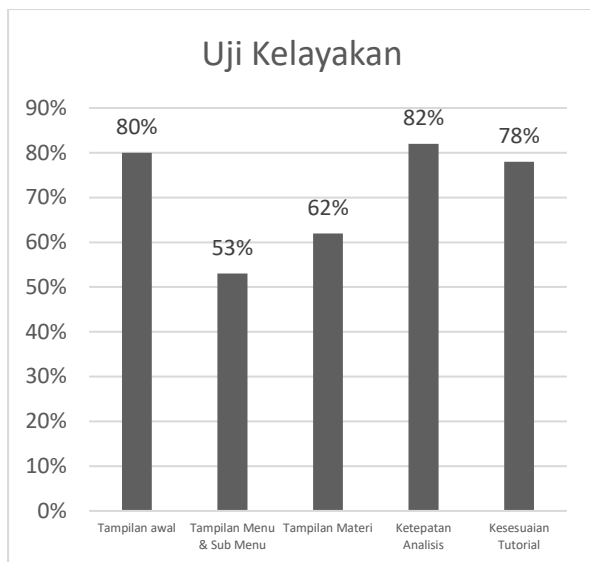
**group\_3 descriptive:**  
 n: 5; Mean: 48.2; SD: 23.424346308915432

**Test Values**  
 F-Hitung: 0.44330381348274034; F-Tabel: 3.8852938346526167

Tidak terdapat perbedaan signifikan antara rata-rata produktivitas di tiga divisi.

Gambar 5 Hasil Output Analisis

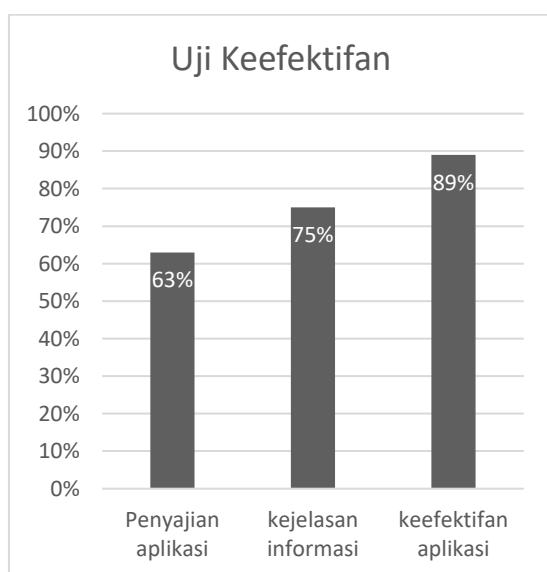
Muchson et, al (2021) meliputi tampilan awal, tampilan menu utama, tampilan sub menu utama, tampilan materi modul, ketepatan analisis, dan kesesuaian tutorial penggunaan.



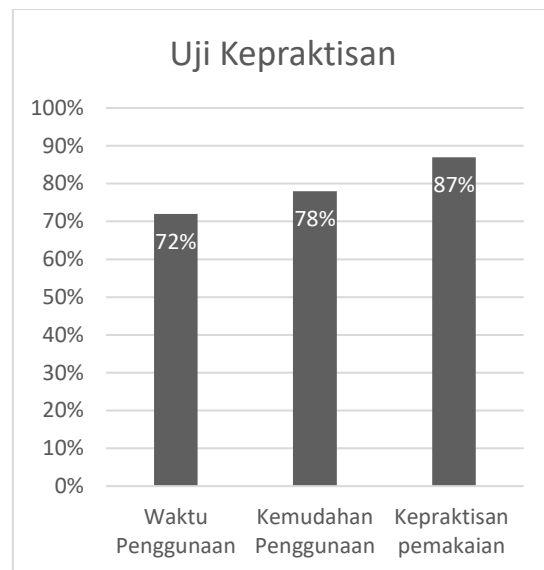
Gambar 6 Uji Kelayakan

Hasil analisis terhadap angket validasi kelayakan aplikasi dari pakar dan pengguna menunjukkan bahwa tampilan menu dan sub menu mendapatkan skor sebesar 53% berbeda dengan ketepatan analisis dengan skor 82%. Berdasarkan kelima indikator tingkat kelayakan menunjukkan rata-rata skor sebesar 71%.

Instrumen keefektifan aplikasi meliputi penyajian aplikasi, kejelasan informasi, dan keefektifan aplikasi. Hasil analisis menunjukkan penyajian aplikasi mendapatkan skor 63%, kejelasan informasi sebesar 75%, dan keefektifan aplikasi sebesar 89%. Rata-rata skor uji keefektifan yakni 75%.



Gambar 7 Uji Keefektifan



Gambar 8 Uji Kepraktisan

Indikator kepraktisan aplikasi merujuk pada Menrisal dan Harmi (2018) aplikasi waktu penggunaan, kemudahan penggunaan, dan kepraktisan pemakaian program. Hasil pengujian menunjukkan skor waktu penggunaan sebesar 72%, kemudahan penggunaan sebesar 78%, dan kepraktisan pemakaian program sebesar 87%. Rata-rata skor tingkat kepraktisan sebesar 79%. Namun demikian, kelayakan, kepraktisan maupun keefektifan aplikasi masih membutuhkan penelitian lebih lanjut. Terutama pada waktu penggunaan maupun penyajian aplikasi beserta tampilan aplikasi yang dinilai masih memiliki skor rendah pada laboratorium statistika ini.

## SIMPULAN

Penelitian menghasilkan produk berupa laboratorium statistika berbasis *Android* yang memudahkan mahasiswa maupun dosen pada mata kuliah statistika. Hasil penelitian menunjukkan Laboratorium Statistika berbasis *Android* memiliki persentase yang cukup signifikan. Pada uji kelayakan aplikasi menghasilkan skor rata-rata sebesar 71%, sementara pada uji kepraktisan memiliki skor sebesar 79%. Skor untuk pengujian keefektifan penggunaan aplikasi sebesar 75%. Hal ini menunjukkan bahwa aplikasi laboratorium statistika berbasis *Android* ini dapat digunakan dan diimplementasikan namun

masih perlu penelitian lebih lanjut untuk menyempurnakan aplikasi. Adapun beberapa saran yang perlu dikaji demi mengembangkan aplikasi ini yaitu Desain aplikasi dapat dikembangkan lagi dengan design yang lebih menarik dan Kapasitas aplikasi disederhanakan kembali agar waktu penggunaannya bisa lebih cepat bagi pengguna

#### DAFTAR PUSTAKA

- Ahmad. 2022. Hasil wawancara dengan dosen Universitas Pattimura.
- Akbarini, N. R., Murtini, W., & Rahmanto, A. N. 2018. "Design of Interactive Learning Multimedia Development in General Administration Subject" 5 (4): 138–148. <https://doi.org/https://doi.org/10.18415/ijmmu.v5i4.218>.
- Dara, F, and U Munirul. 2021. "Rancang Bangun Aplikasi Media Pembelajaran Asmaul Husna Berbasis *Android*."a
- Dinata, R. K. 2018. "Aplikasi Tutorial Resep Masakan Tradisional Aceh Berbasis *Android* Menggunakan Metode Analytical Hierarchy Process (AHP)." *Jurnal Informatika Sunan Kalijaga* 3 (1): 24–33.
- Ferly, A., and S.. Hendra. 2021. "Rancang Bangun Aplikasi Pembelajaran Aksara Lampung Berbasis *Android*." *Jurnal Informasi Dan Komputer* 9 (2): 135–43.
- Lee, W.W., and D.L Owens. n.d. *Multimedia-Based Instructional Design*. San Fransisco: Pfeiffer.
- Liu, SK, L Cao, ZY Shi, and C Feng. 2016. "The Hardware Decoding Implementation of Video Surveillance Based on *Android* System."
- Makransky, G, M. W Thisgaard, and H Gadegaard. 2016. "Virtual Simulations as Preparation for Lab Exercises: Assessing Learning of Key Laboratory Skills in Microbiology and Improvement of Essential Non-Cognitive Skills." *PLoS ONE* 11 (6). <https://doi.org/https://doi.org/10.1371/journal.pone.0155895>.
- Muchson, Munzil, E. W. Betti, and A. Dwi. 2019. "Pengembangan Virtual Lab Berbasis *Android* Pada Materi Asam Basa Untuk Siswa Sma." *Jurnal Pembelajaran Kimia* 4 (1): 51–64.
- Oktavia, Nurul. n.d. Hasil wawancara dengan mahasiswa Jurusan Ekonomi Syariah IAIN Ambon.
- Saputra, Windi Dwi, and Yenni Kurniawati. 2021. "Desain Media Pembelajaran Berbasis *Android* Pada Materi Praktikum Pengenalan Alat Laboratorium Kimia Sekolah Menengah Atas" 4 (2): 268–76.
- Siti, Ode, Nur Alam, Achmad Nur Aliansyah, Feliks Eldad Larobu, Nita Z D L Mulyawati, and Indrayati Galugu. 2021. "Pemanfaatan Smart Thermal Camera Berbasis Arduino Sebagai Upaya Pencegahan Penyebaran COVID-19 Di UPT Teknologi Informasi Dan Komunikasi Universitas Halu Oleo Dan SMK Satria Kendari The Utilization of Smart Thermal Camera Based on Arduino for Preventing COVID-19 Spread at UPT Information and Communication Technology Haluoleo University and SMK Satria Kendari" 5 (4): 634–40.
- Team, Google Developer Training. 2017. "Android Developer Fundamentals Course Concepts." 2017. <https://googledeveloper-training.github.io/Android-developer-fundamentals-courseconcepts/en/Android-developer-fundamentalscourseconcepts-en.pdf>.
- Tri, H., H. Ita, and H. I. Ines. 2019. *Statistika Dasar Panduan Bagi Dosen Dan Mahasiswa*. Purwokerto: Pena Persada.

# GENERATING SYNTHETIC TRAINING DATASETS USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK TO IMPROVE IMAGES SEGMENTATION

Iffati Uzma<sup>1</sup>, Rani Nooraeni<sup>2</sup>, Takdir<sup>3</sup>

<sup>1</sup>Badan Pusat Statistik Kabupaten Bungo, Jambi

<sup>2</sup>Politeknik Statistika STIS

<sup>3</sup>University of Tsukuba

e-mail: [1iffati.uzma@bps.go.id](mailto:iffati.uzma@bps.go.id), [2raninoor@stis.ac.id](mailto:raninoor@stis.ac.id), [3takdir-kde@kde.cs.tsukuba.ac.jp](mailto:takdir-kde@kde.cs.tsukuba.ac.jp)

*Received: April 10, 2023; Revised: June 22, 2023; Accepted: June 27, 2023; Published: June 30, 2023*

## ABSTRACT

A limited amount of training datasets in *deep learning* research could impact the accuracy of the resulting models. This situation can cause *overfit*, so the model cannot work correctly. Conditional Generative Adversarial Network (CGAN) was introduced to generate synthetic data by considering certain conditions. This study aims to generate additional synthetic training datasets to improve the accuracy of the object segmentation model of images. Firstly, we evaluated CGAN-based dataset generator accuracy against several open datasets. Then, we applied the generator to train two object segmentation models, i.e., FCN and CNN U-Net. Our evaluation shows that CGAN can generate synthetic datasets well. Complex datasets require more training iterations. It also improves both segmentation models' validation loss and accuracy, although other metrics still need further improvement.

**Keywords**— *deep learning*, synthetic data, cGAN, U-net, segmentation

## ABSTRAK

Dataset pelatihan dalam jumlah terbatas dalam penelitian pembelajaran mendalam dapat memengaruhi keakuratan model yang dihasilkan. Situasi ini dapat menyebabkan *overfit*, sehingga model tidak dapat bekerja dengan baik. Conditional Generative Adversarial Network (CGAN) diperkenalkan untuk menghasilkan data sintetik dengan mempertimbangkan kondisi tertentu. Penelitian ini bertujuan untuk menghasilkan dataset training sintetik tambahan untuk meningkatkan akurasi model segmentasi objek citra. Pertama, kami mengevaluasi akurasi generator dataset berbasis CGAN terhadap beberapa dataset terbuka. Kemudian, kami menerapkan generator untuk melatih dua model segmentasi objek, yaitu FCN dan CNN U-Net. Evaluasi kami menunjukkan bahwa CGAN dapat menghasilkan dataset sintetik dengan baik. Kumpulan data yang kompleks memerlukan iterasi pelatihan yang lebih banyak. Ini juga meningkatkan kehilangan dan akurasi validasi kedua model segmentasi, meskipun metrik lainnya masih memerlukan peningkatan lebih lanjut.

**Kata kunci**— *deep learning*, data sintetik, cGAN, U-net, segmentasi

## INTRODUCTION

*Deep Learning* Deep is a machine learning concept gaining popularity nowadays. It can automatically extract required features, an essential step in preparation for model training. However, deep learning requires many training datasets to perform well [1][2][3].

The limitation of data is an essential issue in deep learning. Thousands to millions of labeled datasets for each category shall be required to train a model [4]. In fact, we are often faced with the problem of a need for more data [5]. In this situation, the resulting model tends to overfit, i.e., it cannot be generalized for the testing datasets and leads to an inaccurate model [5][6].

A common solution to improve the accuracy and reduce the biases of a deep learning model is to add more datasets [6][7]. Generating new synthetic datasets is one of many available solutions to enlarge dataset size [4]. According to [5], an automatic data generator is used to add more training datasets to improve the performance of the corresponding classification model.

Image segmentation is a fundamental method in object-based image analysis [8]. It aims to recognize objects' existence or location by labeling each pixel in images into a class or category [9]. *Convolutional Neural Network (CNN)* is commonly used to segregate different objects semantically [10]. However, CNN often fails to generalize accurate segmentation when processing satellite images from other areas or weather conditions [11]. In practice, *Fully Convolutional Networks (FCN)* and *U-Net* architectures are the most used CNN architectures to perform semantic segmentations [9].

*Generative Adversarial Network (GAN)* is one of the deep learning methods to generate synthetic data. GAN learns the distributions of the given data or images and creates new ones that have equivalent patterns and characteristics to the existing datasets [4]

This study aims to evaluate CGAN accuracy in generating synthetic datasets. We also assess the accuracy of using CGAN-generated datasets to train CNN-based image segmentation models.

## OBJECTIVE

This study was conducted with the following objectives:

1. To apply and evaluate CGAN to generate synthetic data from publicly available image datasets.
2. To evaluate the accuracy of image segmentation models trained with additional CGAN-generated synthetic datasets.

## RELATED WORK

A sufficient amount of datasets is a common problem in deep learning research. Many studies have been done to overcome this issue. The study in [1] published a literature review to evaluate methods in data acquisition, labeling, and increasing data availability, and each was provided with approaches. One of the mentioned methods to improve data availability is automatic synthetic data generation using GAN.

Many studies related to data generation using GAN also have been conducted. In 2018, Yu Ping et al. [14] reported that CGAN reduces data-generating costs and increases robustness compared to other methods. Heilemann et al. [15] have also conducted a study using CGAN and U-Net related to data unavailability. Their study found that increasing the number of training datasets improves object segmentation accuracy.

The studies above indicate that CGAN is a potential candidate for the object segmentation model. Rezaei et al. [11] applied CGAN for the semantic segmentation of brain tumours. In 2018, Frangi et al. also conducted another medical imaging using CGAN by segmenting mammogram images. The result showed that CGAN significantly improved segmentation accuracy.

## METHODOLOGY

### A. Data Collection

We use open image datasets, i.e., MNIST-fashion, MNIST-digit, CIFAR-10, and Oxford-IIIT Pet, in this study. The first three datasets are available in the

Keras TensorFlow library. MNIST-fashion contains cloth pictures of 10 categories: shirts, trousers, sweaters, dresses, coats, sandals, suits, shoes, bags, and ankle boots. Each category contains 6000 pictures. MNIST-digit is a collection of digit images ('0' to '9' characters). CIFAR-10 includes ten categories of pictures: airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each category contains 6000 pictures.

Figures 1, 2, and 3 are sample images from each category.

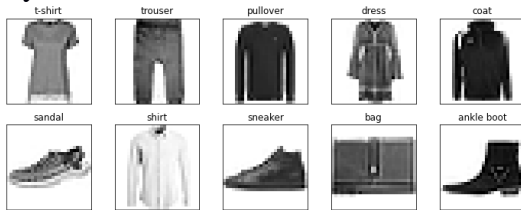


Figure 1. MNIST-Fashion sample images

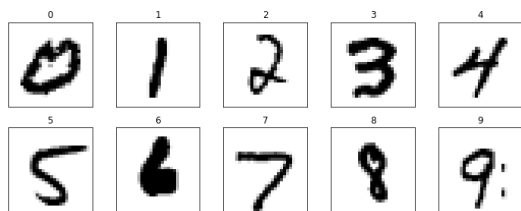


Figure 2. MNIST-Digit sample images

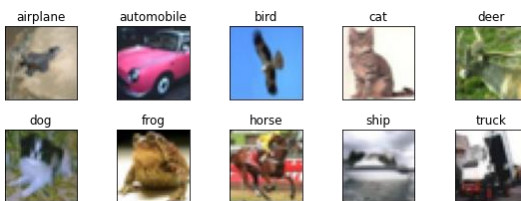


Figure 3. CIFAR-10 sample images

Oxford-IIIT Pet dataset is fetched from Kaggle. It contains 37 pet categories, each with 200 pictures with various scales, layouts, and lighting. We choose 10 out of 37 categories, as shown in Figure 4. Besides synthetic data generation, the Oxford-IIIT Pet dataset was also used to evaluate the segmentation models.



Figure 4. Oxford III-T Pet sample images

### B. CGAN Model

GAN is formed from *generator* and *discriminator* architectures. The generator identifies the distribution of data, whereas the discriminator estimates the possibility of whether a sample originated from real datasets or generators. Conditional GAN (CGAN) is an extended version with an additional feature  $y$ , e.g., data label, in both generator and discriminator. We implemented it by adding  $y$  to the generator and discriminator as a new input layer. Figure 5 depicts the CGAN model training process.

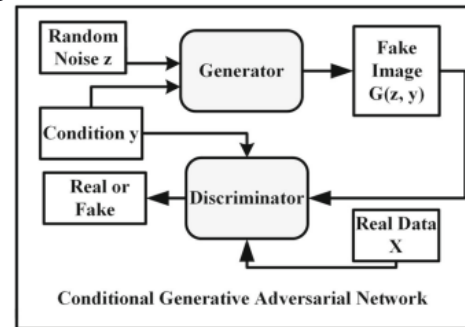


Figure 5. Conditional GAN [17]

The *loss function* of CGAN is formulated as:

$$L_{CGAN} = \min_G \max_D V_{CGAN}(D, G) = E_{x \sim p_{data}(x)} [\log D(y)] + E_{z \sim p_z(z)} \left[ \log \left( 1 - D(G(y)) \right) \right] \dots(1)$$

$V(D, G)$ : is the value function of CGAN

$y$ : condition

$D$ : discriminator

$G$ : generator

$x$ : original data item

$p_{data}(x)$ : data distribution of  $x$

$z$ : random noise vector as the inputs to generator

$p_z(z)$ : prior noise distribution

$D(y)$  means discriminator takes two inputs  $x$  and  $y$  to be used with discriminator function.  $D$  maximizes the accuracy of the classification of samples in detecting whether they are original or generated.  $G(y)$  means the generator produces images based on a condition  $y$ .  $G$  only influence  $1 - D(G(z|y))$ , which is the probability that  $D$  recognizes generated sample, by minimizing its value [4][13][17].



A CGAN model is constructed as follows [18]:

- Load dataset
- Define generator and discriminator functions
  - Generator:
    - Input: *point* in latent space
    - Output: images
    - Uses inverse of convolution (transposed convolution)
    - Activation function: LeakyReLU
    - Optimizer: Adam Optimizer
    - Activation function in the output layer: tanh
  - Discriminator
    - Input: images dan class labels
    - Output: binary (Original=1, Generated=0)
    - Implemented using Convolutional Neural Network
    - Activation function: LeakyReLU
    - Optimizer: Adam optimizer
- Define CGAN model
  - Combination of the generator and discriminator
  - Input: point in *latent space*, uses the generator to produce images. The resulting images are the inputs of the discriminator.
  - Output: a classification, i.e., original or generated
- Define a function to select random samples from the dataset for each update to the CGAN model
- Define input function for the generator from *latent space*

- Perform CGAN training process with predefined parameters.

#### Generating synthetic datasets using CGAN

CGAN model was applied to the MNIST-fashion, MNIST-digit, and CIFAR-10 datasets to evaluate its ability and accuracy in generating synthetic datasets before utilizing it to train object segmentation. This experiment also can show the characteristics of results among different datasets.

The results are evaluated using plot graphs of training loss and training accuracy. CGAN has two loss functions, each in generator and discriminator, respectively. There are two loss values in the discriminator, i.e., loss in detecting that the evaluated data is *original* and loss in detecting that the evaluated data is *generated*.

#### Image Segmentation

We performed image segmentation on the Oxford IIIT Pet dataset. Before segmentation, additional training data is generated using CGAN. We use FCN and CNN U-Net as segmentation models and compare the results with and without synthetic training data on both segmentation models.

The segmentation models are tested and evaluated against the ground truth by collecting the metrics of accuracy, training loss, validation loss, Intersection over Union (IoU), and Dice Score.

## RESULT AND DISCUSSION

### A. Visual Evaluation of Synthetic Data Generation

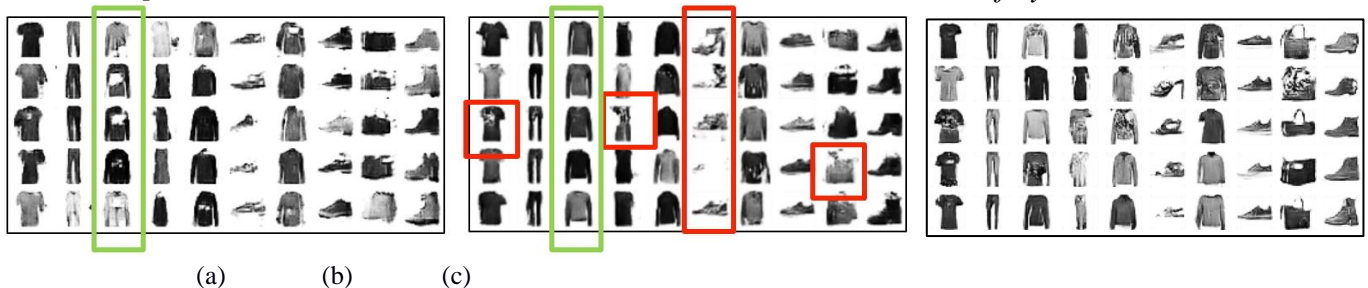


Figure 6. Sample of generated synthetic MNIST-Fashion data using (a) epoch=10, (b) epoch=20, dan (c) epoch=40



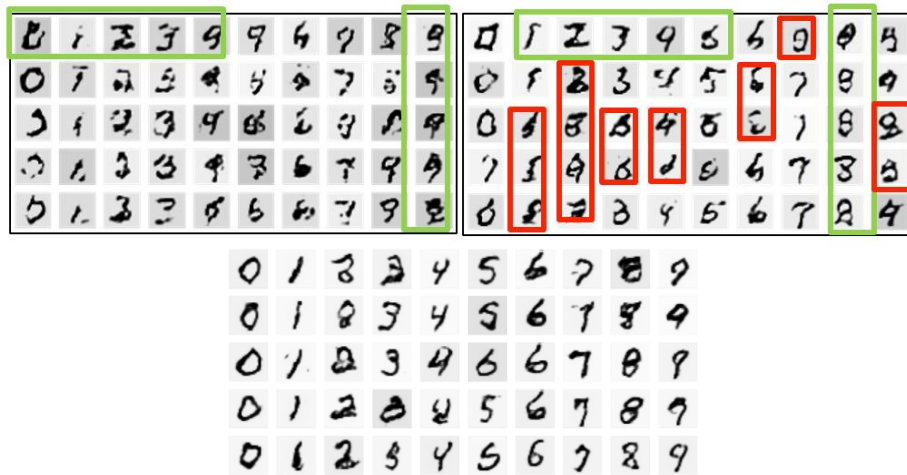


Figure 7. Sample of generated synthetic MNIST-Digit data using (a) epoch=10, (b) epoch=20, dan (c) epoch=40

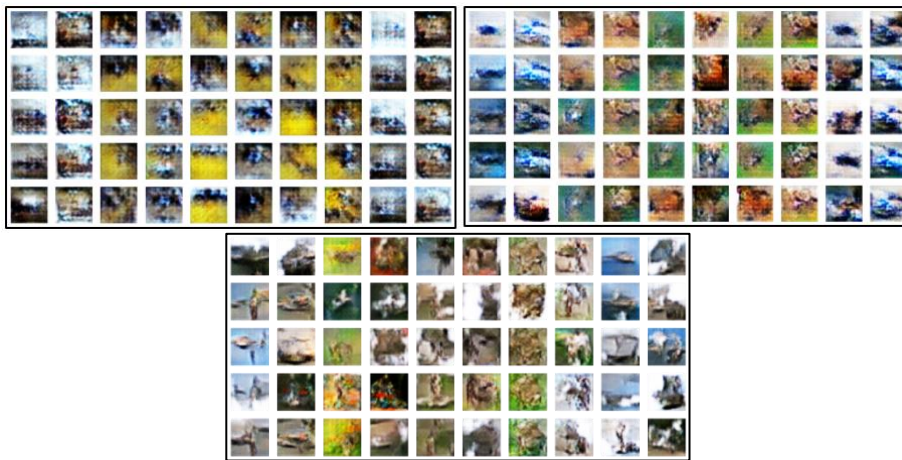


Figure 8. Sample of generated synthetic CIFAR-10 data using (a) epoch=10, (b) epoch=20, dan (c) epoch=40



Figure 9. Sample of generated synthetic Oxford IIIT Pet data using (a) epoch=10, (b) epoch=20, dan (c) epoch=40

We used 6000 images as training data from MNIST-Fashion and MNIST-Digit, respectively. There are 5000 images from

CIFAR-10 dataset. In the Oxford III dataset, there are 10 categories, and we took 200 images from each category.

MNIST-Fashion generated data, as visually shown in Figure 6, has a good similarity with the original. Increasing the epoch from 10 to 20 improves the quality of generated images, especially in the green rectangle-marked images. These images are transformed from images containing incomplete parts to complete images. However, incomplete prints, such as in the red rectangle, remain found. It means that the model can generate synthetic images properly with the configured parameters. Still, more training iterations are needed to improve the quality, as shown in the results of 40 epochs.

A similar process on MNIST-Digit datasets showed equivalent results, as depicted in Figure 7. In the ten epochs, '8' characters offer the best quality. Meanwhile, most of the '1' and '2' characters cannot be seen clearly. Increasing the epoch to 40 improves the whole image quality.

In the CIFAR-10 dataset, generated images in 10 epochs are still blurry, and their class cannot be recognized visually, as shown in Figure 8. Objects can be recognized well if the epoch is increased to 40. The CIFAR-10 dataset contains color images instead of black and white in the former datasets. Therefore, features in this kind of image are more variable and require more training iterations as well.

The lowest quality of generated images is shown in the Oxford IIIT Pet dataset. This dataset has the highest resolution and various scales and lighting. Training the generator until 40 epochs is insufficient to produce good images, as shown in Figure 9.

Generally, increasing the number of training iterations improves generated image quality, as demonstrated by the above experimental results. More iterations are required for richer image features, such as color, resolution, and brightness. Considering this, we increased the number of iterations to 100 for CIFAR-10 and Oxford IIIT Pet datasets. The sample results are presented in Figures 10 and 11, respectively.

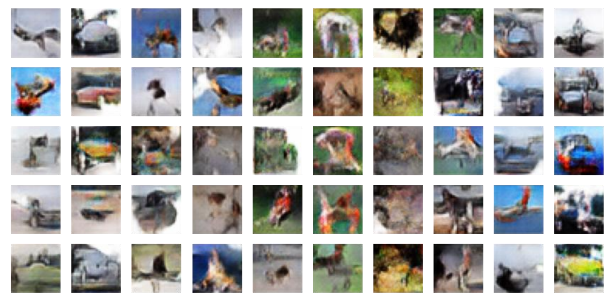


Figure 10. Sample of generated synthetic CIFAR-10 data using epoch=100



Figure 11. Sample of generated synthetic Oxford IIIT Pet data using epoch=100

The resulting images become clearer visually. Both datasets have various colors and backgrounds, meaning more features that need to be accommodated. This led to more iterations, i.e., longer training times, to achieve representative outputs.

### B. Evaluation Metrics of Synthetic Data Generation

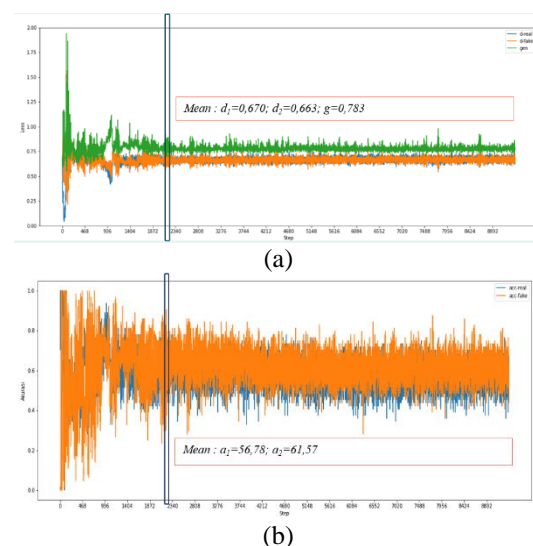


Figure 12. Evaluation metrics of CGAN on MNIST-Fashion dataset: (a) training loss, and (b) discriminator accuracy

When applying CGAN on the MNIST-Fashion dataset, the generator and discriminator were stable at iteration 2000

(5<sup>th</sup> epoch), as shown in the line graph in figure 12 (a). Loss values of the original data discriminator, generated data discriminator, and generator at that point are 0.670, 0.663, 0.783, consecutively. Figure 12 (b) depicts the accuracy of the discriminator in identifying whether data is original or generated. In line with the loss values, the accuracy of either original or generated data identifications becomes stable at iteration 2340 (5<sup>th</sup> epoch). In this stable condition, the accuracy of the original and that of generated data identifications are 56.78% and 61.56% on average, respectively.

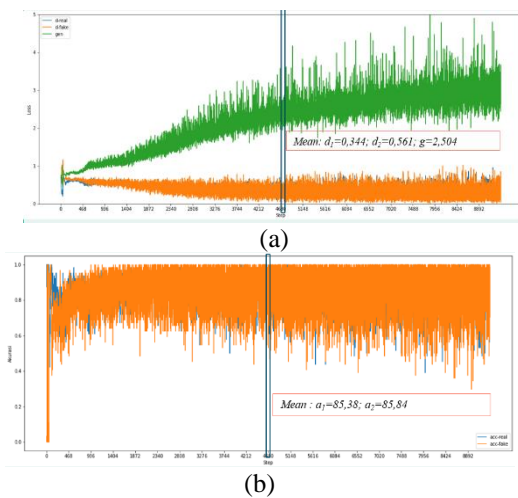


Figure 13. Evaluation metrics of CGAN on MNIST-Digit: (a) training loss, and (b) discriminator accuracy

In MNIST-Digit dataset generation, as shown in figure 13, loss values of the generator are gradually increased. The averages of loss values for the original data discriminator, generated data discriminator, and generator in the last ten epochs are 0.344, 0.561, and 2.504, respectively. The discriminator accuracy graph, i.e., figure 13 (b), shows great averages of both original and generated data which are 85.37% and 85.83% in the last ten epochs. It means that the discriminator model can classify original and generated data well.

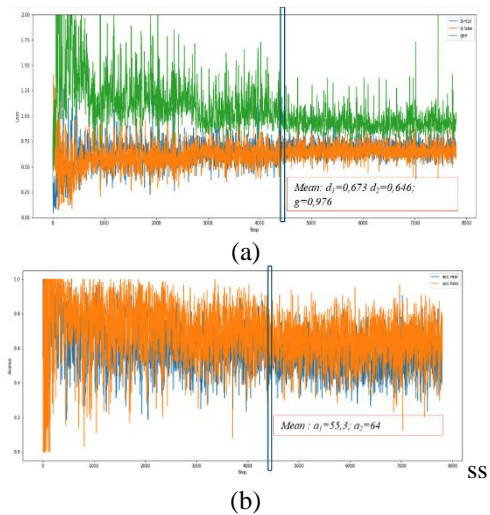


Figure 14. Evaluation metrics of CGAN on CIFAR-10: training loss, and (b) discriminator accuracy

Both graphs in Figure 14 show fluctuations in all evaluation metrics of CIFAR-10 dataset compared to the former datasets.

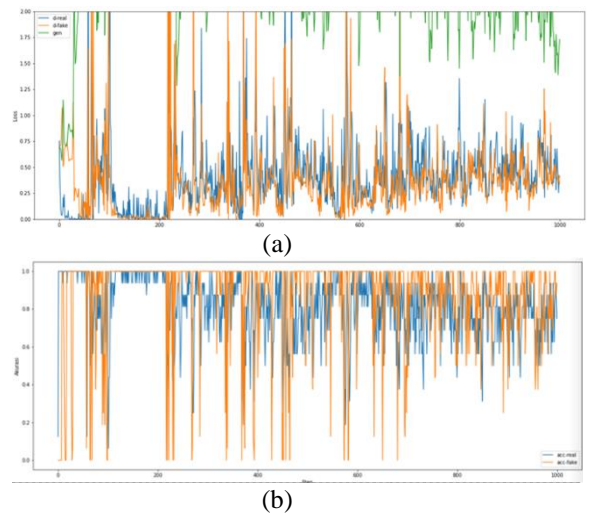


Figure 15. Evaluation metrics of CGAN on Oxford IIIT Pet dataset: (a) training loss, and (b) discriminator accuracy

Oxford IIIT Pet metrics in Figure 15 show unstable values. In the beginning, the loss values of the generator were high and decreased gradually. It means that more training iterations are required to stabilize the metrics.

There is a general pattern from all the above experimental results of loss and accuracy from four different datasets, i.e., the discriminator and generator are opposite. Their loss values would lead to a stable state if iterations were increased. An



optimal GAN model can be achieved when the loss values are stable and convergent, i.e., the model cannot increase or decrease its loss value.

TABEL I  
SUMMARY OF CGAN EVALUATION METRICS

Indikator	Epoch	MNIS	MNIS	CIFA	Oxford IIIIT
		T-fashion n	T-digit	R-10	Pet
Discriminator Loss (Original)	10	0,670	0,368	0,636	0,489
	20	0,672	0,341	0,685	0,488
	40	0,681	0,319	0,70	0,526
Discriminator Loss (Generated)	10	0,665	0,365	0,608	0,379
	20	0,666	0,339	0,663	0,431
	40	0,675	0,302	0,695	0,420
Generator Loss	10	0,779	2,087	1,125	3,091
	20	0,782	2,669	0,917	1,827
	40	0,770	3,22	0,757	1,969
Accuracy (Original)	10	57,56	85,69	59,43	77,52
	20	56,53	84,87	52,87	73,5
	40	54,18	84,7	47,04	70,82
Accuracy (Generated)	10	61,21	87,04	69,64	86,92
	20	60,28	85,56	60,95	89,12
	40	57,70	87,02	52,88	83,04

Table 1 above shows that from all data, the values in the MNIST-fashion data tend to be the most stable. MNIST-digit data is different from other data. In the MNIST digit, the value of the loss discriminator tends to decrease, while the value of the loss generator tends to increase with a fairly high value. However, for the value of discriminator accuracy, it can be seen that the three data seem to have succeeded in reducing the value of discriminator accuracy along with the ongoing training process, meaning that the generator carries out a learning process to make it difficult for the discriminator to classify the original and generated images. Then for the CIFAR-10 data, it can be seen that the loss value and discriminator accuracy are stable. However, the generator needs to produce a better image in the previous visual evaluation. It means that the value that appears in the discriminator loss or accuracy cannot be used to evaluate the resulting image but must be visually checked to determine

whether the resulting image is as desired and whether training is still needed. It is proven that when the number of iterations in training is increased again on CIFAR-10 data, the resulting image generator gets better. The resulting values appear unstable for the Oxford IIIIT pet data, meaning more training is needed.

### C. Segmentation

#### Segmentation without Additional Data

Further segmentation process is performed on Oxford IIIIT pet data. We used 920 images in the training phase. The original data is shown in the following Figure 17.



Figure 17. The original image to be predicted

The segmentation images generated during the training process, i.e. at each of the 5th, 10th, and 20th epochs are depicted in Figures 18 and 19 below.

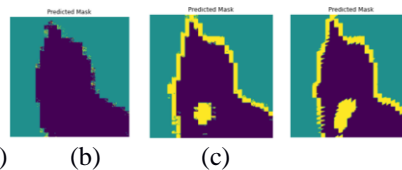


Figure 18. FCN segmentation results without additional data for (a) epoch=5, (b) epoch=10, and (c) epoch=20

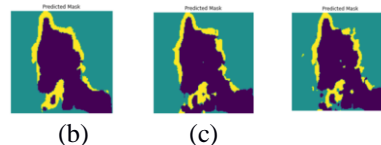


Figure 19. U-Net segmentation results without additional data for (a) epoch=5, (b) epoch=10, and (c) epoch=20

The pictures in Figure 17-19 show that the training process can perform better segmentation and resemble the actual segmentation along with the training process. Visually, with epoch 20 above, it can be obtained that the segmentation results on the U-Net method are better than the FCN method.

#### Segmentation with Additional Data

Before segmentation, the Oxford IIIIT Pet dataset is used to train the conditional-GAN model to generate synthetic data. Figure 20

shows the results of the data training process on cGAN after training with epoch=1000:

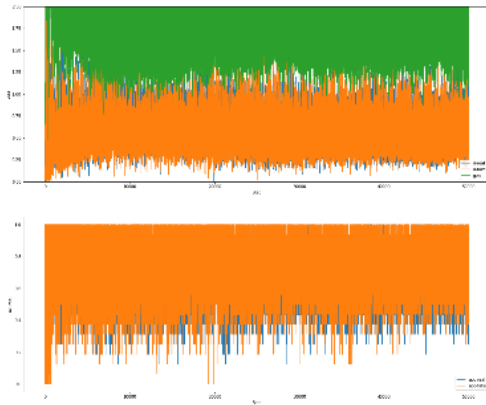


Figure 20. Graph of cGAN Training Loss and Accuracy on the Oxford IIIT Pet Dataset

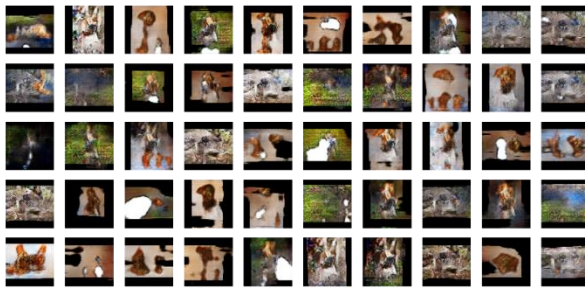


Figure 21. Oxford IIIT cGAN Revival Data Pet Dataset Epoch=1000

TABLE II  
EVALUATION OF REGIONAL DATA FOR THE OXFORD IIIT PET DATASET

Indicator	Epoch 10	Epoch 20	Epoch 40	Epoch 100	Epoch 1000
FI		355,10			205,193
D	381,152	3	332,625	221,914	1,00
	1,000817	1,0014	1,00169	1,00871	701±
	±0,0000	8±0,00	6±0,000	5±	0,00
IS	685	009	111	0,00101	059

Figure 20 shows that the resulting values have not converged and have a wide range. Most animal images (cats and dogs) do not resemble the desired shape, as shown in Figure 21. The evaluation in Table 2 is also in line with those results. In Table 2, the FID value up to epoch=1000 is 205.193. This value is high, meaning the actual and generated images differ. The resulting Inception Score (IS) value is small, meaning the generated data is homogenous

and lacks quality. It is likely because the datasets used have varied in size, color, and lighting, known as complex datasets. The complexity of the dataset in this paper means the complexity of image attribute compositions contained in an image/picture, such as lighting, coloring, and resolution. Each attribute contributes to the computations required to process the images. For instance, higher-resolution images require more computational resources to process than the lower-resolution images. Another example is that rich-colored images are more expensive to process than black-and-white images.

Next, 1000 synthetic data are generated using the cGAN model that has been trained. This data is then fed into segmentation training. The segmentation results are shown in Figures 22 and 23 below.

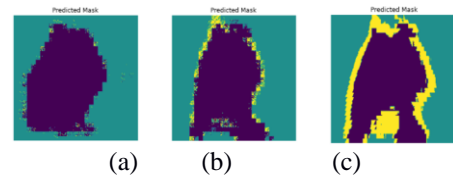


Figure 22. FCN segmentation results with additional data for (a) epoch=5, (b) epoch=10, and (c) epoch=20

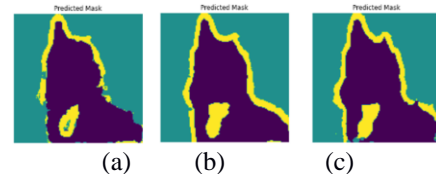


Figure 23. U-Net segmentation results with additional data for (a) epoch=5, (b) epoch=10, and (c) epoch=20

#### D. Comparison of Segmentation between with and without Synthetic Data

The difference in segmentation results between the two conditions with and without adding synthetic data can be seen in Table 3 and Table 4 below:

TABLE III  
COMPARISON OF SEGMENTATION WITH AND WITHOUT  
ADDITIONAL DATA USING THE FCN MODEL





Differentiator	Without Additional Data			With Additional Data		
Segmentation image						
Loss training	0.25	0.17	0.1	0.28	0.21	0.1
Training accuracy	92	51	300	96	10	33
Loss validation	0.86	0.91	0.9	0.86	0.90	0.
Validation accuracy	78	86	396	07	03	93
IoU	0,53066			0,50158		
Dice score	0,68536			0,6564		

TABLE IV  
COMPARISON OF SEGMENTATION WITH AND WITHOUT  
ADDITIONAL DATA USING THE U-NET MODEL

Differentiator	Without Additional Data			With Additional Data		
Segmentation image						
Loss training	0,11	0,05	0,03	0.16	0.093	0.05
Training accuracy	21	80	11	46	3	11
Loss validation	0,94	0,96	0,97	0.92	0.955	0.97
Validation accuracy	89	87	58	73	7	06
IoU	0,70878			0,68498		
Dice score	0,82656			0,80858		

From the two comparisons in Tables 3 and 4, the segmentation results on additional data have generally decreased. Similarly, in FCN and U-Net, the value of training loss increased, training accuracy decreased, and IoU and Dice Score decreased. Only the validation loss and validation accuracy show better results, as shown in the yellow highlights. It is because

the generated data that was previously used lacks similarity to the original data. However, there is still an influence on the results obtained. Decreasing the validation loss value and increasing validation accuracy when data is added shows that the model is getting better at determining model parameters. Therefore, in the future, better data generation can be done to obtain better segmentation results.

## CONCLUSIONS

In this study, the following conclusions are obtained:

1. The cGAN method can generate synthesis data well through training. More complex data requires training with more iterative processes.
2. Adding generated synthetic data with cGAN positively affects loss and validation accuracy, but vice versa for other evaluation indicators. It means there is an effect from the addition of synthetic data, even though not on all evaluation indicators. The insufficient similarity of generated synthetic images from the original images causes it.

## REFERENCES

- [1] Y. Roh, G. Heo och S. E. Whang, "A Survey on Data Collection for Machine Learning: a Big Data -- AI Integration Perspective," IEEE Transactions on Knowledge and Data Engineering, vol. 33, nr 4, pp. 1328 - 1347, 2019.
- [2] S. Mahapatra, "Towards Data Science," 22 March 2018. [Online]. Available: <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>. [Använd 26 November 2021].
- [3] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik och A. Swami, "The Limitations of Deep Learning in Adversarial Settings," IEEE European Symposium on Security & Privacy, IEEE 2016. Saarbrucken, Germany, pp. 372-387, 2016.
- [4] D. Nyberg, "Exploring the Capabilities of Generative Adversarial Networks in Remote

- Sensing Applications,” Linköping University Electronic Press, Linköping, 2021.
- [5] A. Antoniou, A. Storkey och H. Edwards, ”Data Augmentation Generative Adversarial Networks,” pp. 1-14, 2018.
- [6] Ministry of Defence, UK, Machine Learning with Limited Data, London: Data Science and Technology Laboratory, 2020.
- [7] D. Zhu, S. Xia, J. Zhao, Y. Zhou, Y. Zhou, M. Jian, Q. Niu, Q. Niu, R. Yao och Y. Chen, ”Diverse sample generation with multi-branch conditional generative adversarial network for remote sensing objects detection,” *Neurocomputing*, vol. 381, pp. 40-51, 2020.
- [8] M. D. Hossain och D. Chen, ”Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 15-134, 2019.
- [9] T. Kattenborn, J. Leitloff, F. Schiefer och F. Schiefer, ”Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24-49, 2021.
- [10] A. Kulkarni, T. Mohandoss, D. Northrup, B. Hill, S. Louis, E. Mwebaze och H. Alemohammad, ”Semantic Segmentation of Medium-Resolution Satellite Imagery using Conditional Generative Adversarial Networks,” *AI for Earth Sciences Workshop at NeurIPS*, 2020.
- [11] M. Rezaei, K. Harmuth, W. Gierke, K. T. Gierke, M. Fischer, H. Yang och C. Meinel, ”Conditional Adversarial Network for Semantic Segmentation of Brain Tumor,” 2017.
- [12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville och Y. Bengio, ”Generative Adversarial Nets,” *Advances in Neural Information Processing Systems*, pp. 1-9, June 2014.
- [13] M. Mirza och S. Osindero, ”Conditional Generative Adversarial Nets,” *ArXiv*, vol. abs/1411.1784, pp. 1-7, November 2014.
- [14] P. Yu, K. Song och J. Lu, ”Generating Adversarial Examples With Conditional Generative Adversarial Net,” i 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, 2018.
- [15] G. Heilemann, M. Matthewman, P. Kuess, G. Goldner, J. Widder, D. Georg och L. Zimmermann, ”Can Generative Adversarial Networks help to overcome the limited data problem in segmentation?,” *Zeitschrift für Medizinische Physik*, pp. 1-8, 2021.
- [16] A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López och G. Fichtinger, ”Conditional Generative Adversarial and Convolutional Networks for X-ray Breast Mass Segmentation and Shape Classification,” *Medical Image Computing and Computer Assisted Intervention*, vol. 11071, pp. 833-840, 2018.
- [17] K. Cheng, R. Tahir, L. K. Eric och M. Li, ”An Analysis of Generative Adversarial Networks and Variants for Image Synthesis on MNIST Dataset,” *Multimedia Tools and Applications*, vol. 79, nr 19-20, pp. 13725-13750, 2020.
- [18] J. Brownlee, *Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image Translation*, Machine Learning Mastery, 2019.





# KARAKTERISTIK SPASIOTEMPORAL ANGIN KENCANG DI PULAU SULAWESI BERBASIS MODEL REANALISIS ERA5

Achmad Fahrudin Rais<sup>1\*</sup>, Rezky Yunita<sup>2</sup>, Hijrah Kule Musgamy<sup>3</sup>, Rahmat Tauladani<sup>4</sup>,  
Ahmad Kosasih<sup>2</sup>

<sup>1</sup> Pusat Riset Geospasial (PRG)-Badan Riset dan Inovasi Nasional (BRIN), Cibinonga

<sup>2</sup>Badan Meteorologi Klimatologi dan Geofisika (BMKG), Jakarta

<sup>3</sup>Stasiun Meteorologi Sangia Ni Bandera, Sulawesi Tenggara

<sup>4</sup>Stasiun Meteorologi Yogyakarta Internasional Airport

e-mail: <sup>1</sup>achm050@brin.go.id

## ABSTRAK

Pola spasiotemporal angin kencang di Pulau Sulawesi dieksplorasi dalam tulisan ini menggunakan data reanalisis ERA5 periode 1979-2020. Metode yang digunakan adalah *rotated empirical orthogonal function* (REOF) dan fourier harmonik. Hasil yang didapatkan adalah bahwa frekuensi dan magnitudo kecepatan angin terbesar berada di Sulawesi Selatan dan sebagian Sulawesi Tenggara dengan puncak di bulan Agustus dan Januari, pola dominan lainnya adalah di Sulawesi Utara dan Gorontalo dengan pola puncak yang sama yaitu Agustus dan Januari dan di pegunungan tengah Sulawesi dengan puncak pada bulan Januari.

**Kata kunci**— angin kencang, ERA5, REOF

## ABSTRACT

*The spatiotemporal pattern of wind gusts over Sulawesi Island was explored in this paper using ERA5 reanalysis for the period 1979-2020. The methods used were rotated empirical orthogonal function (REOF) and Fourier harmonics. The results obtained were that the highest frequency and magnitude of wind speed were in South Sulawesi and parts of Southeast Sulawesi with peaks in August and January, the other dominant pattern were in North Sulawesi and Gorontalo with the same peak pattern, namely August and January and in the central mountains of Sulawesi with a peak in January.*

**Keywords**— *wind gust, ERA5, REOF*

## PENDAHULUAN

Sulawesi yang terletak di sekitar ekuator adalah area yang termasuk ke dalam wilayah doldrum (Crowe, 1951). Doldrum adalah istilah area dengan angin berkecepatan rendah (Ahrens; & Henson, 2014). Namun Sulawesi adalah merupakan area dengan 7 kabupaten memiliki risiko tertinggi terhadap cuaca ekstrim (BNPB, 2020). Cuaca ekstrim yang dimaksud adalah angin kencang.

Kajian angin kencang juga merupakan suatu hal yang menarik. Pasalnya, bencana ini dapat berpengaruh terhadap badan manusia (Gjeta dkk., 2020), menyebabkan kerusakan bangunan (Sarli dkk., 2020), kerusakan kendaraan (Hemingway, 2020) dan timbulnya gelombang tinggi di laut (Hilmi dkk., 2018).

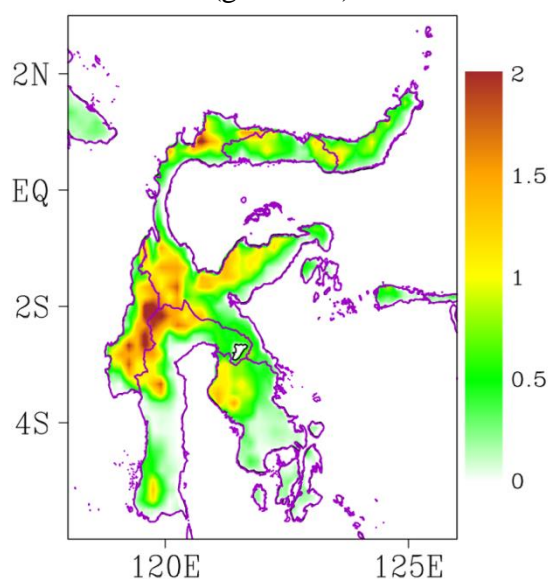
Umumnya, kajian angin kencang di Sulawesi sebelumnya hanya mengkaji angin kencang dalam area yang lebih kecil atau tidak mencakup seluruh Sulawesi dan tidak menganalisis pola temporal klimatologinya. Sulung dkk (2011) hanya mensimulasikan angin kencang di Kota Makassar pada 15-17 Januari 2011. Murlina (2013) dan Oktrifiani (2013) membangun model prediksi hanya di Maros dan Toraja Utara. Syahfitri dkk. (2021) membuat menganalisis tingkat bahaya angin kencang di Sidenreng Rappang saja. Oleh karena itu penulis akan mengkaji pola spasiotemporal angin kencang di seluruh Pulau Sulawesi.

## DATA DAN METODE

### Sumber Data

Data yang digunakan dalam penelitian ini adalah data wind gust *The European Centre for Medium-Range Weather Forecast Reanalysis 5-th Generation* yang biasa disingkat ERA5. Data setiap jam selama periode bulan Februari 1979 – Desember 2020 yang didapatkan dari <https://cds.climate.copernicus.eu>. Data kejadian puting beliung didapat dari data informasi bencana Indonesia Badan Nasional Penanggulangan Bencana (DIBI BNPB) <https://dibi.bnpb.go.id/xdibi> selama

periode 2011-2020. Secara definisi, puting beliung dan angin kencang berbeda dimana angin puting beliung atau tornado memiliki kecepatan >64.4 km/jam dan angin kencang memiliki kecepatan >46,3 km/jam dalam Perka BMKG No.9 Tahun 2010. Tetapi dalam DIBI BNPB, tidak ada kecepatan angin yang dicantumkan, melainkan hanya mencantumkan korban jiwa dan kerusakan. Dalam penelitian Sarli dkk. (2020), angin dengan kecepatan 28.8 km/jam dapat menyebabkan terjadinya kerusakan. Oleh karena itu, laporan puting beliung DIBI BNPB dianggap angin kencang pada penelitian ini. Area penelitian meliputi Pulau Sulawesi (gambar 1).



Gambar 1. Peta ketinggian (km) lokasi penelitian.

### Metode

Dalam mendapatkan pola-pola utama, penulis menggunakan teknik *rotated empirical orthogonal function* (REOF). REOF kurang rentan terhadap fitur buatan yang dihasilkan dari kendala ortogonalitas pada EOF tanpa rotasi (Wilks, 2019). Untuk menghitung REOF (Dommenget, 2015), terlebih dahulu mencari nilai eigen ( $\lambda$ ) dari determinasi kovariansi matriks data angin kencang ( $\Sigma$ ) dan matriks identitas spasial ( $I_s$ ) menggunakan rumus sebagai berikut :

$$\text{Det}(\Sigma - \lambda \cdot I_s) = 0 \quad (1)$$

Kemudian mencari vektor eigen ( $\pi$ ) dengan formula sebagai berikut :

$$(\Sigma - \lambda \cdot I_s)\pi = 0 \quad (2)$$

Vektor eigen terotasi ( $\pi_r$ ) didapatkan dari VARIMAX  $\pi$  dengan formula :

$$S_{var}(\pi_r) = \frac{1}{S} \sum_{i=1}^S \left( \frac{x_i}{\sigma_i} \right)^4 - \frac{1}{S^2} \left( \sum_{i=1}^S \left( \frac{x_i}{\sigma_i} \right)^4 \right)^2 \quad (3)$$

$\sigma$  adalah normalisasi  $\pi$ , S sebagai dimensi spasial dan  $S_{var}(\pi_r)$  merupakan spasial vector eigen VARIMAX. Dengan persamaan 2 didapatkan  $\lambda$  terotasi ( $\lambda_r$ ). Untuk selanjutnya  $\lambda_r$  disebut PC dan  $\pi_r$  disebut REOF. Dalam menganalisis setiap pola yang dihasilkan harus mengkombinasikan REOF dengan PC sehingga mengetahui pola yang lemah dan pola yang kuat.

Penulis menggunakan *fourier* harmonik (Wilks, 2019) sebanyak 4 harmonik (1, 1/2, 1/4 dan 1/8 tahunan) untuk *smoothing* grafik PC sehingga mendapatkan pola PC berdasarkan banyaknya puncak dalam satu tahun (365 hari). PC yang di-*smooth* (PCs) dihitung dengan rumus sebagai berikut :

$$PC_s(t) = \overline{PC} + \sum_{n=1}^4 \left( a_n \cos \frac{n\pi t}{365} + b_n \sin \frac{n\pi t}{365} \right) \quad (4)$$

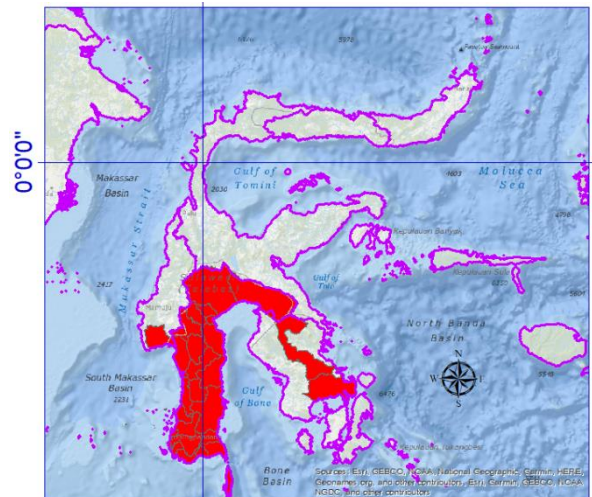
Dengan  $\overline{PC}$ , n,  $\pi$  dan t adalah PC rata-rata dan banyaknya harmonic,  $180^\circ$  dan waktu.

## HASIL DAN DISKUSI

### Analisa Data DIBI BNPB

Pada gambar 2, laporan angin kencang terbanyak berada di Sulawesi Selatan dengan kasus total laporan sebanyak 404 kejadian. Laporan ini tentu berkaitan dengan banyaknya penduduk di Sulawesi Selatan dibandingkan provinsi lain di Sulawesi (BPS, 2021) sehingga banyak yang terdampak dan melaporkan. Angin kencang Sulawesi Selatan umumnya berasal dari kabupaten yang berbatasan dengan laut dengan kejadian terbanyak berasal dari Kabupaten Pangkajene Kepulauan, Takalar serta Bulukumba yang secara berurutan adalah 37, 36 dan 25. Kejadian angin puting beliung paling sedikit berasal dari Enrekang dan Luwu Timur. Delapan puluh empat kejadian angin kencang berasal dari Sulawesi Tenggara dengan 12 dan 11 yang berasal dari Kabupaten Konawe dan Konawe Selatan.

Di Sulawesi Barat, 19 dari 33 kejadian angin kencang pada bulan Oktober berasal dari Kabupaten Polewali Mandar. Di Sulawesi Utara, konsentrasi kejadian terbanyak berada di kepulauan sebelah utara. Sulawesi Tengah dan Gorontalo adalah provinsi dengan angin kencang paling sedikit.



Gambar 2. Peta distribusi angin puting beliung terbanyak di Pulau Sulawesi berdasarkan DIBI BNPB.

### Analisis Model Reanalisis ERA5

Delapan mode REOF dan PCs dianalisis dalam penelitian ini dengan total variansi bernilai 73,9%. Nilai tersebut menunjukkan bahwa mayoritas variansi data sudah diwakili oleh 8 mode REOF dan PCS.

Spasial mode 1 pada gambar 4, yang memiliki variansi sebesar 20,61%, memperlihatkan nilai vektor eigen positif semua dengan nilai terbesar berada di Sulawesi Selatan dan Sulawesi Tenggara. Nilai positif menandakan bahwa grafik temporal tidak terbalik karena rekonstruksi data awal adalah penjumlahan dari spasial vector eigen dikalikan nilai PC-nya pada setiap mode (Wilks, 2019). Hasil ini memiliki pola spasial yang menyerupai pola sebaran kejadian puting beliung DIBI BNPB (gambar 2). Mode 1 memiliki puncak terbesar pada bulan Agustus dan bulan Januari pada gambar 3. Pada bulan Agustus angin permukaan yang kencang

terjadi di sekitar perairan Sulawesi bagian selatan yang bertiup dari tenggara (Alifdini, 2021).

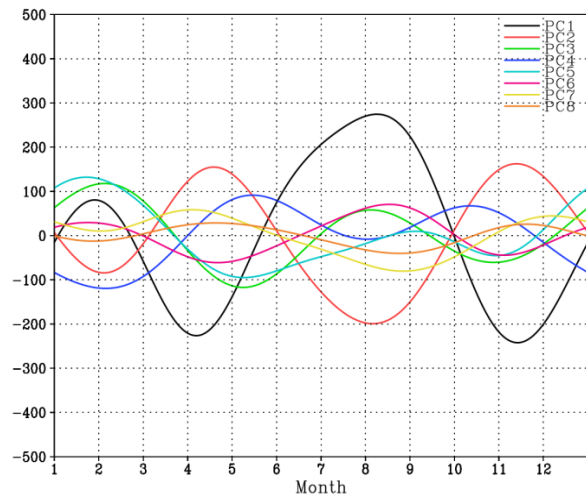
REOF 2 dengan variansi 12.05% menunjukkan nilai vector eigen negatif semua terutama di Gorontalo, Sulawesi Utara dan Sulawesi Tengah bagian barat pada gambar 3. Nilai negatif ini membalikkan grafik PCs 2 menjadi PCs dengan puncak pada bulan Agustus pada gambar 3. Tampaknya, angin permukaan juga bernilai besar pada bulan Agustus di perairan sekitar Gorontalo dan Sulawesi Utara dalam penelitian Alifdini (2021).

REOF 3, 4 dan 8 dengan variansi 11,59%, 9,84%, 2,63% bernilai kecil di Pulau Sulawesi pada gambar 4 secara umum. Nilai REOF yang kecil menandakan lemahnya PCs 3, 4 dan 8 di Pulau Sulawesi.

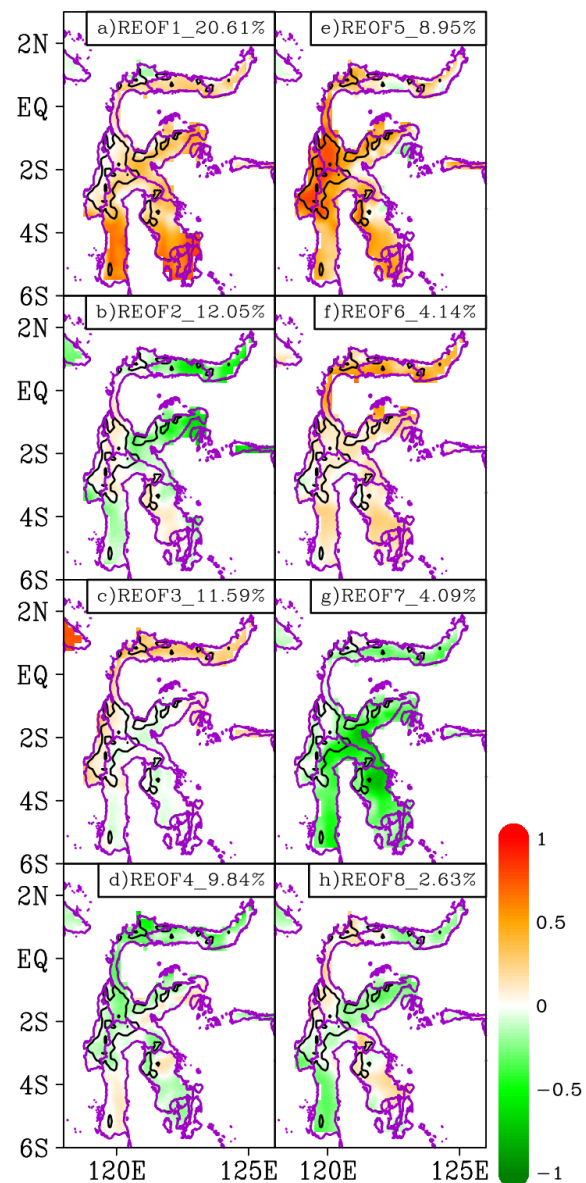
REOF 5 dengan variansi 8,95% memiliki sebaran nilai terbesar berada di dataran tinggi Sulawesi bagian tengah pada gambar 4. Dibandingkan REOF lainnya, hanya REOF 5 yang paling besar di dataran tinggi Sulawesi bagian tengah yang artinya merupakan pola paling dominan di area tersebut. PCs 5 memiliki puncak tertinggi pada bulan Januari pada gambar 3.

REOF 6 (variansi 4,14%) menunjukkan nilai positif semua dan distribusi terbesar berada di Sulawesi bagian utara pada gambar 4. Nilai yang tidak terlalu besar tampak pula di sebagian Sulawesi Selatan dan Tenggara. Sebarannya tampak menyerupai REOF 2, tetapi berbeda di Sulawesi Tengah bagian utara dan barat. PCs 6 juga memiliki pola yang sama dengan PCs 2 dengan REOF 2 negatif.

Pada gambar 4, REOF 7 (variasi 4,09%) bernilai negatif dan memiliki nilai paling kecil di sekitar Sulawesi Selatan, Sulawesi Tenggara dan Sulawesi Tengah bagian barat. Polanya menyerupai REOF 1, kecuali di Sulawesi Tengah. Ketika PCs 7 dikalikan dengan nilai negatif karena REOF 7 negatif, PCs 7 memiliki pola temporal seperti PCs 1 pada gambar 3.



Gambar 3. Grafik PCs 1 sampai PCs 8.



Gambar 4. REOF a) mode 1, b) mode 2, c) mode 3, d) mode 4, e) mode 5, f) mode 6, g) mode 7, dan h) mode 8. REOF dioverlay dengan data ketinggian GLOBE



## KESIMPULAN

Explorasi data angin kencang model reanalisis ERA5 di Pulau Sulawesi dilakukan dalam penelitian ini dengan menggunakan metode REOF dan fourier harmonik yang dikomparasi dengan data DIBI BNPB. Setelah menganalisis data-data tersebut, penulis berkesimpulan bahwa angin kencang di Sulawesi Selatan dan Sulawesi Tenggara adalah yang dominan terjadi di Pulau Sulawesi dan memiliki puncak pada bulan Agustus yang dipengaruhi oleh angin dari tenggara, angin kencang di Sulawesi Utara, Gorontalo dan Sulawesi Tengah bagian barat juga memiliki pola temporal yang sama dengan pola dominan tetapi dengan magnitudo yang lebih kecil serta angin kencang di pegunungan tengah Sulawesi dengan magnitudo yang lebih kecil dibandingkan pola-pola sebelumnya memiliki puncak pada bulan Januari.

Untuk lebih lanjut, kajian yang menganalisis data observasi diperlukan untuk melengkapi kajian angin kencang di Pulau Sulawesi.

## KONTRIBUSI PENULIS

Penulis pertama adalah kontributor utama mengolah data, menganalisis data dan membuat naskah. Penulis ke-2 sampai ke-4 menelaah naskah.

## DAFTAR PUSTAKA

- Ahrens, C. D., & Henson, R. (2014). *Meteorology today : an introduction to weather, climate, and the environment*. Cengage.
- Alifdini, I. (2021). *Seasonal distribution and variability of surface winds in the Indonesian seas using scatterometer and reanalysis data. July 2020*, 1–19. <https://doi.org/10.1002/joc.7101>
- Amin Nur Syafitri, Maru, R., & Invanni, I. (2021). Analisis Tingkat Bahaya Bencana Angin Puting Beliung Berbasis Sistem Informasi Geografis di Kabupaten Sidenreng Rappang. *Jurnal Environmental Science*, 3(April). [ojs.unm.ac.id/JES](https://ojs.unm.ac.id/JES)
- BNPB. (2020). *IRBI Indeks Risiko Bencana Indonesia Tahun 2020*. BNPB.
- BPS. (2021). Hasil Sensus Penduduk 2020. *Berita Resmi Statistik*, 7, 1–12.
- Crowe, P. . R. . (1951). Wind and Weather in the Equatorial. *Institute of British Geographers Transactions*, 17(17), 23–76.
- Dommenget, D. (2015). *An Introduction to Statistical Analysis in Climate*.
- Gjeta, A., Malka, L., Hasanaj, A., & Alcani, M. (2020). *The wind gusts effects on human body based on cfd simulations*. 142(4), 138–142.
- Hemingway, R. (2020). *Developing a hazard-impact model to support impact-based forecasts and warnings : The Vehicle OverTurning ( VOT ) Model*. January. <https://doi.org/10.1002/met.1819>
- Hilmi, I., Ningsih, N. S., Sofian, I., Hanifah, F., & Rizal, A. M. (2018). *The study of MJO impact on wave height and wind speed in Indonesian Seas*. <https://doi.org/10.1088/1755-1315/162/1/012004>
- Murlina, E. (2013). *Prediksi Puting Beliung di Kabupaten Maros*.
- Okstrifiani, N. (2013). *Prediksi Puting Beliung Di Kabupaten Toraja Utara*. Universitas Hasanuddin.
- Sarli, P. W., Abdillah, M. R., & Sakti, A. D. (2020). Relationship between wind incidents and wind-induced damage to construction in West Java , Indonesia Relationship between wind incidents and wind-induced damage to construction in West Java , Indonesia. *IOP Conf. Ser.: Earth Environ. Sci*, November, 13. <https://doi.org/10.1088/1755-1315/592/1/012001>
- Sulung, G., Priyanka, M., Saraswati, N., Nurfiena, S. P., & Ricardo, L. G. (2011). *Pengaruh Parameterisasi Kumulus terhadap Simulasi Angin Kencang di Makassar dengan Menggunakan WRF*.
- Wilks, D. S. (2019). *Statistical Methods in the Atmospheric Sciences*. Elsevier Inc.





# Petunjuk Penulisan

## JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

Naskah dikirim dalam bentuk *softcopy* ke alamat email [pppm@stis.ac.id](mailto: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., Brunsdon, 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*.