

## Jurnal Ilmiah

# Informatika Komputer

ROBOT EDUKASI PERTANIAN AGROBOT-I: RANCANGAN ELEKTRONIKA DAN SISTEM PENGGERAK <b>Yogi Permadi, Sandy Suryo Prayogo, Tubagus Maulana Kusuma</b>	1
PERAMALAN TINGKAT PENGANGGURAN DI INDONESIA MENGGUNAKAN METODE TIME SERIES MODEL ARIMA DAN HOLT-WINTERS <b>Agus Sulaiman, Asep Juarna</b>	13
EVALUASI TATA KELOLA TEKNOLOGI INFORMASI PADA DINAS TENAGA KERJA KABUPATEN BOGOR MENGGUNAKAN PROCESS ASSESSMENT MODEL (PAM) COBIT 5 <b>Yopie Noor Hantoro, Suryarini Widodo</b>	29
PREDIKSI RATA-RATA ZAT BERBAHAYA DI DKI JAKARTA BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY <b>Anisa Oktaviani, Hustinawati</b>	41
KLASIFIKASI AREA GEMPA BUMI MENGGUNAKAN ALGORITMA RANDOM FOREST <b>Ismail</b>	56
KLASIFIKASI PENGADUAN LARAS ONLINE BERBASIS TEXT MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN NAIVE BAYES <b>Muhammad Azis Suprayogi</b>	65
MODEL AUTOREGRESIVE INTEGRATED MOVING AVERAGE (ARIMA) DALAM PERAMALAN NILAI HARGA SAHAM PENUTUP INDEKS LQ45 <b>Devita Priyadi, Iffatul Mardhiyah</b>	78

## **DEWAN REDAKSI JURNAL ILMIAH INFORMATIKA KOMPUTER**

### **Penanggung Jawab**

Prof. Dr. E.S. Margianti, S.E., M.M.  
Prof. Suryadi Harmanto, SSI., M.M.S.I.  
Drs. Agus Sumin, M.M.S.I.

### **Dewan Editor**

Dr. Fitriarningsih, Universitas Gunadarma  
Dr. Rodiah, Universitas Gunadarma  
Dr. Ias Sri Wahyuni, S.Si, M.Si, Universitas Gunadarma

### **Mitra Bebestari**

Prof. Dr. Rer. Nat. A. Benny Mutiara, Universitas Gunadarma  
Prof. Dr. Achmad Nizar Hidayanto, Universitas Indonesia  
Prof. Dr.-Ing. Adang Suhendra, S.Si, S.Kom, M.Sc, Universitas Gunadarma  
Prof. Dr. Sarifuddin Madenda, Universitas Gunadarma  
Surya Sumpeno, PhD, Institut Teknologi Sepuluh Nopember (ITS)  
Noor Akhmad Setiawan, Ph.D, Universitas Gadjah Mada  
Dr. Paulus Insap Santosa, Universitas Gadjah Mada  
Dr. Agus Harjoko, M.Sc, Ph.D, Universitas Gadjah Mada  
Dr. Ernastuti, Universitas Gunadarma  
Dewi Agushinta R., Universitas Gunadarma  
Dr. Eri Prasetyo, Universitas Gunadarma  
Dr. Sunny Arief Sudiro, STMIK Jakarta STI&K  
Dr. Tubagus Maulana Kusuma, S.Kom.,Mengsc, Universitas Gunadarma  
Dr. Lussiana ETP, S.si., M.T, STMIK Jakarta STI&K  
Harry Budi Santoso, Ph.D, Universitas Indonesia

### **Sekretariat Redaksi**

Universitas Gunadarma  
Jalan Margonda Raya No. 100 Depok 16424  
Phone : (021) 78881112 ext 516.

# JURNAL ILMIAH INFORMATIKA KOMPUTER

NOMOR 1, VOLUME 26, APRIL 2021

## DAFTAR ISI

ROBOT EDUKASI PERTANIAN AGROBOT-I: RANCANGAN ELEKTRONIKA DAN SISTEM PENGGERAK Yogi Permadi, Sandy Suryo Prayogo, Tubagus Maulana Kusuma	1
PERAMALAN TINGKAT PENGANGGURAN DI INDONESIA MENGGUNAKAN METODE TIME SERIES DENGAN MODEL ARIMA DAN HOLT-WINTERS Agus Sulaiman, Asep Juarna	13
EVALUASI TATA KELOLA TEKNOLOGI INFORMASI PADA DINAS TENAGA KERJA KABUPATEN BOGOR MENGGUNAKAN PROCESS ASSESSMENT MODEL (PAM) COBIT 5 Yopie Noor Hantoro, Suryarini Widodo	29
PREDIKSI RATA-RATA ZAT BERBAHAYA DI DKI JAKARTA BERDASARKAN INDEKS STANDAR PENCEMAR UDARA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY Anisa Oktaviani, Hustinawati	41
KLASIFIKASI AREA GEMPA BUMI MENGGUNAKAN ALGORITMA RANDOM FOREST Ismail	56
KLASIFIKASI PENGADUAN LARAS ONLINE BERBASIS TEXT MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN NAÏVE BAYES Muhammad Azis Suprayogi	65
MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DALAM PERAMALAN NILAI HARGA SAHAM PENUTUP INDEKS LQ45 Devita Priyadi, Iffatul Mardhiyah	78

# ROBOT EDUKASI PERTANIAN AGROBOT-I: RANCANGAN ELEKTRONIKA DAN SISTEM PENGGERAK

<sup>1</sup>Yogi Permadi, <sup>2</sup>Sandy Suryo Prayogo, <sup>3</sup>Tubagus Maulana Kusuma  
<sup>123</sup>Fakultas Teknologi Industri Universitas Gunadarma,  
<sup>123</sup> Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat  
{yogi, sandy\_sr, mkusuma}<sup>123</sup>@staff.gunadarma.ac.id

## Abstrak

Berkurang peminat generasi muda terhadap pertanian konvensional berdampak pada penurunan ketersediaan pangan, untuk mengatasi permasalahan tersebut, maka teknologi otomasi dibidang pertanian perlu dikembangkan selain untuk mempermudah juga untuk meningkatkan minat generasi penerus pertanian di Indonesia. Pada penelitian ini dirancang dan dibangun sebuah prototipe robot pertanian untuk keperluan edukasi dan penelitian dalam hal otomasi tanam dan panen tanaman padi yang diberi nama Agrobot-I. Robot ini dapat bergerak medan tanah lahan pertanian untuk melakukan proses tanam, perawatan tanaman dari gangguan gulma dan proses panen tanaman padi yang dilengkapi dengan mekanik gripper yang menyerupai lengan sebagai alat bantu untuk melakukan ketiga pekerjaan tersebut. Robot yang memiliki tujuan utama untuk sarana edukasi dan pengenalan terhadap aplikasi teknologi pada bidang pertanian ini diharapkan dapat memberikan gambaran proses pertanian yang sesungguhnya, meskipun hanya dalam bentuk simulasi di lingkungan buatan. Pengujian dilakukan terhadap fungsi masing-masing sistem penggerak yang dikendalikan menggunakan mikrokontroler Arduino dari pergerakan motor DC yang menggunakan sistem differensial drive. Pengujian terhadap lengan robot dari pneumatik untuk menaik turunkan lengan, pengujian cartesian untuk sb-x dan sb-y dari lengan, dan juga lengan itu sendiri yang menggunakan motor servo. Hasil pengujian terhadap fungsi robot secara keseluruhan telah berhasil dilakukan, yaitu dari proses tanam, pencabutan gulma, dan panen.

**Kata Kunci:** Arduino, Otomasi, Pertanian, Robot, Sistem Penggerak.

## Abstract

Reducing the interest of young people in conventional agriculture has an impact on reducing the availability of food, to overcome these problems, the automation technology in agriculture needs to be developed in addition to making it easier to increase the interest of the next generation of agriculture in Indonesia. In this research, a prototype of an agricultural robot was designed and built for the purposes of education and research in terms of automation of planting and harvesting rice plants named Agrobot-I. This robot can move the land of agricultural land to do the planting process, care of plants from weed disturbance and the process of harvesting rice plants equipped with mechanical gripper that resembles an arm as a tool to do all three jobs. The robot, which has the main goal of educational facilities and introduction to technology applications in agriculture, is expected to provide an overview of the actual agricultural process, even if only in the form of simulations in an artificial environment. The test is carried out on the function of each drive systems which is controlled using an Arduino microcontroller from the movement of a DC motor using a differential drive system. Tests on the pneumatic robotic arms for raising and lowering arms, cartesian testing for sb-x and sb-y from the arms, and also the arms themselves using servo motors. The results of testing of the overall robot function have been successfully carried out, namely from the process of planting, weeding, and harvesting.

**Keywords:** Agriculture, Arduino, Automation, Drive Systems, Robots.

## PENDAHULUAN

Pengamat Ekonomi Institute for Development of Economics and Finance (INDEF), Bhima Yudistira, menyebut teknologi pertanian di Indonesia masih tertinggal di era digitalisasi. Sebab, kesadaran teknologi di kalangan petani masih rendah. [1] Rendahnya tingkat pendidikan dan minimnya pelatihan penggunaan teknologi modern merupakan faktor utama sektor pertanian masih bersifat tradisional. "Di level daerah pendidikan banyak yang tamatan SMP ke bawah. Selain itu, memang diperlukan bantuan dari pemerintah. Seperti, pelatihan teknologi terbaru," ungkapnya. [1] Kebutuhan pangan akan terus bertambah seiring bertambahnya jumlah penduduk, jika masih menggunakan system pertanian tradisional tidak tertutup kemungkinan target pemenuhan pangan tidak akan tercapai.

Kemajuan teknologi yang semakin modern menuntut sektor pertanian untuk mengikuti kemajuan yang terjadi pada dinamika perkembangan teknologi di industri. Untuk dapat menerapkan system pertanian modern harus didukung juga oleh sumber daya manusia yang dapat beradaptasi dengan perubahan teknologi pertanian. Hal ini membutuhkan kerjasama antara generasi muda yang lebih fokus di bidang teknologi dengan para petani konvensional yang lebih mengenal lapangan dan kendala yang ada, sehingga menghasilkan sinergi untuk meningkatkan hasil produksi pangan di Indonesia. Dengan

menggabungkan teknologi ke dalam pertanian maka akan menarik minat generasi muda dan pelajar untuk memulai belajar bertani maupun penelitian dibidang teknologi pertanian.

Salah satu masalah teknis yang dihadapi oleh petani adalah penggunaan pestisida sintetik skala besar untuk mengendalikan gulma, serangga, dan penyakit. Penggunaan pestisida berlebihan akan sangat berbahaya kepada produk pangan yang akan dikonsumsi manusia. Oleh karena itu ada kebutuhan untuk menemukan cara-cara baru untuk menghasilkan tanaman yang tidak memerlukan atau mengurangi penggunaan pestisida. Disamping itu dengan pertanian tradisional butuh banyak tenaga manusia untuk dapat merawat tanaman pada area lahan yang luas dan pekerjaan rutinitas, hal ini menjadi tidak efisien.

Mengacu pada beberapa permasalahan di atas, untuk meningkatkan produktifitas pertanian dan membentuk sumber daya manusia yang dapat menerapkan dan mengembangkan teknologi pertanian yang tepat sesuai kebutuhan. Fokus terbaru dari agri-robotika adalah penerapan aplikasi otomatisasi untuk tugas yang berulang lebih efisien atau efektif dari pada pendekatan manusia atau mesin besar tradisional. [2] Penelitian diperlukan ke platform robot yang dapat beroperasi dekat dengan tanaman (baik di tanah atau di ketinggian) dan manipulasi lanjutan, terutama dengan sifat interaktif atau taktis. Penggunaan teknologi robotika untuk membantu para petani menyelesaikan tugas

menanam dan merawat hingga memanen sudah banyak digunakan di negara maju. Saat ini inovasi robot untuk penyiangan tanaman yang mengurangi kebutuhan herbisida dengan mengerahkan lengan mekanis yang dipandu oleh kamera, untuk mencabut gulma. Meskipun masih baru disimulasikan, teknologi ini menunjukkan harapan besar. Robot juga dapat digunakan sebagai bagian dari sistem pengelolaan hama terpadu, misalnya, untuk penyiangan gulma yang akurat dan murah untuk menangkal hama dan penyakit tanaman. Sebagai langkah awal di rancang robot prototipe pertanian dengan nama *Agrobot-I* adalah sebuah robot yang dapat melakukan tiga tugas yaitu menanam bibit padi, mencabut rumput liar dan memanen padi, ketiga fungsi tersebut dapat di kendalikan langsung oleh manusia atau secara otomatis melakuakan pekerjaan secara mandiri. Robot yang di rancang ini masih bersifat simulasi pertanian untuk tujuan edukasi dan penelitian untuk pengembangan robot pertanian yang nyata. Rancangan robot dengan sistem beroda empat diharapkan dengan mudah dapat melintasi tanah lading, serta rancangan mekatronika lain seperti halnya untuk menanam benih yang juga berfungsi untuk mencabut rumput liar berupa lengan mekanik bisa berfungsi sesuai kebutuhan. Keseluruhan system mekatronika pada robot akan di kendalikan oleh sebuah mikrokontroler yang bertugas mengtur semua pergerakan actuator secara manual maupun otomatis. Penggunaan lengan mekanis untuk menanggulangi gulma mengacu pada

penelitian sebelumnya tinjauan terhadap penelitian dan literatur komersial menemukan bahwa mayoritas (75%) dari robot gulma menggunakan kontrol mekanis [3]. Penelitian lain mengenai robot edukasi simulasi pertanian dilakukan oleh Ade Reza Septiadi dan Syaiful Amri tentang Rancang bangun dan Analisa Robot Tematik Simulasi Pertanian dengan Kendali Wireless [4] Robot tematik pertanian adalah robot yang dirancang untuk melakukan simulasi menanam padi yang dikendalikan dengan jarak jauh secara nirkabel, mencabut rumput di antara batang padi dan memanen padi. Desain mekanis roda penggerak menggunakan roda jenis *rubber track* ini dapat bergerak meskipun berada pada permukaan yang tidak rata. Mekanis penanam menggunakan sistem *gripper* yang menggunakan motor servo untuk membuka dan menjepit bibit padi dan menggunakan sebuah *cylinder pneumatik* berukuran 30 cm.

## METODE PENELITIAN

Perancangan sistem elektronika dan sistem penggerak robot dibagi menjadi tiga bagian, yaitu pembuatan rangka robot sebagai struktur utama bentuk fisik robot dengan menggunakan bahan aluminium profile. Selanjutnya adalah pembuatan sistem gerak robot dengan model differencial drive untuk memudahkan robot bergerak di area ladang dan juga sistem gerak mekanis untuk fungsi menanam benih dan mencabut gulma dengan model lengan mekanis. Bagian terakhir

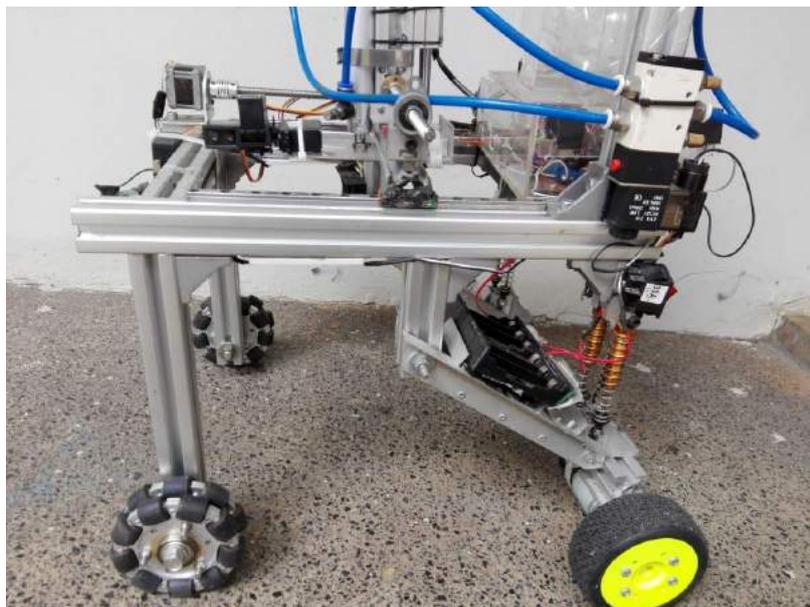
adalah pembuatan sistem elektronika untuk dapat mengendalikan semua sistem penggerak pada robot sesuai dengan data yang dikirim dari.

### **Bentuk Fisik Robot**

AGROBOT-I dibuat berbentuk robot mobil beroda empat dengan menggunakan roda karet *off-road* pada bagian belakang dan sistem suspensi untuk stabilitas bergerak di antara tanaman dan beroperasi pada permukaan lahan pertanian padat.

Rangka robot pada Gambar 1. dirakit menggunakan bahan batang aluminium *profile* tipe 30x30, dengan dimensi Panjang lebar tinggi 50 x 40 x 60 cm dan bobot robot 12 kg. Pada bagian tengah robot memanjang dari bagian depan sampai belakang dibuat jalur agar bibit padi yang telah ditanam melawati bagian bawah robot sampai ke

belakang dan keluar dari bagian robot. Teknik seperti ini bertujuan agar bagian roda robot tidak melindas bibit padi yang sudah ditanam. Roda karet bagian belakang terhubung langsung ke motor DC (*Direct Current*) yang sudah dilengkapi dengan *gearbox* dan roda bagian depan menggunakan mekanisme roda *omnidirectional*. Robot dilengkapi dengan *Gripper* sebagai lengan untuk menaruh bibit padi dan mengambil rumput liat pada lahan yang dapat mengganggu pertumbuhan padi. Terdapat dua jenis mekanis pada manipulator tersebut, yang pertama adalah sistem cartesian dengan *sliding linear bearing* yang di gerakkan oleh motor *stepper* agar dapat mengubah posisi *gripper* terhadap sumbu x dan y dan untuk pergerakan terhadap sumbu z menggunakan mekanisme *pneumatic* sehingga dapat menurunkan dan menaikkan posisi *gripper*.



**Gambar 1. Bentuk Fisik AGROBOT-I**

## Sistem Gerak Robot

Dikendalikan mekanisme *Differential drive* atau berpengerak dua buah roda yang dikemudikan secara terpisah menjadi mekanisme *steering* Untuk mengubah orientasi robot [5].

Roda penggerak biasanya terletak pada masing-masing sisi samping robot, dan menggunakan dua roda sehingga membutuhkan suatu alat tambahan agar robot dapat menyeimbangkan diri.

Robot *Agrobot-I* menggunakan roda pendukung yang diletakan di depan dan roda pendukung ini diharapkan tidak mengganggu pergerakan robot mobil sehingga roda pendukung harus dapat bergerak bebas. Penggerak utama robot menggunakan dua buah Motor DC (*Direct Current*) gearbox dengan tipe PG36 *torque* 10kgfcm 600rpm yang membutuhkan tegangan DC sebesar 24 volt. Sedangkan pada roda bagian depan menggunakan mekanisme *omnidirectional* yang dapat bergerak ke sisi kiri dan kanan dari roda utamanya untuk mengendalikan

arah orientasi robotnya (Persamaan 1 dan Persamaan 2).

Postur robot

$$P = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \quad (1)$$

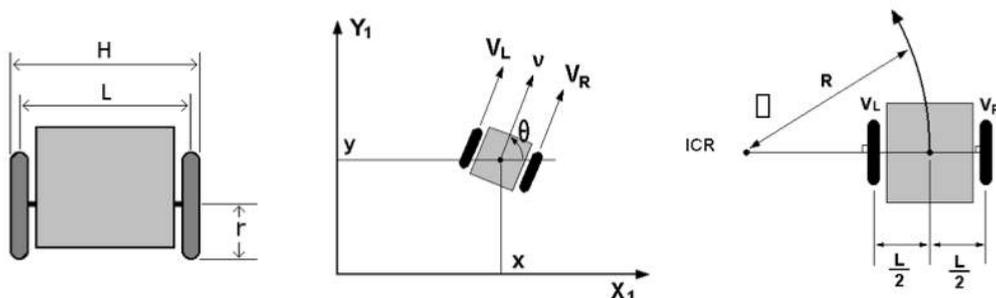
(x,y) : posisi robot,  $\theta$  : Orientasi robot

Control Input

$$U = \begin{pmatrix} v \\ \omega \end{pmatrix} \quad (2)$$

$v$  : kecepatan linear robot,  $\omega$  : kecepatan sudut robot

Bagaimana robot bermanuver dengan menggunakan dua roda penggerak utama di tunjukan Gambar 2. Kelebihan menggunakan sistem tersebut adalah cukup murah untuk dibuat, mudah direalisasikan dan desain cukup sederhana, tetapi ada kerugiannya juga yaitu sukar untuk melakukan pergerakan lurus *speed control* dan jika diameter roda tidak konsisten dapat menyebabkan kesalahan posisi [6]. Penggunaan roda dalam sebuah mekanik dimanfaatkan untuk mengurangi friksi yang terjadi.



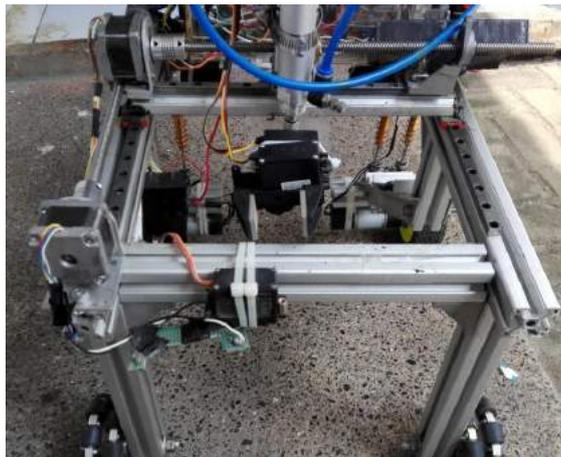
Gambar 2. *Differential Drive*

### Manipulator Lengan

Konfigurasi *cartesian* pada struktur robot ini terdiri dari tiga sumbu linier (*prismatic*) [7]. Masing-masing sumbu dapat bergerak ke area sumbu x-y-z.

Keuntungan sistem ini adalah pengontrolan posisi *gripper* agar tepat pada area tanam dan mempunyai struktur yang lebih kokoh. Pada Gambar 3 terdapat dua jenis mekanis pada manipulator tersebut, yang pertama adalah sistem *cartesian* dengan *sliding linear bearing* yang di gerakkan oleh besi ulir yang dapat mengubah posisi *gripper* terhadap sumbu x dan y untuk mencari posisi

yang tepat agar dapat melakukan penanaman padi dan mencabut rumput liar untuk pergerakan terhadap sumbu z menggunakan mekanisme *pneumatic* sehingga dapat menurunkan dan menaikkan posisi *gripper*. *Cylinder pneumatic* dengan panjang cylinder 30 cm sehingga posisi tanaman dapat menyentuh tanah dan *gripper* dapat melepaskan bibit padi. Penggunaan sistem *pneumatic* untuk menurunkan *gripper* sampai posisi tanam dapat di lihat pada Gambar 4 karena sistem *pneumatic* bekerja lebih cepat dibandingkan *linear actuator* yang bekerja dengan prinsip ulir.



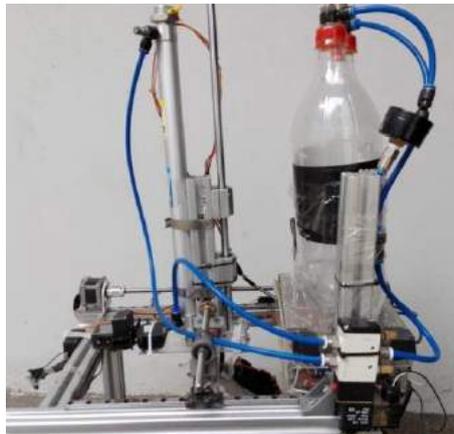
**Gambar 3. Konfigurasi Cartesian**



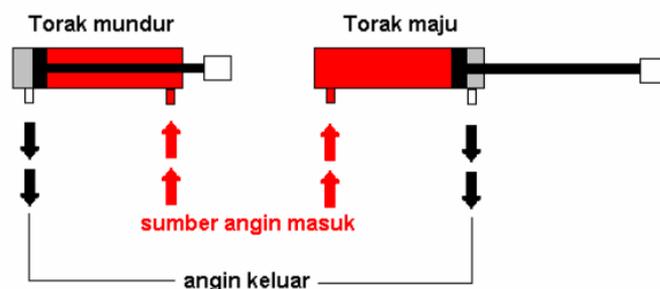
**Gambar 4. Posisi Gripper Turun**

Untuk mengontrol sistem pneumatic membutuhkan dua kondisi saja sehingga dapat menggunakan modul relay untuk mengaktifkan solenoid *valve* maka *pneumatic* sudah bisa dikendalikan melalui program *arduino*. Tetapi sistem *pneumatic* juga membutuhkan tekanan udara, pada robot ini digunakan 2 botol plastik minuman soda berukuran 1,5 liter masing-masing untuk menampung udara bertekanan, dapat dilihat pada Gambar 5. Ada beberapa pilihan model *Cylinder Pneumatic* diantaranya model *Single*

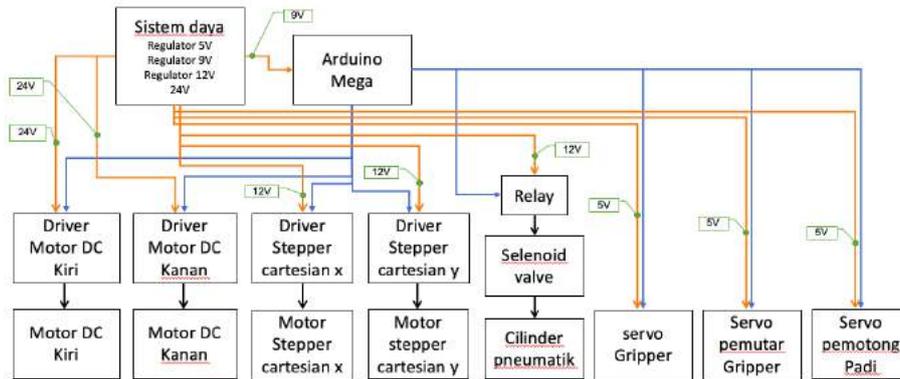
*Acting Cylinder* dan *Double Acting Cylinder*. *Double Acting Cylinder* (Silinder Kerja Ganda) ini mendapat suplai udara dari dua sisi. konstruksinya hampir sama dengan silinder kerja tunggal. Keuntungannya adalah bahwa silinder ini dapat memberikan tenaga kepada dua belah sisinya dapat dilihat pada Gambar 6. Silinder kerja ganda ada yang memiliki batang torak (*piston road*) pada satu sisi dan ada kedua pula yang pada kedua sisi [8]. konstruksinya yang mana akan dipilih dengan menyesuaikan kebutuhan.



Gambar 5. Sistem *Pneumatic*



Gambar 6. Ilustrasi Cara Kerja Silinder Kerja Ganda



**Gambar 7. Blok Diagram Elektromekanik AGROBOT-I**

### Sistem Elektronik Robot

Sistem Elektronika merupakan komponen utama sebagai penopang kinerja dari setiap system mekanik yang terdapat pada robot. Selain membutuhkan suplai tegangan listrik arus DC, system penggerak yang terdapat pada robot juga harus bisa dikendalikan sehingga dibutuhkan perangkat pengendali dan jalur komunikasi data, berikut ini gambar skema perangkat elektronik. Pada Gambar 7, setiap sistem penggerak memerlukan sistem kontrol terpusat yang langsung dikendalikan oleh *Arduino mega* dan juga pasokan daya yang berbeda-beda, untuk jalur panah berwarna biru merupakan jalur data keluaran dari *Arduino mega* untuk mengendalikan sistem penggerak, sedangkan jalur panah berwarna jingga adalah jalur pasokan daya untuk masing-masing perangkat elektronik dan sistem penggerak. Pada sistem daya yang bersumber dari baterai akan di turunkan tegangannya sesuai kebutuhan masing-masing perangkat elektronika. Kebutuhan motor DC (*Direct Current*) sebagai penggerak utama roda membutuhkan

tegangan 24 volt yang dipasok melalui *driver* motor.

### HASIL DAN PEMBAHASAN

Pengujian yang dilakukan terhadap sistem mekanik dan elektronik dari *AGROBOT-I* meliputi beberapa komponen mekanik untuk memastikan semua komponen mekanik dan elektronik dapat bekerja dengan baik. Pertama pengujian pergerakan robot, yaitu menguji manuver robot yang menggunakan sistem penggerak differential drive untuk dapat berjalan sehingga dapat di kendalikan dengan mudah. Kedua sisi akurasi dan tingkat keberhasilan robot terhadap perintah yang seharusnya di eksekusi seperti pergerakan motor DC, motor *stepper* z-axis, dan juga pergerakan satu set motor servo *gripper*.

### Pengujian Driver Motor dan Motor DC PG36

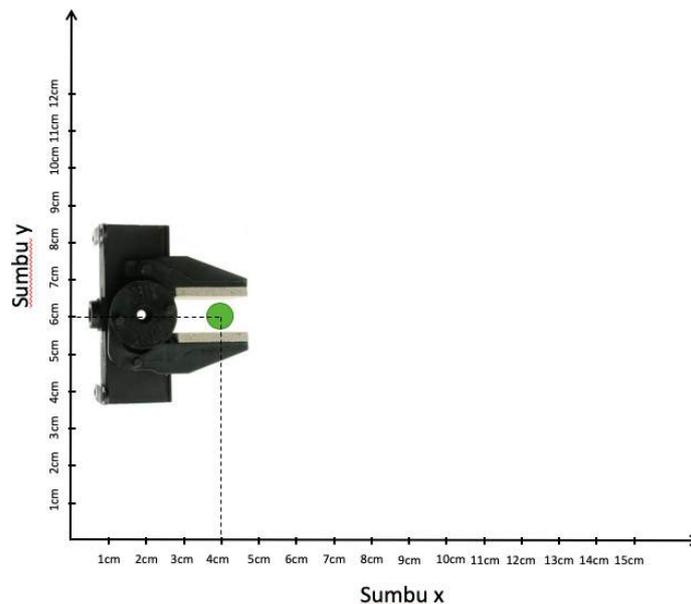
Pengujian *driver* motor dilakukan dengan keluaran motor PG36 24V. Kecepatan

dan arah perputaran motor sudah bisa dikontrol melalui *source code* program pada *Arduino*. Setelah dilakukan pengujian pada kecepatan kedua motor DC dan arah perputaran motor sehingga dapat mengubah orientasi arah robot agar dapat di kendalikan dengan mudah. Pengujian dua buah motor dapat dilihat pada Tabel 1. Dari data pengujian manuver robot jika diberikan nilai PWM yang sama untuk kedua motor DC (*Direct Current*) maka untuk gerakan maju lurus lebih condong ke sebelah kanan, sehingga untuk nilai PWM motor sebelah kanan di beri nilai lebih besar

sedikit agar robot dapat bergerak maju dengan lurus. Ini merupakan salah satu kelemahan menggunakan sistem *differential drive* untuk menggerakkan kedua roda. Pada Gambar 8 menunjukkan posisi awal *gripper* sebelum melakukan pergerakan pada motor stepper untuk mengubah posisi *gripper*. Setiap selesai melakukan penanaman maka posisi *gripper* akan dikembalikan pada posisi awal atau semua untuk memudahkan melakukan tugas berikutnya, sehingga menghasilkan keakuratan dalam mencari posisi yang tepat untuk menurunkan bibit pada area tanam.

**Tabel 1. Hasil Pengujian Manuver Robot**

Roda Kiri		Roda Kanan		Pergerakan Robot		
Logika	PWM	Logika	PWM			
0	1	150	0	1	160	Maju
1	0	150	1	0	160	Mundur
0	1	150	1	0	160	Putar Kanan
1	0	150	0	1	160	Putar Kiri
0	1	150	0	0	0	Serong kanan
0	0	0	0	1	160	Serong Kiri



**Gambar 8. Posisi Awal Gripper**

Sumbu *x* pada *cartesian* untuk mengubah posisi griper maju ke arah depan dan sumbu *y* pada *cartesian* untuk mengubah posisi *gripper* ke kiri dan kanan. Pengujian sistem *cartesian* dilakuakn untuk mendapatkan hitungan yang akurat terhadap penempatan posisi *gripper* ditunjukkan pada Tabel 2. Proses instalasi silinder *pneumatic* dan *solenoid valve* sangat perlu diperhatikan, karena apabila pemasangan selang angin pada silinder *pneumatic* dan *solenoid valve* dengan arah yang terbalik maka pada saat *solenoid valve* aktif piston silinder *pneumatik* tidak akan terdorong, tetapi piston silinder *pneumatik* akan tertarik. Setelah melakukan pengujian *solenoid valve* dengan cara manual, selanjutnya pengujian dengan cara otomatis.

Pengujian ini dilakukan dengan menggunakan modul *relay* dan suplai tegangan 12V. Untuk menghubungkan dan memutuskan tegangan baterai pada *solenoid valve* maka socket pada modul *relay* yang digunakan adalah COM dan NC. Pengujian *solenoid valve* dapat dilihat pada Tabel 3. Dari hasil pengujian di atas peng-gunaan *pneumatic* seluruh proses tanam dan penyiangan membutuhkan 6 kali untuk menggerakkan *gripper* ke bawah yaitu pada saat menanam bibit padi sebanyak 3 bibit dan mencabut gulma sebanyak 3 rumput, jadi untuk tabung penampung tekanan angina cukup di isi udara bertekanan 3 Bar. Setelah 1 putaran proses penanaman dan penyiangan maka tabung penampung akan di isi ulang kembali.

**Tabel 2. Hasil Pengujian Fungsi Pergerakan Cartesian**

Jarak (cm)	Waktu tempuh (detik)	Kecepatan geser (mm/dtk)	Kecepatan putaran motor (rpm)
5	0,30	62,5	75
10	1,6	62,5	75
15	2,4	62,5	75
20	3,2	62,5	75

**Tabel 3. Hasil Pengujian Pneumatik**

Tekanan Tabung (bar)	Piston Maju (kali)	Piston mundur (kali)
2	7	6
2,5	9	9
3	11	11
3,5	13	12

**Tabel 4. Hasil Pengujian Servo**

Pengujian Ke-	data	Kondisi gripper
1	0	menutup
2	75	Buka separuh
3	150	Buka maksimal
4	0	Menutup
5	75	Buka separuh
6	150	Buka maksimal

Pada Tabel 4 dapat dilihat bahwa ada tiga perintah yang digunakan untuk menutup *gripper* dengan mengirimkan data 0 kepada *servo*, untuk membuka separuh dengan mengirimkan data 75, dan untuk membuka maksimal *gripper* dengan data 150, data tersebut didapat dari hasil pengujian *servo* terhadap gripper yang di sesuaikan dengan kinetik pergerakan *gripper*. Untuk penggunaan *gripper servo* di gabung dengan perintah mekanik lainnya seperti *pneumatic* dan sistem *cartesian* untuk melakukan “tanam” pada *control joystick* dalam sekali tekan, sistem akan merespon dan menjalankan perintah sesuai dengan urutan program pada *Arduino* dan akan berhenti setelah semua perintah tanam selesai dikerjakan yang telah ditentukan dalam program. Rata-rata keberhasilan dalam melakukan proses penanaman yaitu 90%. Pengujian tersebut dilakukan dengan berkali-kali melakukan *tuning* pada hardware dan juga *software*, seperti besarnya sudut, posisi end effektor, dan posisi statik lainnya.

## **KESIMPULAN DAN SARAN**

Paper ini mempresentasikan rancangan elektronik dan mekanik sebuah robot prototipe dan sekaligus sebagai robot edukasi di sektor industri pertanian. Rancangan elektronik disajikan dalam bentuk blok diagram yang dilengkapi dengan jalur distribusi sumber daya dan komunikasi untuk di kendalikan dari *Arduino* ke setiap perangkat elektronika sebelum diteruskan ke

sistem penggerak. Perancangan elektronika menyesuaikan fungsi sistem penggerak untuk dapat mengendalikan sistem penggerak. Perancangan system penggerak dengan model Differential drive dapat diandalkan untuk pergerakan robot pada lahan simulasi pertanian.

Setelah melakukan proses pembuatan dan melakukan pengujian ke setiap komponen sistem penggerak terdapat kendala di bagian sistem penggerak utama yaitu motor DC sebagai penggerak roda belakang, dengan menggunakan sistem *differential drive* mengalami sedikit kesulitan untuk robot berjalan lurus karena terdapat perbedaan kecepatan dan presisi dari bagian roda robot, dapat juga disebabkan karena karet ban yang tidak presisi. Untuk penggunaan sistem *pneumatic* bisa bekerja dengan baik tanpa kendala dan bisa menghasilkan gerakan presisi tetapi harus mengisi ulang tekanan udara setiap akan digunakan. Untuk keseluruhan komponen sistem penggerak dapat berfungsi dengan baik sesuai dengan kebutuhan dari pekerjaan robot yaitu menanam, mencabut rumput dan memanen.

Penelitian selanjutnya untuk merancang sistem gerak yang lebih handal di permukaan tanah dan lebih presisi sehingga dapat difungsikan secara otomatis penuh dan menambahkan beberapa sensor navigasi. Perlu tambahan kompresor angin untuk mensuplai tekanan angin pada sistem *pneumatic* dalam operasional dan sistem *cartesian* yang lebih presisi. Komponen

penggerak robot untuk proses panen otomatis juga perlu diperbaiki hingga memperoleh tingkat keberhasilan yang lebih tinggi.

#### DAFTAR PUSTAKA

- [1] Sulaeman, “merdeka.com,” Surya Citra Media, 11 February 2020. [Online]. Available: <https://www.merdeka.com/undang-pengamat-ungkap-penyebab-teknologi-di-industri-pertanian-tertinggal.html>. [Diakses 1 June 2020].
- [2] G.Z Yang, T. Duckett, S. Pearson, S. Blackmore dan B. Grieve, “Agricultural Robotics: The Future of Robotics Agriculture,” UK-RAS Robotics and Autonomous System White papers, ISSN 2398-4414, 2018.
- [3] B. L. Steward, J. Gai dan L. Tang, “The use of agricultural robots in weed management and control,” Agricultural and Bio system Engineering Publication, 2019.
- [4] A. R. Septiadi dan S. Amri, “Rancang Bangun dan Analisa Robot Tematik Simulasi Pertanian dengan Kendali Wireless,” Jurnal Infomedia: Teknik Informatika, Multimedia dan Jaringan, vol. 4, p. 1, 2019.
- [5] A. Z. Arifin dan Subehan, “Simulasi Pengendalian Gerak Robot Mobil Berpenggerak Differensial Dengan Metode  $\square, \square$  Tracking Control Berbasis Proportional Derivative,” Jurnal Teknik POMITS, vol. 1, pp. 1-6, 2012.
- [6] C. Delgado-Mata, R. Valazquez dan C. A. Gutierrez, “A Differential-Drive Mobile Robot Driven by an Ethology Inspired Behaviour Architecture,” The 2012 Iberoamerican Conference on Electronics Engineering and Computer Science, pp. 157-166, 2012.
- [7] Suroso, Sujatno dan R. G. Tambati, “Rancang Bangun Sistem Mekanik Dua Axis Berbasis Kendali Arduino Untuk Peraga Praktikum,” Seminar Nasional XI SDM Teknologi Nuklir, 2015.
- [8] M. Subhan dan A. Satmoko, “Penentuan Dimensi Dan Spesifikasi Silinder Pneumatik Untuk Pergerakan Tote Iradiator Gamma Multiguna Batan,” Jurnal Perangkat Nuklir, vol. 10, p. 2, 2016.
- [9] T. Citradi dan A. Gunawan, “Pak Syahrul, Usul 4 Langkah untuk Mendongkrak Pertanian RI,” CNBC Indonesia, 26 October 2019. [Online]. Available: <https://www.cnbcindonesia.com/news/20191025194550-4-110303/pak-syahrul-usul-4-langkah-untuk-mendongkrak-pertanian-ri/2>. [Diakses 28 April 2020].

# PERAMALAN TINGKAT PENGANGGURAN DI INDONESIA MENGUNAKAN METODE *TIME SERIES* DENGAN MODEL ARIMA DAN HOLT-WINTERS

<sup>1</sup>Agus Sulaiman, <sup>2</sup>Asep Juarna

<sup>1,2</sup>Fakultas Ilmu Komputer Universitas Gunadarma,  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat  
<sup>1</sup>sulaiman.agus12@gmail.com, <sup>2</sup>ajuarna@staff.gunadarma.ac.id

## Abstrak

Beberapa penyebab terjadinya pengangguran di Indonesia ialah, tingkat urbanisasi, tingkat industrialisasi, proporsi angkatan kerja SLTA dan upah minimum provinsi. Faktor-faktor tersebut turut serta mempengaruhi persentase data terkait tingkat pengangguran menjadi sedikit fluktuatif. Berdasarkan pergerakan persentase data tersebut, diperlukan sebuah prediksi untuk mengetahui persentase tingkat pengangguran di masa depan dengan menggunakan konsep peramalan. Pada penelitian ini, peneliti melakukan analisis peramalan *time series* menggunakan metode Box-Jenkins dengan model Autoregressive Integrated Moving Average (ARIMA) dan metode Exponential Smoothing dengan model Holt-Winters. Pada penelitian ini, peramalan dilakukan dengan menggunakan dataset tingkat pengangguran dari tahun 2005 hingga 2019 per 6 bulan antara Februari hingga Agustus. Peneliti akan melihat evaluasi Range Mean Square Error (RMSE) dan Mean Square Error (MSE) terkecil dari setiap model *time series*. Berdasarkan hasil penelitian, ARIMA(0,1,12) menjadi model yang terbaik untuk metode Box-Jenkins sedangkan Holt-Winters dengan  $\alpha(\text{mean}) = 0.3$  dan  $\beta(\text{trend}) = 0.4$  menjadi yang terbaik pada metode Exponential Smoothing. Pemilihan model terbaik dilanjutkan dengan perbandingan nilai akurasi RMSE dan MSE. Pada model ARIMA(0,1,12) nilai RMSE = 1.01 dan MSE = 1.0201, sedangkan model Holt-Winters menghasilkan nilai RMSE = 0.45 dan MSE = 0.2025. Berdasarkan data tersebut terpilih model Holt-Winters sebagai model terbaik untuk peramalan data tingkat pengangguran di Indonesia.

**Kata Kunci:** ARIMA, Holt-Winters, Pengangguran, Peramalan, Time Series

## Abstract

The causes of unemployment in Indonesia are urbanization, industrialization, high school graduate workers, and the minimum wage of province. These causes influenced unemployment rate to fluctuate slightly. Therefore, a forecast is needed to determine percentage of unemployment using the forecasting concept. In this study, the researcher analysed *time series* forecasting using the Box-Jenkins method with the Autoregressive Integrated Moving Average (ARIMA) model and the Exponential Smoothing method with the Holt-Winters model. The forecasting was done using dataset of unemployment rate from 2005 to 2019, every 6 months between February and August. Furthermore, the researcher evaluates smallest value of Range Mean Square Error (RMSE) and Mean Square Error (MSE) for each *time series* model. As results, ARIMA (0,1,12) is best model for the Box-Jenkins, meanwhile best of Holt-Winters with  $\alpha(\text{mean}) = 0.3$  and  $\beta(\text{trend}) = 0.4$  is the Exponential Smoothing method. The best model selection is followed by a comparison of RMSE and MSE accuracy values. For ARIMA (0,1,12), the RMSE value = 1.01 and MSE = 1.0201, whereas the Holt-Winters resulting the RMSE value = 0.45 and MSE = 0.2025. Based on these data, the Holt-Winters was selected as the best model for forecasting unemployment rate in Indonesia.

**Keywords:** ARIMA, Forecasting, Holt-Winters, Time Series, Unemployment

## PENDAHULUAN

Definisi pengangguran menurut *International Conference of Labour Statisticians* (ICLS) adalah orang-orang yang tidak memiliki pekerjaan, yang akhir-akhir ini mencari pekerjaan, dan pada saat ini bisa bekerja. Beberapa penyebab terjadinya pengangguran di Indonesia [1] yaitu, tingkat urbanisasi, tingkat industrialisasi, proporsi angkatan kerja SLTA, upah minimum provinsi dan persentase perubahan pemakaian tenaga kerja sebagai akibat perubahan harga input tenaga kerja/upah, serta output yang akan diproduksi (elastisitas lapangan kerja). Berdasarkan pergerakan persentase data [2], suatu prediksi dapat dibuat untuk mengetahui persentase tingkat pengangguran di masa depan dengan menggunakan konsep peramalan (*forecasting*).

Peramalan deret waktu (*time series*) merupakan metode yang dapat digunakan untuk menghasilkan suatu prediksi informasi di masa yang akan datang berdasarkan riwayat data pada masa sebelumnya. Hasil dari peramalan *time series* ini berguna bagi pihak yang terkait dengan informasi tersebut. Pada penelitian ini, peneliti melakukan analisis peramalan *time series* menggunakan model *Autoregressive Integrated Moving Average* (ARIMA) dan Holt-Winters. Pemilihan kedua model ini didasarkan pada kelebihan masing-masing model, kelebihan dari metode *exponential smoothing Holt-Winters* adalah metode ini sangat baik

meramalkan pola data yang berpengaruh musiman dengan unsur *trend* yang timbul secara bersamaan, metode yang sederhana dan mudah dimasukkan ke dalam praktek dan kompetitif terhadap model peramalan yang lebih rumit. Sedangkan kelebihan ARIMA adalah memiliki sifat fleksibel (mengikuti pola data), tingkat akurasi peramalan cukup tinggi dan cocok digunakan untuk memprediksi dengan cepat, sederhana, akurat dan murah [3]

Beberapa penelitian terdahulu terkait dengan model ARIMA dan Holt-Winters dapat digunakan untuk memprediksi tingkat pengangguran. Penelitian yang dilakukan [4] bertujuan untuk menemukan model yang paling sesuai dalam meramalkan tingkat pengangguran di Sulawesi Selatan dengan menggunakan model ARIMA. Parameter estimasi yang digunakan pada model ARIMA ialah *Sum Squared Residual* dan *Mean Square Error* (MSE). Hasil penelitian menyimpulkan bahwa model (ARIMA)(1,2,1) memiliki nilai MSE terkecil sebesar 2.0474. Penelitian [5] mengidentifikasi faktor-faktor yang mempengaruhi tingkat pengangguran di Malaysia. Peneliti melakukan perbandingan ARIMA dengan Holt-Winters dengan hasil model ARIMA (2,1,2) lebih unggul dibanding model Holt-Winters, hal ini terlihat dari perbandingan MSE antar kedua model yaitu ARIMA (2,1,2) bernilai 0.2623 dan Holt-Winters bernilai 0.3344.

Penelitian [6] menggunakan model SARIMA untuk meramalkan tingkat

pengangguran di USA. Hasil penelitian menyimpulkan bahwa penggabungan model SARIMA (1,1,2) (1,1,1) - GARCH (1,1) dengan estimasi model menggunakan metode optimasi non-linier *Maximum Likelihood*, serta penggunaan algoritma optimasi numerik BFGS memiliki kekuatan prediksi terbaik. Penelitian [7] ini melakukan prediksi terhadap jumlah harian kasus covid-19 yang terkonfirmasi di Sudan dengan menggunakan model ARIMA dan model Holt-Winters. Pada penelitian ini diterapkan tes *Ljung-Box* untuk mengecek *residual* data yang ada. Hasil penelitian menyimpulkan bahwa model ARIMA (2,1,2) direkomendasikan dalam peramalan jumlah kasus harian covid-19 di Sudan dibandingkan dengan model Holt-Winters.

Penelitian [8] melakukan peramalan tingkat pengangguran terhadap negara-negara Eropa terpilih. Dataset diambil dari database Eurostat untuk 5 negara, yaitu Kroasia, Yunani, Italia, Portugal dan Spanyol. Hasil penelitian menyimpulkan bahwa model yang optimal untuk meramalkan tingkat pengangguran di Yunani adalah metode Holt-Winters *additive*, untuk kasus Spanyol model yang optimal menurut *Mean Absolute Percentage Error* (MAPE) adalah *double exponential smoothing*. Selanjutnya untuk Kroasia dan Italia model peramalan terbaik untuk tingkat pengangguran adalah model Holt-Winters *multiplicative*. Terakhir, model peramalan tingkat pengangguran yang cocok untuk negara Portugal adalah *double*

*exponential smoothing*. Pada penelitian [9] dilakukan prediksi terhadap curah hujan di provinsi Nakhon Ratchasima, Thailand dengan menggunakan model ARIMA dan model Holt-Winters, parameter estimasi yang digunakan pada model ARIMA ialah *Sum Squared Residual*. Model ARIMA (1,0,1)(1,0,1)<sub>12</sub> lebih unggul dibanding model Holt-Winters, hal ini terlihat dari perbandingan *Mean Square Error* (MSE), MAE dan MAPE antar kedua model yaitu ARIMA (1,0,1)(1,0,1)<sub>12</sub> memiliki MSE = 775427.300, MAE = 616.918 dan MAPE = 3,23622% sedangkan untuk model Holt-Winters memiliki MSE = 1694700.000, MAE = 943.19 dan MAPE = 11,95938%.

Penelitian [10] bertujuan untuk menganalisis perbandingan peramalan antara model ARIMA dan Holt-Winters terhadap aliran masuk waduk yang ada pada waduk Krishnagiri di negara bagian Tamilnadu, India. Parameter estimasi yang digunakan model ARIMA ialah metode *maximum likelihood estimation* yang diikuti dengan pemeriksaan diagnostik dari perkiraan nilai model. Peneliti menganggap model ARIMA memiliki beberapa keterbatasan dalam memodelkan data, oleh karena itu difokuskan evaluasi akurasi model Holt-Winters antara model *multiplicative seasonal* dan model *additive seasonal*. Berdasarkan hasil evaluasi akurasi didapatkan model *additive seasonal* sebagai yang terbaik dengan nilai *Mean* = 3.002852, *R-Squared* = 0.965939, *Mean (Error)* = 0.005832289 dan *Mean (Percent*

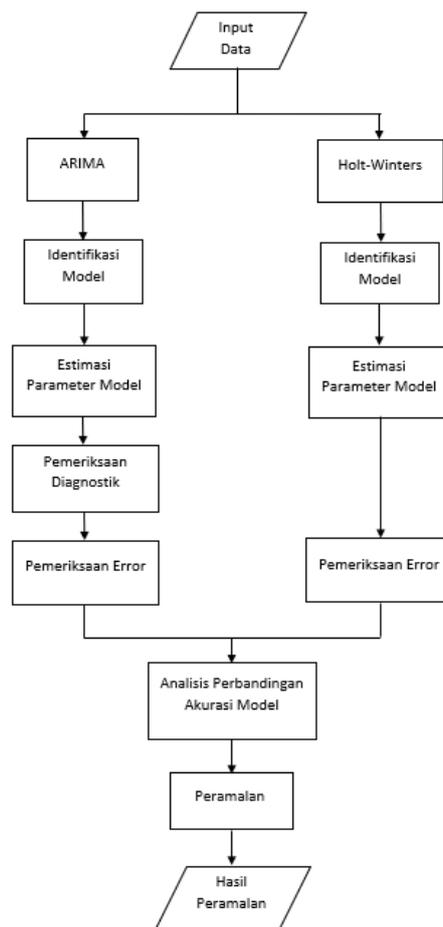
*Error*) = 0.1953426. Penelitian yang dilakukan [11] bertujuan untuk memprediksi kasus positif di beberapa negara terdampak penyebaran COVID-19. Penelitian dilakukan dengan menggunakan metode *time series* dengan beberapa model: ARIMA, *Holt-Winters Additive Models* (HWAAS), *TBAT*, Facebook's Prophet, Deep AR dan N-Beats dengan parameter *Root Mean Squarred Error* (RMSE). Hasil penelitian ini menyimpulkan bahwa model ARIMA dan *TBAT* memiliki nilai peringkat statistik RMSE 1.70000 dan 2.90000 lebih kecil dibandingkan model Prophet, Deep AR dan N-Beats. Penelitian [12] dilakukan untuk memprediksi konsentrasi polusi udara harian di Surabaya. Peramalan dilakukan dengan menggunakan model ARIMA dan Holt-Winters untuk menemukan mana model yang terbaik. Model Holt-Winters direkomendasikan untuk prediksi konsentrasi CO berdasarkan kriteria kecocokan model untuk tiga percobaan yang berbeda dengan berbagai jumlah data yang hilang, yaitu *Mean Error* (ME) (0.039; -0.878; -1106), *Root Mean Square Error* (RMSE) (0,315; 0,985; 1,175), koefisien determinasi ( $R^2$ ) (0,516; 0,612; 0,785) dan korelasi (0,719; 0,782; 0,886). Penelitian [13] dilakukan untuk mengidentifikasi beberapa variabel prediksi seperti permintaan pembangkitan angin, suhu dan kecepatan

angin yang dapat mempengaruhi harga listrik per jam. Penelitian dilakukan dengan menggabungkan model regresi dengan Holt-Winters dan ARIMA. Parameter evaluasi yang digunakan adalah MAPE terkecil. Penelitian ini menyimpulkan bahwa kombinasi antara ARIMA dan Holt-Winters mengungguli metode lainnya dengan akurasi peramalan sebesar 70%.

Berdasarkan permasalahan yang ada, maka penelitian ini akan melakukan peramalan deret waktu (*time series*) dengan model ARIMA dan Holt-Winters untuk memprediksi tingkat pengangguran di Indonesia dengan menggunakan *dataset* tingkat pengangguran dari tahun 2005 hingga 2019 per 6 bulan antara Februari hingga Agustus. Berdasarkan hasil dari analisis kedua model tersebut, akan dilakukan evaluasi *Root Mean Square Error* (RMSE) dan *Mean Square Error* (MSE) terkecil dari setiap model *time series* untuk memilih model terbaik yang dapat digunakan dalam peramalan tingkat pengangguran di Indonesia pada 2 tahun ke depan.

## **METODE PENELITIAN**

Secara garis besar, metode pada penelitian ini terdiri atas beberapa tahapan proses seperti dapat dilihat pada Gambar 1.



**Gambar 1. Metode Penelitian**

**Tabel 1. Sampel Data 6 Bulanan Tingkat Pengangguran di Indonesia**

Bulan	Persentase
Februari 2005	10.26
Agustus 2005	11.24
Februari 2006	10.45
Agustus 2006	10.28
Februari 2007	9.75
Agustus 2007	9.11
Februari 2008	8.46
Agustus 2008	8.39

[Sumber: BPS, 2020]

### Input Data

Pada tahapan ini, data yang digunakan adalah data publik tingkat pengangguran di Indonesia bersumber dari situs Badan Pusat Statistik (BPS) [2] dari tahun 1986 hingga tahun 2019. *Dataset* ini memiliki nilai dalam

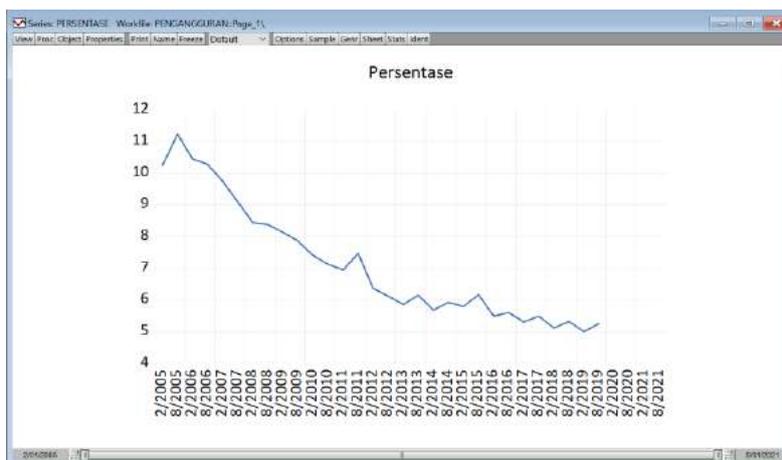
bentuk persentase yang diambil berdasarkan provinsi yang ada di Indonesia. Data diambil secara tahunan dari tahun 1986 hingga tahun 2004, sisanya yaitu mulai dari tahun 2005 hingga tahun 2019 yang diambil secara 6 bulanan (semester) setiap bulan Februari dan

Agustus. Pada penelitian ini, data yang digunakan hanya data dari tahun 2005 hingga tahun 2019 agar keseragaman data dapat tercapai. Berikut contoh sampel data 6 bulanan tingkat pengangguran di Indonesia yang dapat dilihat pada Tabel 1.

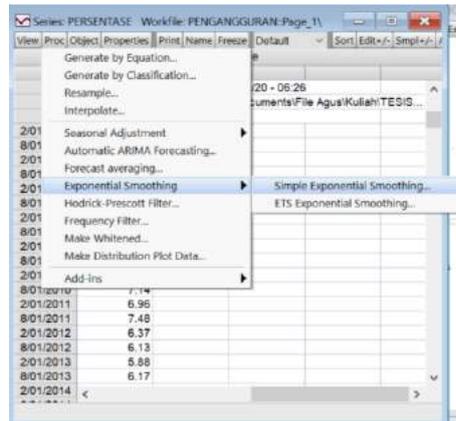
### Identifikasi Model

Pada model ARIMA, tahapan ini dilakukan untuk mendapatkan data yang stasioner dan menghasilkan model terbaik dengan menggunakan *autocorrelation function* (ACF), *partial autocorrelation function* (PACF) dan transformasi (*differencing* dan *logging*). Data deret untuk peramalan menggunakan metode ARIMA harus stasioner terhadap rata-rata dan varian. Gambar 2 merupakan data persentase tingkat pengangguran di Indonesia. Pembentukan model pada ARIMA membutuhkan kondisi

*stasioneritas* data yang harus dipenuhi [14]. Pada penelitian ini, uji *stasioneritas* menggunakan *Unit Root Tests* dengan tipe pengujian *Augmented Dickey-Fuller* (ADF). Parameter yang digunakan pada pengujian ADF adalah *Schwarz Info Criterion* (SIC) dengan nilai probabilitas (*p-value*) lebih kecil dari 0.05 dan nilai statistik (*t-statistic*) lebih kecil dari nilai kritis pengujian (*t-critical values*) 1%-10%. Pengujian terdiri dari 3 pilihan diferensiasi dimulai dari level 0 (Level), level 1 (*1st difference*) dan level 2 (*2nd difference*). Pengecekan dimulai dari level paling awal, yaitu level 0 (Level). Apabila data dianggap sudah *stasioner* maka proses *diferensiasi* dapat dihentikan pada level yang sedang diuji. Pada Holt-Winters, proses identifikasi model dilakukan pada aplikasi EViews menggunakan Menu *Simple Exponential Smoothing* seperti Gambar 3.



Gambar 2. Data Persentase Tingkat Pengangguran



Gambar 3. Menu Simple Exponential Smoothing

Tabel 2. Estimasi Nilai Parameter Alpha dan Beta Data Tingkat Pengangguran

Peramalan	$\alpha$	$\beta$
P1	0.1	0.2
P2	0.2	0.3
P3	0.3	0.4
P4	0.6	0.7
P5	0.7	0.8
P6	0.8	0.9
P7	0.9	0.9

Metode yang digunakan adalah *Holt-Winters Exponential Smoothing - no seasonal* dengan konfigurasi dua parameter yaitu *alpha* dan *beta* seperti pada Gambar 3. Peneliti dapat mengubah nilai parameter Alpha (mean) dan Beta (trend) untuk mendapatkan nilai terbaik dalam proses peramalan data tingkat pengangguran.

#### Estimasi Parameter Model

Estimasi model ARIMA yang telah dihasilkan dapat dilakukan perbandingan dengan beberapa parameter yaitu *Akaike's Information Criterion (AIC)* dan *Schwartz's Bayesian Criterion (SBC)*, *Sum Squared Residual* dan *Adjusted R-Squared*. Penentuan estimasi model terbaik dilihat dari nilai

*Adjusted R-squared* yang besar serta nilai *S.E. of regression*, *Akaike Info Criterion* dan *Schwarz Criterion* yang kecil. Pada model Holt-Winters peneliti menggunakan estimasi beberapa nilai konstanta *alpha* dan *beta* pada setiap objek penelitian seperti pada Tabel 2 menampilkan estimasi nilai parameter *alpha* dan *beta* untuk data tingkat pengangguran dengan 7 kali peramalan dimana P1 adalah estimasi ke-1, P2 adalah estimasi ke-2, P3 adalah estimasi ke-3, P4 adalah estimasi ke-4, P5 adalah estimasi ke-5, P6 adalah estimasi ke-6, P7 adalah estimasi ke-7.

#### Pemeriksaan Diagnostik

Pemeriksaan diagnostik hanya dilakukan pada model ARIMA, tahapan ini

untuk menganalisis nilai probabilitas untuk setiap estimasi model. Tahap ini dilakukan untuk memastikan apakah model yang digunakan sudah baik dengan melihat *residual*. Analisis *residual* dapat dianggap baik jika memiliki *white noise* dengan melihat nilai probabilitas pada ACF dan PACF yang tidak signifikan ( $p\text{-value} > \alpha 0.05$ ).

### Peramalan dan Pemeriksaan Error

Pada model ARIMA, peramalan dan pemeriksaan error merupakan tahapan untuk menganalisis nilai error terkecil dari setiap parameter evaluasi akurasi. Parameter yang digunakan untuk menentukan model ARIMA terbaik dalam penelitian ini adalah nilai *Root Mean Square Error* (RMSE), *Mean Square Error* (MSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) dengan nilai terkecil. Setelah mendapatkan nilai *correlogram residual* data persentase pengangguran pada tahap sebelumnya, proses selanjutnya adalah melakukan *forecasting* dan memeriksa error dari model ARIMA.

Pada model Holt-Winters, peneliti mengimplementasikan metode pada tahap identifikasi model dengan melakukan pemeriksaan error berdasarkan nilai *Sum of Square Residuals* dan *Root Mean Squared Error* (RMSE) terkecil.

### Analisis Perbandingan Evaluasi Akurasi

Tahap analisis ini, dilakukan perbandingan serta evaluasi akurasi dari kedua model. Model pilihan terbaik untuk setiap

metode nantinya akan dibandingkan satu sama lain, dengan pertimbangan nilai akurasi RMSE dan MSE terkecil antara kedua model. *Root Mean Square* RMSE merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat. Sedangkan MSE merupakan nilai kuadrat dari RMSE.

## HASIL DAN PEMBAHASAN

### ARIMA

#### Identifikasi Model

Pada proses identifikasi model ARIMA membutuhkan data deret yang *stasioner*, yaitu data konstan terhadap rata-rata dan varian serta tidak memiliki tren. Dilakukan proses *diferensiasi* atau dalam ARIMA dilambangkan dengan  $d$  dengan uji ADF untuk mencapai data yang *stasioner*. Proses *diferensiasi* dilakukan mulai dari level 0 hingga level 1, berikut hasil pengujian tersebut.

Berdasarkan Gambar 4 (a) dapat dilihat bahwa proses *diferensiasi* level 0 dilambangkan dengan  $d = 0$ , menghasilkan  $p\text{-value } 0.7962 > 0.05$  dan  $t\text{-statistic } 5\% - 1.519849 > -3.595026$  yang berarti data belum *stasioner*. Pada Gambar 4 (b) dilakukan proses *diferensiasi* level 1 ( $d = 1$ ) dengan menghasilkan  $p\text{-value } 0.0030 < 0.05$  dan  $t\text{-statistic } 5\% - 4.887989 > -3.595026$  yang berarti data sudah *stasioner*. Tahapan selanjutnya pada identifikasi model adalah

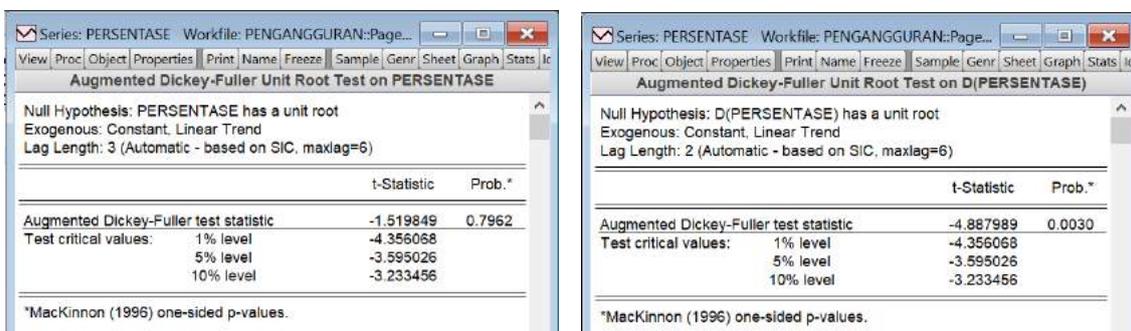
pengujian ACF dan PACF dengan melihat lag yang melewati *flag*.

Pengujian ACF dan PACF dilakukan melalui *correlogram* dengan *diferensiasi* data deret yang sudah *stasioner* yaitu diferensiasi level 1 seperti terlihat pada Gambar 5.

Pada Gambar 5. terlihat bahwa ACF mengalami signifikan di lag-12 setelahnya mengalami *cut off* sedangkan PACF

mengalami *cut off* dari awal hingga akhir lag.

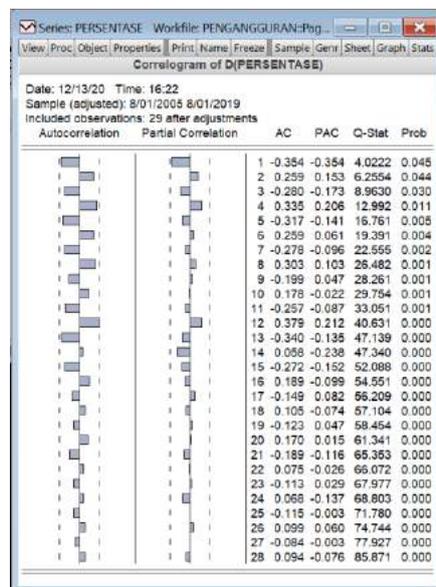
Berdasarkan plot ACF dan PACF tersebut, peneliti melakukan pendugaan model ARIMA dengan nilai MA(12). Oleh karena itu, terdapat 3 model yang sesuai untuk data tingkat pengangguran yaitu ARIMA(0,1,12), ARIMA(1,1,12) dan ARIMA(1,1,0).



(a)

(b)

**Gambar 4. Hasil Uji Stasioneritas (a) Diferensiasi Level 0 (b) Diferensiasi Level 1**

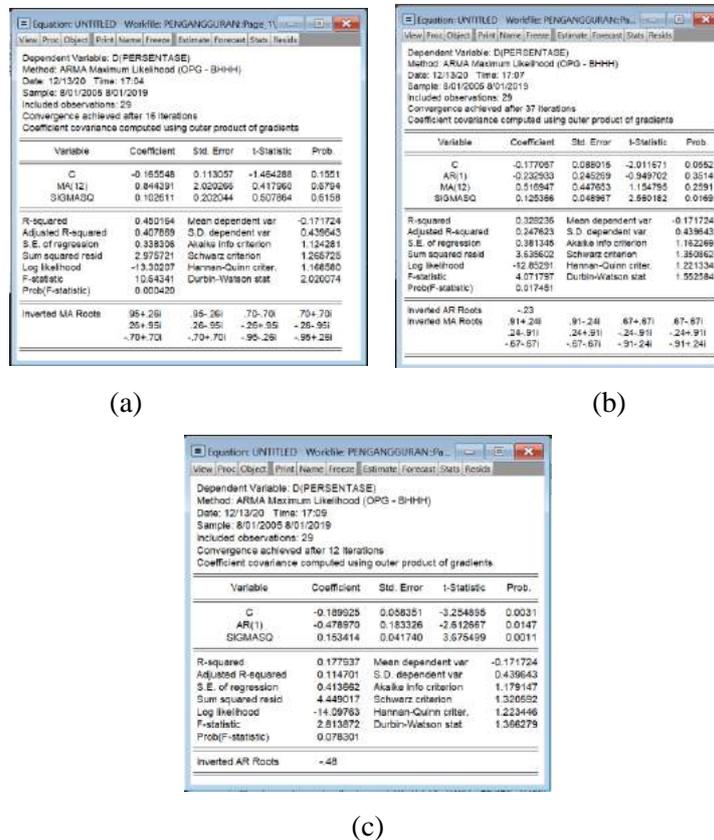


**Gambar 5. Hasil Correlogram Tingkat Pengangguran**

## Estimasi Parameter Model

Tahapan ini menguji ketiga model dugaan yaitu ARIMA(0,1,12), ARIMA(1,1,12) dan ARIMA(1,1,0). Hasil dari pengujian ketiga model dugaan dapat terlihat pada Gambar 6. Hasil parameter uji dari setiap estimasi model ARIMA dapat dilihat pada Tabel 3. Tabel 3 menunjukkan 4 jenis nilai parameter yang dihasilkan dari setiap estimasi model ARIMA data tingkat

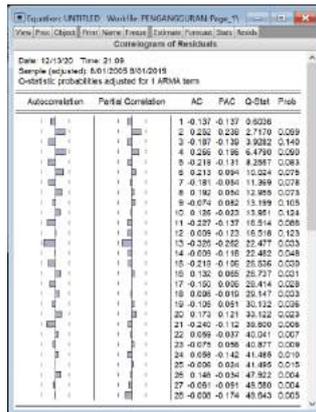
pengangguran. Pada penelitian ini nilai bagus dari parameter *Adjusted R-squared* adalah yang bernilai besar, sedangkan untuk tiga parameter lainnya yaitu, *S.E. of regression*, *Akaike Info Criterion* dan *Schwarz Criterion* nilai paling bagus adalah yang terkecil. Berdasarkan ketentuan tersebut, model ARIMA(0,1,12) dapat digunakan pada proses selanjutnya yaitu, pemeriksaan diagnostik untuk mendapatkan model ARIMA terbaik.



Gambar 6. Estimasi Model (a) ARIMA(0,1,12) (b) ARIMA(1,1,12) (c) ARIMA(1,1,0)

Tabel 3. Hasil Parameter Uji Estimasi ARIMA Data Tingkat Pengangguran

Estimasi Model	Adjusted R-squared	S.E. of regression	Akaike Info Criterion	Schwarz Criterion
ARIMA(0,1,12)	0.407869	0.338306	1.124281	1.265725
ARIMA(1,1,12)	0.247623	0.381345	1.162269	1.350862
ARIMA(1,1,0)	0.114701	0.413662	1.179147	1.320592



Gambar 7. Hasil Correlogram Residual ARIMA(0,1,12)

### Pemeriksaan Diagnostik

Tahap ini memastikan apakah model yang digunakan sudah baik dengan melihat *residual*.

Analisis *residual* dapat dikatakan baik jika memiliki *white noise* dengan melihat nilai probabilitas pada ACF dan PACF yang tidak signifikan ( $p\text{-value} > \alpha 0.05$ ). Pada Gambar 7 tidak ada *lag* yang signifikan pada *correlogram residual* ARIMA(0,1,12). Hal ini menunjukkan tidak ada korelasi antar *residual* dan tidak ada pola yang ditampilkan, artinya ARIMA(0,1,12) sudah *white noise* dan terdistribusi dengan baik.

### Pemeriksaan Error

Tahap pemeriksaan *error* model ARIMA pada penelitian ini menggunakan beberapa parameter evaluasi akurasi, yaitu RMSE, MSE, MAE dan MAPE.

MAE merupakan hasil nilai absolut dari selisih antara nilai output model dengan data sebenarnya, sedangkan MAPE merupakan persentase kesalahan mutlak rata-rata dari peramalan.

### Holt-Winters

#### Identifikasi Model

Berdasarkan *dataset* tingkat pengangguran yang menunjukkan perubahan yang ada di *mean* ( $\alpha$ ) serta *trend* non-musiman ( $\beta$ ) maka metode yang digunakan adalah metode *holt-winter - no seasonal exponential smoothing*.

#### Estimasi Parameter Model

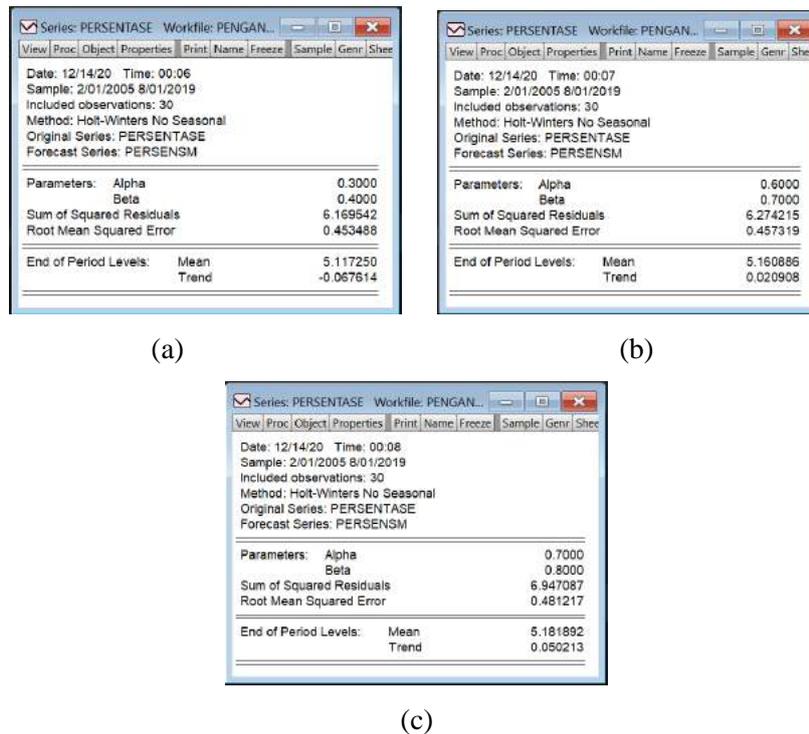
Pengujian estimasi parameter model dilakukan dengan memasukkan kemungkinan angka acak pada kolom  $\alpha$  dan  $\beta$ , dengan ketentuan kedua kolom harus dimasukkan angka lebih dari 0 atau kurang dari 1. Pada Gambar 8 memperlihatkan 3 model terbaik dengan RMSE terkecil. Persentase(0.3, 0.4) dengan nilai  $\alpha = 0.3$  dan  $\beta = 0.4$  menghasilkan RMSE = 0.453488, persentase(0.6, 0.7) dengan nilai  $\alpha = 0.6$  dan  $\beta = 0.7$  menghasilkan RMSE = 0.457319, persentase(0.7, 0.8) dengan nilai  $\alpha = 0.7$  dan  $\beta = 0.8$  menghasilkan RMSE = 0.481217. Hasil beberapa estimasi ditunjukkan pada Tabel 4 menggunakan metode *holt-winter -*

*no seasonal exponential smoothing* untuk menentukan model Holt-Winters yang terbaik.

### Pemeriksaan Error

Berdasarkan hasil RMSE yang sudah dilakukan pada tahap estimasi parameter model dapat disimpulkan bahwa model

persentase(0.3, 0.4) merupakan model terbaik untuk menjadi model Holt-Winters dengan RMSE = 0.45 dan MSE = 0.2025. Model Holt-Winters ini yang akan dibandingkan dengan model ARIMA yang sebelumnya sudah diuji. Hasil dari pemeriksaan error ditunjukkan pada Tabel 4.



Gambar 8. Estimasi Model Holt-Winters (a) persentase(0.3, 0.4) (b) persentase(0.6, 0.7) (c) persentase(0.7, 0.8)

Tabel 4. Hasil Pemeriksaan Error Holt-Winters Data Tingkat Pengangguran

Peramalan	$\alpha$	$\beta$	RMSE	MSE (RMSE <sup>2</sup> )
P1	0.1	0.2	0.66	0,4356
P2	0.2	0.3	0.50	0,25
P3	0.3	0.4	0.45	0,2025
P4	0.6	0.7	0.46	0,2116
P5	0.7	0.8	0.48	0,2304
P6	0.8	0.9	0.53	0,2809
P7	0.9	0.9	0.57	0,3249

Keterangan : P1 adalah estimasi ke-1, P2 adalah estimasi ke-2, P3 adalah estimasi ke-3, P4 adalah estimasi ke-4, P5 adalah estimasi ke-5, P6 adalah estimasi ke-6, P7 adalah estimasi ke-7

### Analisis Perbandingan Model

Pada proses uji model ARIMA telah didapat model ARIMA(0,1,12) dengan RMSE dan MSE terbaik diantara model ARIMA lainnya.

Pada proses uji model Holt-Winters didapatkan model  $\alpha = 0.3$  dan  $\beta = 0.4$  dengan nilai terbaik pada RMSE dan MSE. Tabel 5 menunjukkan nilai perbandingan RMSE dan MSE antara kedua model. Selain pada Tabel 5, hasil peramalan pada masing-masing model dapat dilihat dalam bentuk grafik yang memuat informasi tentang data aktual dan hasil peramalan untuk 2 tahun ke depan. Grafik data tingkat pengangguran aktual serta data peramalan untuk 2 tahun ke depan menggunakan model ARIMA(0,1,12) dan Holt-Winters dengan  $\alpha = 0.3$  dan  $\beta = 0.4$  dapat dilihat seperti pada Gambar 9.

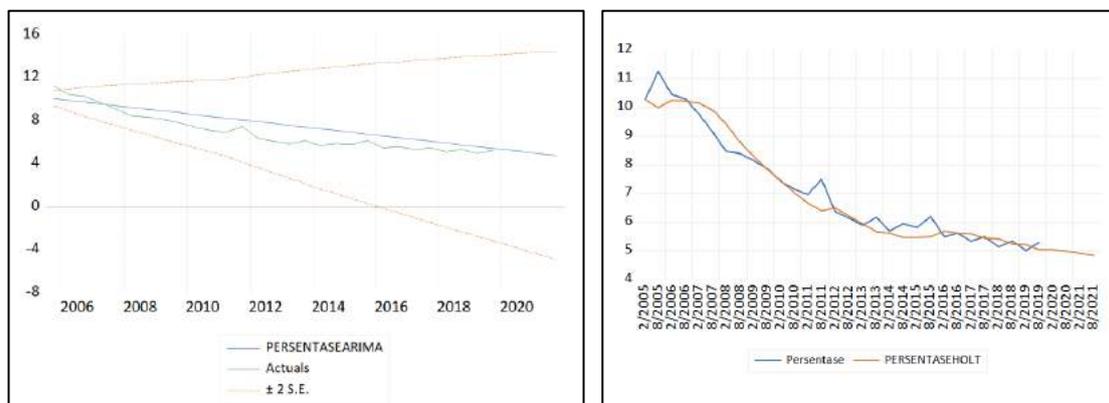
### Hasil Peramalan Tingkat Pengangguran

Proses perbandingan model menemukan model Holt-Winters sebagai model terbaik dalam melakukan prediksi, hal ini dibuktikan dari nilai RMSE dan MSE yang terendah. Oleh karena itu dilakukan peramalan dengan menggunakan model Holt-Winters. Berikut hasil peramalan untuk data tingkat pengangguran di Indonesia yang ditunjukkan pada Tabel 6.

Pada Tabel 6 ditunjukkan hasil peramalan data tingkat pengangguran di Indonesia menggunakan model Holt-Winters dengan  $\alpha = 0.3$  dan  $\beta = 0.4$ . Pada record ke-1 terlihat data pada bulan Februari 2005 memiliki data aktual dan data peramalan secara persentase, untuk nilai aktual memiliki angka sebesar 10.26 dan nilai hasil peramalan sebesar 10.26.

**Tabel 5. Perbandingan RMSE serta MSE Model ARIMA dan Model Holt-Winters**

Model	RMSE	MSE
ARIMA(0,1,12)	1.01	1.0201
Holt-Winters	0.45	0.2025



(a)

(b)

**Gambar 9. Grafik Hasil Peramalan (a) ARIMA(0,1,12) (b) Holt-Winters(0,3,0,4)**

**Tabel 6. Hasil Peramalan Data Tingkat Pengangguran Model Holt-Winters**

Bulan	Aktual	Peramalan
2014 - Februari	5.70	5.62
2014 – Agustus	5.94	5.45
2015 - Februari	5.81	5.47
2015 – Agustus	6.18	5.48
2016 - Februari	5.50	5.69
2016 – Agustus	5.61	5.60
2017 - Februari	5.33	5.58
2017 – Agustus	5.50	5.45
2018 - Februari	5.13	5.41
2018 – Agustus	5.34	5.24
2019 - Februari	5.01	5.20
2019 – Agustus	5.28	5.05
2020 - Februari	-	5.05
2020 – Agustus	-	4.98
2021 - Februari	-	4.91
2021 – Agustus	-	4.85

## KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian dapat ditarik kesimpulan bahwa model Holt-Winters memiliki kesalahan yang kecil dibanding model ARIMA terhadap *dataset* tingkat pengangguran di Indonesia. Hal ini berdasarkan nilai RMSE dan MSE pada kedua model tersebut.

Beberapa saran dari peneliti diharapkan dapat membuat penelitian selanjutnya menjadi lebih baik, antara lain

1. Penggunaan sampel data (*dataset*) disarankan lebih banyak dari yang ada sekarang. Hal ini dapat mempengaruhi tingkat akurasi, karena pada *forecasting* semakin banyak *dataset* maka semakin besar tingkat akurasi yang akan dihasilkan.
2. Perbandingan metode dapat dilakukan selain dengan kedua metode pada penelitian, dengan harapan penggunaan

metode lain dapat meningkatkan efektivitas serta akurasi dalam melakukan *forecasting* data tingkat pengangguran di Indonesia.

3. Penggunaan variabel seperti jenis pekerjaan, cakupan umur dan strata pekerjaan dapat disertakan dalam peramalan pada penelitian selanjutnya.
4. Jika memungkinkan, data pengangguran dapat diambil per bulan agar hasil peramalan tingkat pengangguran di Indonesia dapat cukup representatif dan valid.

## DAFTAR PUSTAKA

- [1] E. A. R. Puspajuita, "Factors that Influence the Rate of Unemployment in Indonesia," *Int. J. Econ. Financ.*, vol. 10, no. 1, p. 140, 2017.
- [2] Badan Pusat Statistik, "Tingkat Pengangguran Terbuka (TPT) Menurut

- Provinsi, 1986 -2019”, Badan Pusat Statistik, 2020. [Online]. Available : <https://www.bps.go.id/dynamictable/2020/02/19/1774/tingkat-pengangguran-terbuka-tpt-menurut-provinsi-1986---2019.html>. [Accessed: Oct. 08, 2020]
- [3] I. R. Akolo, “Perbandingan Exponential Smoothing Holt-Winters Dan Arima Pada Peramalan Produksi Padi Di Provinsi Gorontalo,” *J. Technopreneur*, vol. 7, no. 1, pp. 20–26, 2019.
- [4] D. Didiharyono and M. Syukri, “Forecasting with arima model in anticipating open unemployment rates in south sulawesi,” *Int. J. Sci. Technol. Res.*, vol. 9, no. 3, pp. 3838–3841, 2020.
- [5] S. F. Ramli, M. Fidaus, H. Uzair, M. Khairi, and A. Zharif, “Prediction of the Unemployment Rate in Malaysia,” *Int. J. Mod. Trends Soc. Sci.*, vol. 1, no. 4, pp. 38–44, 2018.
- [6] N. Dritsakis and P. Klazoglou, “Forecasting Unemployment Rates in USA Using Box-Jenkins Methodology,” *Int. J. Econ. Financ. Issues*, vol. 8, no. 1, pp. 9–20, 2018.
- [7] F. Al, R. Idress, and A. Sameeh, “Prediction the daily number of confirmed cases of covid-19 in Sudan with arima and holt winter exponential smoothing PREDICTION THE DAILY NUMBER OF CONFIRMED CASES OF COVID-19 IN SUDAN WITH ARIMA,” no. August, 2020.
- [8] K. Dumi, Č. Č. Anita, and B. Žmuk, “Forecasting Unemployment Rate in Selected European Countries Using Smoothing Methods,” *Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng.*, vol. 9, no. 4, pp. 1073–1078, 2015.
- [9] N. Sopipan, “Forecasting Rainfall in Thailand: A Case Study of Nakhon Ratchasima Province,” vol. 8, no. 11, pp. 712–716, 2014.
- [10] N. Vijayakumar, “A Comparative Analysis of Forecasting Reservoir Inflow using ARMA Model & Holt Winters Exponential Smoothing Technique,” *Int. J. Innov. Sci. Math. Educ.*, vol. 4, no. May 2016, pp. 85–90, 2017.
- [11] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, “COVID-19: A comparison of time series methods to forecast percentage of active cases per population,” *Appl. Sci.*, vol. 10, no. 11, pp. 1–15, 2020.
- [12] A. D. Syafei, N. Ramadhan, J. Hermana, A. Slamet, R. Boedisantoso, and A. F. Assomadi, “Application of exponential smoothing holt winter and ARIMA models for predicting air pollutant concentrations,” *EnvironmentAsia*, vol. 11, no. 3, pp. 251–262, 2018.
- [13] D. Bissing, M. T. Klein, R. A. Chinnathambi, D. F. Selvaraj, and P. Ranganathan, “A Hybrid Regression

Model for Day-Ahead Energy Price Forecasting,” *IEEE Access*, vol. 7, no. March, pp. 36833–36842, 2019.

[14] S. W. Gikungu, “Forecasting Inflation

Rate in Kenya Using SARIMA Model,” *Am. J. Theor. Appl. Stat.*, vol. 4, no. 1, p. 15, 2015.

# EVALUASI TATA KELOLA TEKNOLOGI INFORMASI PADA DINAS TENAGA KERJA KABUPATEN BOGOR MENGGUNAKAN PROCESS ASSESSMENT MODEL (PAM) COBIT 5

<sup>1</sup>Yopie Noor Hantoro, <sup>2</sup>Suryarini Widodo

<sup>1,2</sup>Manajemen Sistem Informasi, Program Pasca Sarjana, Universitas Gunadarma  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat  
<sup>1</sup>yopie\_noor@yahoo.co.id, <sup>2</sup>srini@staff.gunadarma.ac.id

## Abstrak

Dinas Tenaga Kerja Kabupaten Bogor merupakan salah satu organisasi perangkat daerah yang telah menerapkan sistem informasi dalam melayani kebutuhan masyarakat melalui program Bogor Career Center (BCC). Sebuah tata kelola Teknologi Informasi (TI) yang memadai diperlukan untuk memaksimalkan program ini diperlukan sehingga memberikan hasil yang maksimal dan sesuai dengan tujuan organisasi. Penelitian ini bertujuan untuk melakukan evaluasi terhadap tata kelola TI pada Dinas Tenaga Kerja Kabupaten Bogor dan memberikan rekomendasi perbaikan tata kelola. Metode yang digunakan mengacu pada Process Assessment Model (PAM) kerangka kerja COBIT 5. Hasil penelitian menunjukkan bahwa domain tingkat kapabilitas proses TI yang sesuai dengan prioritas organisasi adalah EDM01, EDM02, EDM04, EDM05, DSS01, DSS02, DSS03, DSS04, DSS06 dan MEA01. Hasil pengukuran menunjukkan bahwa domain MEA01 berada pada tingkat kapabilitas 0 (incomplete process) yang artinya proses tidak diimplementasikan atau gagal mencapai tujuan prosesnya. Sedangkan sisanya berada pada tingkat kapabilitas 1 (performed process) yang artinya proses telah diimplementasikan dan mencapai tujuan prosesnya. Sedangkan tingkat kapabilitas yang diharapkan adalah pada level 3 (established process) yang artinya proses memiliki dokumentasi terhadap proses baik pada perencanaan, kebijakan, standar dan dokumen kinerja.

**Kata Kunci:** COBIT 5, Dinas Tenaga Kerja, Process Assessment Model (PAM), Tata Kelola Teknologi Informasi, Tingkat Kapabilitas

## Abstract

The Bogor Regency Manpower Office is one of the regional apparatus organizations that has implemented an information system in serving the needs of the community through the Bogor Career Center (BCC) program. An adequate Information Technology (IT) governance is required to maximize this program in order to provide maximum results and in accordance with organizational goals. This study aims to evaluate IT governance at the Bogor Regency Manpower Office and provide recommendations for improved governance. The method used refers to the Process Assessment Model (PAM) COBIT 5 Framework. The result showed that the IT process capability level domains in accordance with organizational priorities are EDM01, EDM02, EDM04, EDM05, DSS01, DSS02, DSS03, DSS04, DSS06. and MEA01. The measurement results show that the MEA01 domain is at the capability level of 0 (incomplete process), which means that the process is not implemented or fails to achieve its process objectives. While the rest are at capability level 1 (performed process), which means that the process has been implemented and has achieved its process objectives. While the expected level of capability is at level 3 (established process), which means that the process has documentation of the process both in planning, policies, standards and performance documents.

**Keywords:** COBIT 5, Department of Manpower, Process Assessment Model (PAM), Information Technology Governance, Capability Level

## PENDAHULUAN

Pada era revolusi industri 4.0 telah terjadi perkembangan dunia digital yang begitu pesat di mana memberikan pengaruh besar terhadap kehidupan manusia sehingga tidak bisa dipungkiri bahwa penerapan teknologi informasi dapat memberikan peluang kepada organisasi baik yang bersifat komersial maupun bukan komersial untuk meningkatkan kinerja dan produktivitas. Dinas Tenaga Kerja (Disnaker) Kabupaten Bogor merupakan organisasi perangkat daerah yang bertanggungjawab kepada Bupati yang bertugas untuk melaksanakan urusan pemerintah daerah dalam bidang ketenagakerjaan. Salah satu program yang dijalankan oleh Disnaker adalah mengembangkan sistem informasi *Bogor Career Center* (BCC). Namun penerapan sistem informasi ini belum memberikan hasil yang optimal sesuai tujuan dibentuknya program kerja ini misalnya masih terbatasnya informasi lowongan kerja yang disajikan oleh sistem informasi ini. Untuk itu perlu dilakukan evaluasi terhadap tata kelola teknologi informasi untuk mengukur tingkat kapabilitasnya dan memperbaiki tata kelolanya agar sistem informasi dapat memberikan hasil yang optimal sesuai dengan tujuan organisasi.

Tata kelola teknologi informasi (TI) merupakan upaya pengelolaan teknologi informasi agar mendukung dan selaras dengan strategi bisnis suatu perusahaan di

mana tata kelola teknologi informasi berkonsentrasi pada kinerja dan transformasi teknologi untuk memenuhi kebutuhan bisnis saat ini dan saat yang akan datang, baik dari sudut internal maupun eksternal bisnis [1]. Segala sesuatu yang berhubungan dengan penggunaan TI dapat bersinergi dan terukur kinerjanya sesuai dengan tujuan organisasi jika melaksanakan tata kelola TI yang baik. Penerapan TI pada suatu organisasi dapat dilakukan dengan baik apabila ditunjang dengan suatu tata kelola TI mulai dari perencanaan sampai implementasinya [2]. Suatu organisasi dapat menggunakan alat pendukung yang telah berkembang saat ini untuk mengukur kemampuan tata kelola TI, salah satunya adalah kerangka kerja COBIT.

COBIT (*Control Objectives For Information and Related Technology*) merupakan salah satu panduan standar praktik manajemen TI yang disusun agar dapat menghasilkan kerangka dasar dalam mewujudkan TI sesuai dengan kepentingan organisasi dengan tidak mengabaikan faktor-faktor penting lainnya [3]. COBIT menyediakan serangkaian langkah, seperti ukuran, indikator, proses dan praktik terbaik, pada pimpinan, auditor, dan pengguna teknologi informasi dalam memaksimalkan manfaat yang diperoleh melalui penggunaan teknologi informasi dan pengembangan tata kelola teknologi informasi [4].

Pada penelitian ini menggunakan *Process Assessment Model* (PAM) yang ada pada kerangka kerja COBIT 5. Kerangka

kerja COBIT 5 ini mampu mendefinisikan tujuan organisasi yang terkait dengan realisasi manfaat, pengelolaan resiko dan sumber daya menjadi tujuan terkait TI dan proses-proses bisnis yang sesuai dengan tujuan terkait TI tersebut. Beberapa penelitian terdahulu terkait evaluasi tata kelola TI telah dilakukan.

Penelitian [5] melakukan evaluasi terhadap tata kelola TI dengan mengukur tingkat kematangan desain Sistem Informasi / TI pada empat *domain* yaitu APO, EDM, BAI dan DSS. Penentuan *domain* tidak berdasarkan tujuan organisasi sehingga pengukuran dilakukan pada semua *domain* dan tidak fokus pada tujuan organisasi. Penelitian tidak melakukan analisis kesenjangan antara tingkat kematangan yang didapatkan terhadap tingkat kematangan yang diharapkan. Hasil penelitian menunjukkan tingkat kematangan pada *domain* DSS berada pada level 2 (*repeatable but intuitive*), sedangkan *domain* APO, EDM dan BAI berada pada level 3 (*defined proses*). Penelitian [6] melakukan audit terhadap tata kelola TI pada *domain* EDM04, DSS01, APO07 dan APO01. Penentuan *domain* berdasarkan pemetaan antara misi organisasi dan tujuan bisnis berdasarkan COBIT 5. Analisis kesenjangan dilakukan terhadap tingkat kematangan yang didapatkan dan yang diharapkan. Hasil penelitian menunjukkan tingkat kapabilitas proses pada *domain* DSS berada pada level 1 (*performed process*) untuk EDM dan APO berada pada level 2 (*managed proses*), sedangkan level

yang diharapkan adalah 3 untuk *domain* DSS dan 4 untuk *domain* EDM dan APO. Rekomendasi perbaikan diberikan namun tidak menunjukkan bagaimana cara untuk mencapai setiap tingkatan level secara berjenjang. Rekomendasi tersebut tidak menjelaskan bagaimana mencapai level 2, 3 dan seterusnya sampai pada level yang diharapkan. Penelitian [7] melakukan evaluasi tata kelola TI pada *domain* EDM01, EDM04, APO02, APO04, APO06, BAI08 dan DSS03. Penentuan *domain* berdasarkan pemetaan antara visi dan misi organisasi dan disesuaikan dengan relevansi terhadap masalah yang ada. Analisis kesenjangan dilakukan terhadap tingkat kematangan yang didapatkan dan yang diharapkan. Hasil penelitian menunjukkan tingkat kapabilitas proses pada semua *domain* yang terpilih berada pada level 1 (*performed process*). Tingkat kapabilitas yang diharapkan adalah pada level 4 (*predictable process*) sehingga kesenjangan yang ada adalah 3 level. Rekomendasi tidak diberikan untuk mencapai level yang diharapkan.

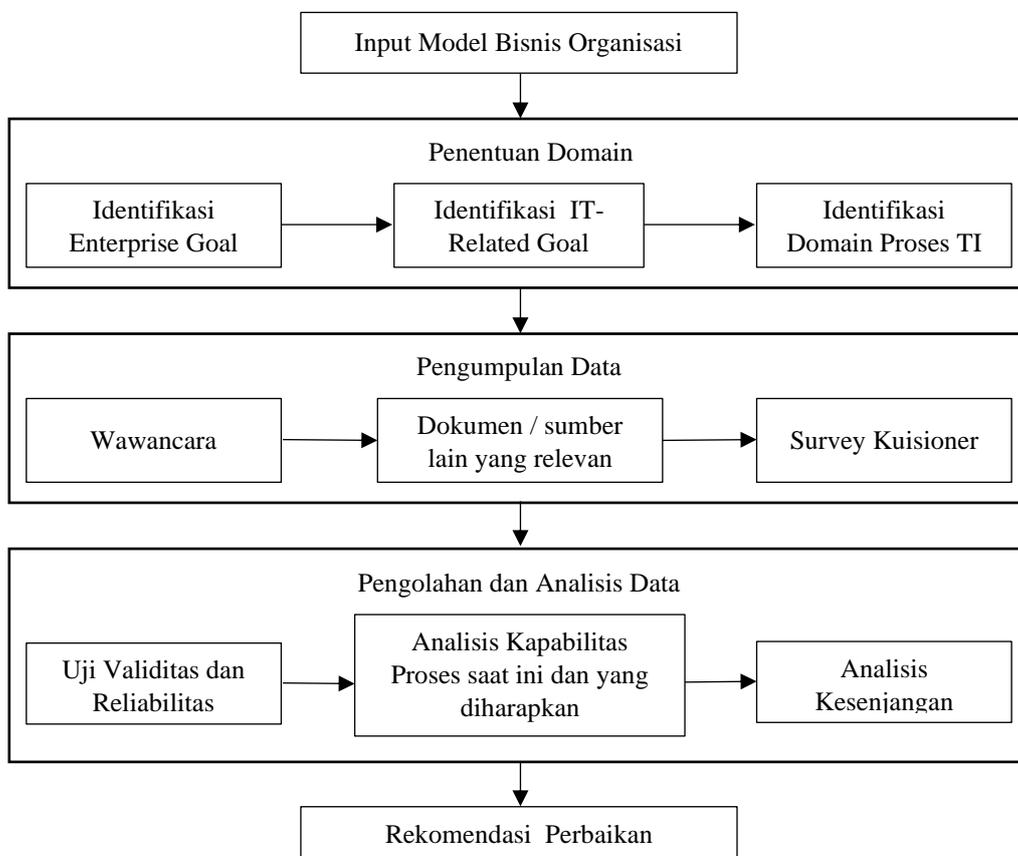
Pada penelitian ini akan melakukan evaluasi tata kelola TI Disnaker Kabupaten Bogor berdasarkan *Process Assessment Model* (PAM) COBIT 5. Penelitian ini melakukan pengukuran tingkat kapabilitas proses terhadap tata kelola yang sudah berjalan dan melakukan Analisis kesenjangan sesuai dengan tingkat kapabilitas yang diharapkan. Penentuan *domain* proses dilakukan melalui *goal cascading*

berdasarkan tujuan organisasi. Hasil yang diharapkan dari penelitian ini adalah mendapatkan nilai kapabilitas proses tata kelola TI yang berjalan pada organisasi saat ini dan memberikan rekomendasi untuk mencapai setiap tingkatan level secara berjenjang hingga mencapai tingkat kapabilitas yang diharapkan. Implementasi sebuah sistem informasi pada organisasi akan memberikan hasil yang maksimal jika tingkat kapabilitas yang diharapkan tercapai.

## METODE PENELITIAN

Penelitian dilakukan pada Disnaker Kabupaten Bogor dalam pengelolaan sistem

informasi *Bogor Career Center* (BCC) melalui beberapa tahapan proses seperti pada Gambar 1. Tahapan awal dari penelitian ini adalah memperoleh masukkan terkait model bisnis organisasi. Data diperoleh melalui penelusuran dokumen yang berlaku pada organisasi seperti dokumen rencana strategis organisasi dan *Standard Operating Procedure* (SOP) serta wawancara terhadap pihak organisasi. Tahap selanjutnya adalah menentukan *domain* yang sesuai dengan model bisnis organisasi. Penentuan *domain* dilakukan dengan beberapa tahap yaitu mengidentifikasi *enterprise goal*, mengidentifikasi *IT-related goal* dan identifikasi domain proses TI.



**Gambar 1. Metode Penelitian**

Tahap Pengumpulan data dilakukan dengan mengumpulkan dokumen-dokumen yang terkait, melakukan wawancara dan menyebarkan kuisisioner kepada responden yang berhubungan dengan pengelolaan sistem informasi BCC. Wawancara dilakukan dengan Kepala Bidang Penempatan dan Perluasan Kerja untuk menggali informasi mengenai prioritas tujuan organisasi, level tata kelola yang diharapkan dan informasi lain yang relevan.

Kuesioner disusun berdasarkan indikator kinerja proses *domain* yang terpilih. Indikator kinerja tersebut diadopsi dari *Process Assessment Model* (PAM) COBIT 5. Kuisisioner kemudian disebar kepada sembilan orang yang terdiri dari seorang Kepala Bidang Penempatan dan Perluasan Kerja, seorang Kepala Seksi Penempatan Tenaga Kerja, seorang Kepala Sub Bagian Program dan Pelaporan, dan 6 orang staff Seksi Penempatan Tenaga Kerja.

Proses pada tahap pengolahan dan analisis data meliputi uji validitas dan reliabilitas terhadap hasil kuisisioner, mengukur tingkat kapabilitas proses yang telah dicapai saat ini dan melakukan analisis kesenjangan terhadap tingkat kapabilitas yang diharapkan oleh organisasi. Proses pengukuran tingkat kapabilitas saat ini dan analisis kesenjangan dilakukan dengan menggunakan metode *Process Assessment Model* (PAM) COBIT 5. Rekomendasi perbaikan disusun berdasarkan analisis kesenjangan yang telah dilakukan

sebelumnya. Rekomendasi diberikan untuk mencapai setiap tingkatan level secara berjenjang hingga mencapai level yang diharapkan. Rekomendasi ini dibuat berdasarkan *Generic Practices* (GPs) dan *Generic Work Products* (GWPs) pada *Process Assessment Model* (PAM) COBIT 5. *Generic Practices* (GPs) merupakan praktek / kegiatan yang diperlukan untuk mencapai atribut pada setiap level kapabilitas proses, sedangkan *Generic Work Products* (GWPs) merupakan hasil kerja / output dari *Generic Practices* (GPs) tersebut.

## HASIL DAN PEMBAHASAN

Hasil dan pembahasan terdiri dari hasil kajian model bisnis organisasi, penentuan *domain* proses yang sesuai, hasil pengolahan dan analisis data serta penyusunan rekomendasi perbaikan.

Penyusunan rekomendasi perbaikan berbeda dengan peneliti sebelumnya dimana rekomendasi diberikan secara berjenjang untuk mencapai setiap tingkat kapabilitas dari yang terendah ke tingkat yang diharapkan. Rekomendasi ini disusun berdasarkan hasil analisa kesenjangan antara tingkat kapabilitas saat ini dengan tingkat kapabilitas yang diharapkan.

### Model Bisnis Organisasi

Model bisnis organisasi diperoleh melalui wawancara dan penelusuran dokumen yang berlaku dalam organisasi. Hasil yang

didapatkan adalah prioritas tujuan organisasi berupa optimasi sumber daya pada internal organisasi.

Prioritas tujuan organisasi ini dipetakan ke dalam *enterprise goal* COBIT 5 dan yang terpilih adalah optimalisasi fungsi proses bisnis. Pemetaan *IT-related goal* yang sesuai dengan *enterprise goal* COBIT 5 dapat dilihat pada Tabel 1.

Pengolahan dan Analisis Data.

Langkah-langkah yang dilakukan pada tahapan pengolahan data ini adalah melakukan uji validitas dan reliabilitas, menghitung tingkat kapabilitas proses yang telah dicapai saat ini dan melakukan analisis kesenjangan terhadap tingkat kapabilitas yang diharapkan oleh organisasi.

**Tabel 1. Enterprise Goal dan IT-Related Goal**

<b>Dimensi TI</b>	<b>Enterprise Goal</b>	<b>IT-Related Goal</b>
Internal	11 Optimalisasi fungsi proses bisnis	1, 7, 8, 9, 12
	12 Optimalisasi biaya proses bisnis	5, 6, 11
	14 Produktivitas operasional dan staf	8, 16

**Tabel 2. Domain Proses Bisnis Terpilih**

<b>Id Domain</b>	<b>Deskripsi Domain</b>
EDM01	Memastikan pengaturan dan pemeliharaan kerangka tata kelola
EDM02	Memastikan pemberian manfaat
EDM04	Memastikan pengoptimalan sumber daya
EDM05	Memastikan transparansi pemangku kepentingan
DSS01	Mengelola operasi
DSS02	Mengelola permintaan layanan dan insiden
DSS03	Mengelola masalah
DSS04	Mengelola kontinuitas
DSS06	Mengelola kontrol proses bisnis
MEA01	Memantau, mengevaluasi dan menilai kinerja dan kesesuaian

**Tabel 3. Hasil Pengujian Validitas**

<b>Item</b>	<b>r hitung</b>	<b>r tabel</b>	<b>Keterangan</b>
EDM01	0,921	0,6664	Valid
EDM02	0,906	0,6664	Valid
EDM04	0,936	0,6664	Valid
EDM05	0,995	0,6664	Valid
DSS01	0,946	0,6664	Valid
DSS02	0,995	0,6664	Valid
DSS03	0,995	0,6664	Valid
DSS04	0,995	0,6664	Valid
DSS06	0,936	0,6664	Valid
MEA01	0,959	0,6664	Valid

## Uji Validitas dan Reliabilitas

Uji validitas dan reliabilitas dilakukan terhadap hasil kuisioner dengan menggunakan perangkat lunak SPSS 22. Pengujian validitas menggunakan teknik korelasi Pearson dengan tingkat signifikansi pada  $r$  tabel sebesar 0,05. Hasil pengujian adalah valid seperti ditunjukkan pada Tabel 3.

Uji reliabilitas menggunakan teknik Alfa Cronbach dan hasil pengujian menunjukkan nilai Alfa Cronbach sebesar 0,985. Hal ini menunjukkan instrumen penelitian adalah reliabel karena hasil lebih besar dari 0,7 sebagai batas minimum. Hasil perhitungan korelasi ( $r$  hitung) pada masing-masing *domain* menunjukkan nilai yang lebih besar dari nilai  $r$  tabel. Nilai  $r$  tabel diperoleh dari tabel korelasi dengan signifikansi 0,05. Hal ini menunjukkan bahwa instrumen kuisioner adalah valid untuk digunakan dalam pengambilan data.

## Hasil Penghitungan Tingkat Kapabilitas

Tingkat kapabilitas ini terdiri dari beberapa tingkatan sesuai dengan kerangka kerja COBIT 5 seperti pada Tabel 4. Untuk mencapai tingkat kapabilitas tertentu, atribut prosesnya harus *Largely achieved* (L) atau *Fully achieved* (F), dan atribut proses pada tingkat di bawahnya mencapai *Fully achieved* (F). Level 1 tercapai jika atribut PA 1.1 terpenuhi dengan kategori L atau F. Level 2 tercapai jika atribut PA 2.1 dan PA 2.2 terpenuhi dengan kategori L atau F. Level 3 tercapai jika atribut PA 3.1 dan PA 3.2 terpenuhi dengan kategori L atau F. Level 4 tercapai jika atribut PA 4.1 dan PA 4.2 terpenuhi dengan kategori L atau F. Level 5 tercapai jika atribut PA 5.1 dan PA 5.2 terpenuhi dengan kategori L atau F. Skala peringkat pencapaian sesuai dengan kerangka kerja COBIT 5 yang mengadopsi dari standar ISO / IEC 15504 seperti pada Tabel 5.

**Tabel 4. Tingkat Kapabilitas dan Atribut Proses**

Tingkat Kapabilitas	ID Atribut Proses	Atribut Proses
Level 0: <i>Incomplete process</i>	-	-
Level 1: <i>Performed process</i>	PA 1.1	<i>Process performance</i>
Level 2: <i>Managed process</i>	PA 2.1	<i>Performance management</i>
	PA 2.2	<i>Work product management</i>
Level 3: <i>Established process</i>	PA 3.1	<i>Process definition</i>
	PA 3.2	<i>Process deployment</i>
Level 4: <i>Predictable process</i>	PA 4.1	<i>Process measurement</i>
	PA 4.2	<i>Process control</i>
Level 5: <i>Optimizing process</i>	PA 5.1	<i>Process innovation</i>
	PA 5.2	<i>Process optimization</i>

[Sumber: ISACA, 2013 [8]]

**Tabel 5. Skala Peringkat**

ID	Deskripsi	Pencapaian (%)
N	<i>Not achieved</i>	0 – 15
P	<i>Partially achieved</i>	>15 – 50
L	<i>Largely achieved</i>	>50 – 85
F	<i>Fully achieved</i>	>85 – 100

[Sumber: ISACA, 2013 [8]]

Pengukuran tingkat kapabilitas dilakukan dengan mengambil data melalui kuisioner.

Responden yang diambil adalah pihak-pihak yang terkait dengan operasional teknologi informasi sehingga mengetahui permasalahan yang ada. Tabel 6 menunjukkan hasil penghitungan tingkat kapabilitas. Dari hasil penghitungan menunjukkan bahwa tingkat kapabilitas tata kelola TI berada pada level 1 (*performed process*) untuk domain EDM dan DSS. Hal ini menunjukkan bahwa Disnaker sebagian besar telah mengimplementasikan dan mencapai tujuan dari proses domain EDM (*Evaluate, Direct and Monitor*) dan DSS (*Deliver, Service and Support*). Tingkat kapabilitas domain MEA berada pada level 0 (*incomplete process*) yang menunjukkan bahwa Disnaker baru belum mengimplementasikan dan mencapai tujuan dari proses domain MEA (*Monitor, Evaluate and Assess*).

Penghitungan pencapaian tingkat kapabilitas menggunakan persamaan (1).

$$\% \text{ Pencapaian} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{z} * 100\% \quad (1)$$

di mana:

x = tingkat kapabilitas yang dipilih (0 – 5)

n = jumlah responden

z = nilai maksimal

### Analisis Kesenjangan

Analisis kesenjangan merupakan analisis untuk menilai kesenjangan antara kondisi saat ini dengan tingkat kapabilitas yang diharapkan. Berdasarkan wawancara dengan Kepala Bidang yang membawahi pengelolaan BCC, didapatkan informasi tingkat kapabilitas yang diharapkan adalah pada level 3 (*established process*). Hasil analisis ini digunakan sebagai dasar untuk memberikan rekomendasi kepada Disnaker Kabupaten Bogor.

**Tabel 6. Hasil Penghitungan Tingkat Kapabilitas**

Id Proses	Atribut									Tingkat Kapabilitas
	PA 1.1	PA 2.1	PA 2.2	PA 3.1	PA 3.2	PA 4.1	PA 4.2	PA 5.1	PA 5.2	
EDM01	L	P	P	N	N	N	N	N	N	1
EDM02	L	P	P	N	N	N	N	N	N	1
EDM04	F	P	P	N	N	N	N	N	N	1
EDM05	L	P	P	N	N	N	N	N	N	1
DSS01	F	P	P	N	N	N	N	N	N	1
DSS02	L	P	P	N	N	N	N	N	N	1
DSS03	L	P	P	N	N	N	N	N	N	1
DSS04	L	N	N	N	N	N	N	N	N	1
DSS06	L	P	P	N	N	N	N	N	N	1
MEA01	P	N	N	N	N	N	N	N	N	0

**Tabel 7. Kesenjangan Tingkat Kapabilitas**

Id Proses	Nama Proses	Tingkat Kapabilitas		Kesenjangan
		Saat ini	Diharapkan	
EDM01	Memastikan pengaturan dan pemeliharaan kerangka tata kelola	1	3	2
EDM02	Memastikan pemberian manfaat	1	3	2
EDM04	Memastikan pengoptimalan sumber daya	1	3	2
EDM05	Memastikan transparansi pemangku kepentingan	1	3	2
DSS01	Mengelola operasi	1	3	2
DSS02	Mengelola permintaan layanan dan insiden	1	3	2
DSS03	Mengelola masalah	1	3	2
DSS04	Mengelola kontinuitas	1	3	2
DSS06	Mengelola kontrol proses bisnis	1	3	2
MEA01	Memantau, mengevaluasi dan menilai kinerja dan kesesuaian	0	3	3

**Tabel 8. Rekomendasi Proses MEA01**

<i>Generic Practices (GPs)</i>	<i>Generic Work Products (GWPs)</i>
Menetapkan sebuah pendekatan pemantauan	<ul style="list-style-type: none"> <li>• Kebutuhan pemantauan</li> <li>• Tujuan dan matrik pemantauan</li> </ul>
Menetapkan target kinerja dan ditinjau ulang secara berkala	Target pemantauan pelayanan
Melakukan pengumpulan dan pemrosesan data kinerja secara akurat dan tepat waktu	Data pemrosesan yang telah diolah
Secara berkala melakukan analisis dan melaporkan kinerja	Laporan Kinerja

Tabel 7 menunjukkan kesenjangan tingkat kapabilitas di mana tingkat kapabilitas yang diharapkan berada pada tingkat 3 dan tingkat kapabilitas saat ini sesuai dengan hasil perhitungan seperti pada Tabel 6.

Hasil menunjukkan bahwa kesenjangan yang paling besar ada pada proses MEA01 yaitu sebesar 3 tingkat, sedangkan kesenjangan proses *domain* yang lainnya sebesar 2 tingkat.

Rekomendasi perlu diberikan untuk mencapai tingkat kapabilitas yang diinginkan dengan meminimalisir kesenjangan yang ada.

### Rekomendasi

Rekomendasi disusun berdasarkan *Generic Practices (GPs)* dan *Generic Work Products (GWPs)* pada *Process Assessment Model (PAM) COBIT 5* untuk mencapai tingkat kapabilitas yang diharapkan. Rekomendasi terdiri dari 3 tahap yaitu meningkatkan level 0 ke level 1 (untuk proses MEA01), menaikkan level 1 ke level 2 (untuk semua proses *domain*) dan menaikkan level 2 ke level 3 (untuk semua proses *domain*). Rekomendasi untuk meningkatkan level 0 ke level 1 adalah menyusun *Generic Practices (GPs)* dan *Generic Work Products (GWPs)*

pada proses MEA01 seperti terlihat pada Tabel 8. *Generic Practices* (GPs) merupakan praktek / kegiatan yang diperlukan untuk mencapai atribut pada setiap level kapabilitas proses MEA01, sedangkan *Generic Work Products* (GWPs) merupakan hasil kerja / output dari *Generic Practices* (GPs) tersebut.

Rekomendasi untuk meningkatkan level 1 ke level 2 adalah melaksanakan manajemen kinerja dan manajemen hasil kerja pada semua *domain* proses. Manajemen kinerja meliputi mengidentifikasi target kinerja dengan output berupa dokumentasi dan rencana proses, merencanakan dan memantau kinerja dengan output berupa rencana proses dan catatan kinerja proses, menyesuaikan kinerja untuk memenuhi perencanaan dengan output berupa catatan kualitas, mendefinisikan, menugaskan dan mengkomunikasikan tanggung jawab dan wewenang dengan output berupa dokumentasi dan rencana proses, mendefinisikan, menyediakan, mengalokasikan dan menggunakan sumber daya dan informasi yang diperlukan untuk menjalankan aktivitas dengan output berupa rencana proses, mengelola pihak yang terlibat dalam pelaksanaan proses untuk memastikan komunikasi yang efektif dan penugasan yang jelas dengan output berupa dokumentasi dan rencana proses. Manajemen hasil kerja meliputi mendefinisikan kebutuhan / persyaratan hasil kerja dengan output berupa rencana kualitas, mendefinisikan kebutuhan untuk dokumentasi dan proses kontrol

terhadap hasil kerja dengan output berupa dokumentasi proses dan rencana kualitas, mengidentifikasi, mendokumentasi dan mengontrol hasil kerja dengan output berupa rencana kualitas, meninjau dan memastikan hasil kerja sesuai dengan perencanaan dengan output berupa catatan kualitas.

Rekomendasi untuk meningkatkan level 2 ke level 3 adalah melaksanakan definisi proses dan penerapan proses pada semua *domain* proses. Definisi proses meliputi menentukan standar proses yang akan mendukung penerapan semua proses *domain* dengan output berupa kebijakan dan standar proses, menentukan urutan dan interaksi antar setiap proses sehingga dapat bekerja sebagai sistem proses yang terintegrasi dengan output berupa kebijakan dan standar proses, mengidentifikasi peran dan kompetensi untuk menjalankan setiap standar proses dengan output berupa kebijakan dan standar proses, mengidentifikasi infrastruktur dan lingkungan kerja yang diperlukan untuk melakukan setiap standar proses dengan output berupa kebijakan dan standar proses, menentukan metode yang sesuai untuk memantau efektivitas dan kesesuaian setiap standar proses dengan output berupa kebijakan dan standar proses. Penerapan proses meliputi menerapkan setiap proses yang telah ditentukan berdasarkan standar proses yang dipilih atau yang disesuaikan dengan output berupa kebijakan dan standar proses, menetapkan dan mengkomunikasikan peran,

tanggung jawab, dan wewenang untuk melakukan setiap proses yang ditentukan dengan output berupa kebijakan dan standar proses, memastikan kompetensi yang diperlukan untuk melakukan setiap proses dengan output berupa dokumentasi dan rencana proses, menyediakan sumber daya dan informasi untuk mendukung kinerja setiap proses dengan output berupa rencana proses, menyediakan infrastruktur setiap proses yang memadai untuk mendukung kinerja proses dengan output berupa rencana proses, mengumpulkan dan menganalisis data terkait kinerja setiap proses untuk menunjukkan kesesuaian dan efektivitasnya dengan output berupa catatan kualitas dan catatan kinerja proses

## KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil berdasarkan hasil evaluasi terhadap tata kelola TI Disnaker Kabupaten Bogor adalah terdapat 10 proses *domain* COBIT 5 yang sesuai dengan prioritas tujuan organisasi yaitu EDM01, EDM02, EDM04, EDM05, DSS01, DSS02, DSS03, DSS04, DSS06 dan MEA01. Hasil evaluasi menunjukkan bahwa satu proses *domain* masih berada pada level 0 (*incomplete process*) yaitu MEA01. Hal ini menunjukkan bahwa aktivitas pemantauan, evaluasi dan penilaian kinerja masih dibawah 50% (*partially achieved / P*) terhadap kriteria COBIT 5. *Domain* EDM01, EDM02, EDM05, DSS02, DSS03, DSS04 dan DSS06

berada pada level 1 (*performed process*) dengan kategori (*largely achieved / L*) yang menunjukkan bahwa aktivitas proses *domain* tersebut telah dijalankan 50%-85% terhadap kriteria COBIT 5. *Domain* EDM04 dan DSS01 berada pada level 1 (*performed process*) dengan kategori (*Fully achieved / F*) yang menunjukkan bahwa aktivitas proses *domain* tersebut telah dijalankan lebih dari 85% terhadap kriteria COBIT 5.

Disnaker Kabupaten Bogor disarankan untuk menjalankan rekomendasi perbaikan yang diberikan agar tata kelola TI sesuai dengan tingkat kapabilitas yang diharapkan. Penelitian selanjutnya disarankan dilaksanakan pada dimensi pelanggan yaitu budaya layanan berorientasi pelanggan dengan tujuan mampu meningkatkan kinerja organisasi terhadap pelayanan kepada masyarakat mengingat bahwa Disnaker Kabupaten Bogor merupakan organisasi perangkat daerah yang berfungsi untuk memberikan pelayanan kepada masyarakat.

## DAFTAR PUSTAKA

- [1] N. Fitri, "Perancangan Tata Kelola Teknologi Informasi Menggunakan Kerangka Kerja COBIT 5 pada PT. Berlian Jasa Terminal Indonesia," MMT. thesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2015.
- [2] P. H. Sinta, I. P. A. Swastika, and I. G. L. A. R. Putra, "Evaluasi Tata Kelola Teknologi Informasi berbasis COBIT 5

- pada Badan Pendapatan Daerah Kabupaten Badung,” *J. Penelit. Tek. Inform.*, vol. 2, no. 2, 2019.
- [3] R. P. Aji and Sarmini, “Penilaian Tata Kelola Teknologi Informasi Pada Dinas Kesehatan Kabupaten Banyumas Mengacu Pada Kerangka Kerja COBIT 5,” *J. Pro Bisnis*, vol. 12, no. 2, pp. 13–24, 2019.
- [4] O. Liandi and Fitria, “Evaluasi Tata Kelola Framework COBIT 5 Pada Dinas Kependudukan Dan Pencatatan Sipil,” *J. Sist. dan Teknol. Inf.*, vol. 5, no. 2, 2019.
- [5] I. N. Putra, A. Hakim, S. H. Pramono, and H. Tolle, “Adopted COBIT-5 Framework for System Design of Indonesia Navy IS/IT : An Evaluation,” *Int. J. Appl. Eng. Res.*, vol. 12, no. 17, pp. 6420–6427, 2017.
- [6] N. L. Adriani, I. M. S. S. Mahardika, and N. W. S. Aryani, “Audit of Certification System Governance Using COBIT 5,” *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 2, 2018.
- [7] S. R. Pasaribu and Harisno, “IT Governance Assessment at Presidential Secretariat using COBIT 5 Framework,” *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 532–537, 2020.
- [8] ISACA, *COBIT 5 : Process Assessment Model (PAM): Using COBIT 5*. USA: ISACA, 2013.

# PREDIKSI RATA-RATA ZAT BERBAHAYA DI DKI JAKARTA BERDASARKAN INDEKS STANDAR PENCEMAR UDARA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY

<sup>1</sup>Anisa Oktaviani, <sup>2</sup>Hustinawati

<sup>1,2</sup>Manajemen Sistem Informasi, Program Pasca Sarjana, Universitas Gunadarma  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

<sup>1</sup>92318030@student.gunadarma.ac.id, <sup>2</sup>hustina@staff.gunadarma.ac.id

## Abstrak

Indonesia menempati peringkat ke-6 dari 98 negara paling berpolusi di dunia pada tahun 2019. Di tahun tersebut, rata-rata AQI (Air Quality Index) sebesar 141 dan rata-rata konsentrasi PM<sub>2.5</sub> sebesar 51.71 µg/m<sup>3</sup> yang lima kali lipat diatas rekomendasi World Health Organization (WHO). Salah satu kota penyumbang polusi udara yaitu Jakarta. Berdasarkan data ISPU (Indeks Standar Pencemar Udara) yang diambil dari SPKU (Stasiun Pemantau Kualitas Udara) Dinas Lingkungan Hidup DKI Jakarta melampirkan pada tahun 2019, Jakarta memiliki kualitas udara sangat tidak sehat. Oleh karena itu perlu adanya model Artificial Intelligence dalam memperdiksi rata-rata tingkat zat berbahaya pada udara di DKI Jakarta. Salah satu algoritma yang dapat diterapkan dalam membuat model prediksi dengan menggunakan data timeseries adalah Long Short-Term Memory (LSTM). Tujuan dari penelitian ini membangun model prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM yang berguna bagi para pemangku kepentingan dibidang lingkungan hidup khususnya mengenai polusi udara. Penelitian mengenai prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM, menghasilkan nilai evaluasi MAPE 12.28%. Berdasarkan hasil evaluasi MAPE yang diperoleh, model LSTM yang digunakan untuk prediksi rata-rata ISPU di DKI Jakarta masuk kedalam kategori akurat.

**Kata Kunci:** DKI Jakarta, Indeks Standar Pencemar Udara, Long Short-Term Memory, Prediksi

## Abstract

Indonesia was ranked 6th out of 98 most polluted countries in the world in 2019. In that year, the average AQI (Air Quality Index) was 141 and the average PM<sub>2.5</sub> concentration was 51.71 µg / m<sup>3</sup> which was five times that above the recommendations of the World Health Organization (WHO). One city that contributes to air pollution is Jakarta. ISPU data (Air Pollutant Standard Index) taken from the SPKU (Air Quality Monitoring Station) of the DKI Jakarta Environmental Service said that in 2019, Jakarta had very unhealthy quality. Therefore it is necessary to have an Artificial Intelligence model in predicting the average level of hazardous substances in the air in DKI Jakarta. One of the algorithms that can be applied in making prediction models using time series data is Long Short-Term Memory (LSTM). The purpose of this study is to build a prediction model for the average ISPU in DKI Jakarta using the LSTM method which is useful for stakeholders in the environmental sector related to air pollution. Research on the prediction of the average ISPU in DKI Jakarta using the LSTM method, resulted in a MAPE evaluation value of 12.28%. Based on the results of the MAPE evaluation obtained, the LSTM model used to predict the average ISPU in DKI Jakarta is in the accurate category.

**Keywords:** DKI Jakarta, Indeks Standar Pencemar Udara, Long Short-Term Memory, Prediction

## PENDAHULUAN

Dewasa ini, bidang ilmu *Artificial Intelligence* (AI) semakin banyak diminati dan terus berkembang. Kecerdasan Buatan atau Artificial Intelligence (AI) adalah teknik yang digunakan untuk meniru kecerdasan yang dimiliki oleh makhluk hidup maupun benda mati untuk menyelesaikan sebuah persoalan[1]. Salah satu fungsi *Artificial Intelligence* yang banyak dimanfaatkan adalah melakukan prediksi. Prediksi adalah proses perhitungan sistematis dalam memperkirakan suatu objek dimasa yang akan datang berdasarkan data masa lalu dan sekarang. Algoritma yang dapat diterapkan dalam membuat model prediksi dengan menggunakan data *timeseries* adalah *Long Short-Term Memory* (LSTM).

LSTM adalah jenis arsitektur pengembangan dari *Recurrent Neural Network* (RNN), sehingga terdapat kelebihan yang dimiliki yaitu dapat menyimpan informasi dalam jangka waktu yang panjang dengan menambahkan *memory cell*. Selain itu, LSTM dapat menjadi solusi saat memproses data sekuensial yang panjang sehingga mengatasi terjadinya *vanishing gradient*. Permasalahan *vanishing gradient* merupakan kekurangan dari RNN dalam menangkap *long term dependencies* yang mengurangi akurasi dari suatu prediksi.

Prediksi yang berhubungan dengan aktifitas kehidupan bermasyarakat sehari-hari adalah prediksi zat berbahaya penyebab

pencemaran udara. Menurut data yang dapat dilihat dari website IQAir, pada tahun 2019 Indonesia menempati peringkat ke-6 dari 98 negara paling berpolusi di dunia [2] dengan rata-rata AQI 141 dan rata-rata konsentrasi PM2.5 51.71  $\mu\text{g}/\text{m}^3$  yang lima kali lipat diatas rekomendasi World Health Organization (WHO). Salah satu kota yang penyumbang polusi udara di Indonesia yaitu Jakarta[2].

Pada tahun 2019, Jakarta menempati peringkat ke-3 dari 10 kota paling berpolusi di Indonesia[3]. Berdasarkan data ISPU (Indeks Standar Pencemar Udara) yang diambil dari SPKU (Stasiun Pemantau Kualitas Udara) Dinas Lingkungan Hidup Jakarta menunjukkan pada tahun 2019, Jakarta memiliki 2 hari (1%) kualitas udara baik, 172 hari (48%) kualitas udara sedang, 183 hari (50%) kualitas udara tidak sehat dan 8 hari (2%) kualitas udara sangat tidak sehat [4].

Perkembangan *Artificial Intelligence* semakin masif diaplikasikan dalam penyelesaian masalah di kehidupan sehari-hari. Salah satunya dengan membangun model prediksi rata-rata ISPU di DKI Jakarta. Prediksi tersebut salah satunya dapat berguna bagi para pemangku kepentingan dibidang lingkungan hidup khususnya mengenai polusi udara. Model prediksi yang dibangun dapat dimplementasikan menggunakan metode *Long Short-Term Memory* (LSTM) karena data *time series* kualitas udara menghasilkan tingkat akurasi yang tinggi dan hasil evaluasi pengujian yang rendah [5].

Penelitian terkait mengenai prediksi terhadap kualitas udara telah banyak dilakukan oleh peneliti terdahulu. Penelitian mengimplementasikan RNN-LSTM dalam melakukan prediksi konsentrasi polutan udara di kota New Delhi dan Agra. Hasil penelitian menunjukkan bahwa metode RNN-LSTM menghasilkan akurasi prediksi yang baik untuk 1 jam dibanding 6 – 12 jam berikutnya [6]. Penelitian menerapkan metode LSTM untuk sistem monitoring real-time memprediksi Influenza Like Illness (ILI) dan kadar zat PM<sub>2.5</sub> di Taiwan. Hasil evaluasi dari model prediksi ILI dan AQI adalah nilai MAPE kurang dari 10% yang berarti nilai akurasi tinggi [5]. Penelitian mengimplementasikan LSTM untuk memprediksi dan meramalkan debit air untuk lima hari kedepan di Sungai Cimandiri dalam membantu masyarakat mengambil keputusan dalam pengendalian operasi irigasi. Hasil dari penelitian menunjukkan model memperoleh hasil RSME kurang dari 10%, sehingga model yang dibuat baik untuk prakiraan jangka pendek [7]. Penelitian mengimplementasikan metode CNN-LSTM dalam memprediksi parameter polutan PM<sub>2.5</sub> di Kota Beijing. Hasil penelitian menggunakan 10 sample menunjukkan bahwa model multivariate CNN-LSTM merupakan model yang memiliki akurasi tertinggi dengan RSME 17.9306 dan MAE 13.9697 [8]. Penelitian mengimplementasikan model RNN-LSTM dalam peramalan polusi udara di Beijing. Hasil penelitian menunjukkan bahwa

algoritma RNN-LSTM menghasilkan akurasi yang setara dengan data asli, tetapi hanya sampai batas waktu tertentu yaitu 5 sampai 6 jam kedepan. Setelah batas waktunya lewat hasilnya akan menurun drastis [9]. Penelitian membandingkan beberapa model yaitu SVR-KERNEL-RBF, SVR-KERNEL-POLY, SVR-KERNEL-LINERA, dan RNN-LSTM untuk memprediksi kualitas udara di kota Visakhapatnam, India. Hasil penelitian menunjukkan metode RNN-LSTM akurasi tinggi dan kinerja lebih baik dalam matrik evaluasi [10]. Penelitian mengimplementasikan kombinasi model RNN, GRU dan LSTM untuk memprediksi polutan PM<sub>2.5</sub> di Distrik Keçiören, Turki dengan kisaran waktu 1, 2, dan 3 jam kedepan. Hasil penelitian terbaik diperoleh untuk prediksi 1 jam dari model terpilih dengan R<sup>2</sup> 0.832, RMSLE 0.404, RMSE 6.272 dan MAE 4.21 [11]. Penelitian membandingkan antara LSTM-RNN dengan ARIMA dalam mempelajari jeda waktu antara paparan konsentrasi tinggi PM<sub>2.5</sub> dan respon kondisi kesehatan tubuh manusia. Selain itu menguji mode lag terdistribusi non-linier (DLNM) untuk mensimulasikan hubungan eksposur-respon dan distribusi waktu yang digunakan untuk mendeteksi efek lag dari PM<sub>2.5</sub> pada kunjungan darurat rumah sakit pada bagian penyakit pernafasan. Hasil menunjukkan bahwa model berbasis LSTM dapat digunakan sebagai metode baru untuk analisis lag yang bermanfaat untuk mendukung kebijakan darurat rumah sakit [12]. Penelitian

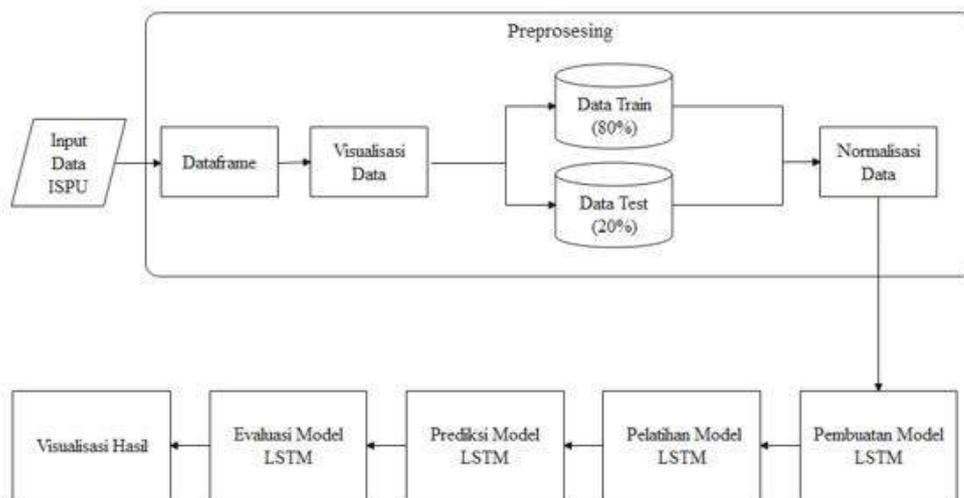
menimplementasikan metode LSTM dan DAE untuk memprediksi konsentrasi PM10 dan PM2.5. Hasil yang didapatkan adalah metode LSTM lebih akurat dibandingkan metode DAE untuk prediksi konsentrasi PM10 dan PM2.5 [13]. Penelitian mengimplementasikan metode LSTM untuk memprediksi kualitas udara di Kota Shanghai, Cina. Evaluasi pada model LSTM menghasilkan RSME 10.95 dan deviasi kuadrat 7.56 [14]. Berdasarkan telaah teoritis, hasil-hasil penelitian terdahulu, maka dapat dikembangkan hipotesis yaitu pengolahan data time series khususnya mengenai kualitas udara menghasilkan tingkat akurasi yang tinggi jika diimplementasi menggunakan metode long short-term memory.

Hipotesis lain dari penelitian ini adalah H\hasil evaluasi yang didapat dari model prediksi ISPU menggunakan metode LSTM menghasilkan nilai MAPE dengan presentasi rendah.

## **METODE PENELITIAN**

Metode penelitian dalam melakukan prediksi rata-rata ISPU (Indeks Standar Pencemaran Udara) di DKI Jakarta menggunakan metode *Long Short-Term*

*Memory* terdiri dari beberapa tahapan-tahapan. Bagan umum metode penelitian ini dapat dilihat pada Gambar 1. Berdasarkan Gambar 1 tahapan awal penelitian adalah memasukkan dataset Indeks Standar Pencemar Udara (ISPU). Dataset ISPU yang didapat dikelola oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta dan dapat di unduh melalui portal resmi data terbuka Indonesia yaitu website Satu Data Indonesia. Dataset ISPU kemudian diproses ke Google Colab untuk memulai *pre-processing*. Tahapan kedua adalah melakukan *pre-processing* pada dataset. *Preprocessing* adalah proses manipulasi dataset ISPU sebelum dimasukkan ke dalam model. Tujuan dari *pre-processing* adalah menyesuaikan data agar lebih kompatibel dengan *library* yang akan digunakan. Pada tahapan *pre-processing* prediksi rata-rata ISPU terdiri dari proses dataframe, visualisasi data, pembagian data, dan normalisasi. Tahapan ketiga adalah proses pembuatan model prediksi rata-rata ISPU di DKI Jakarta. Pembuatan model dengan mengimplementasikan metode LSTM, kemudian menambahkan layer. Penambahan layer disesuaikan dengan jumlah data, agar menghasilkan model dengan nilai akurasi tinggi.



**Gambar 1. Metode Penelitian**

Tahapan keempat adalah melakukan pelatihan model LSTM. Pelatihan model bertujuan untuk melihat nilai loss dari data dan melatih model yang dibangun agar mengenal data-data yang telah dimasukkan. Pada tahapan ini juga ditambahkan beberapa *layer* tambahan seperti *layer epoch* dan *batch size*. *Epoch* menandakan bahwa seluruh dataset telah melalui satu kali putaran proses training pada Neural Network.

*Batch size* adalah sample data yang didistribusikan ke Neural Network. *Layer-layer* tersebut disesuaikan nilainya hingga menghasilkan akurasi tinggi dan nilai loss yang rendah.

Tahapan kelima adalah melakukan prediksi dari model LSTM, dengan membandingkan data asli dengan data hasil prediksi model LSTM. Prediksi ISPU menggunakan data testing yang dihasilkan pada tahap preprocessing. Hasil dari prediksi akan menampilkan data aktual dengan data

hasil prediksi, sehingga dapat terlihat perbedaannya.

Tahapan keenam adalah evaluasi model dengan tujuan untuk mengetahui jumlah error yang dihasilkan dan melihat kelayakan dari model yang dibuat.

Evaluasi model LSTM menggunakan perhitungan nilai Mean Percentage Error (MAPE). Hasil dari perhitungan MAPE yang rendah menandakan bahwa model prediksi menghasilkan nilai yang mendekati nilai aktual atau data real. Tahapan ketujuh yaitu menampilkan visualisasi dari hasil prediksi model tersebut. Visualisasi data adalah teknik mengkomunikasikan data atau informasi dengan menggambarkannya ke dalam objek visual. Hasil dari prediksi rata-rata ISPU DKI Jakarta menggunakan pemodelan LSTM akan di visualisasi dalam bentuk plot diagram. Dari hasil visualisasi tersebut dapat dilihat gambaran visual dari hasil prediksi model dengan nilai aktualnya.

## HASIL DAN PEMBAHASAN

### Memasukkan Data Indeks Standar Pencemar Udara

Data input ISPU (Indeks Stanadar Pencemar Udara) yang digunakan berasal dari data Dinas Lingkungan Hidup DKI Jakarta yang dapat di download dari website Satu Data Indonesia. Data-data tersebut digabungkan sehingga menghasilkan *dataset* dengan jumlah observasi sebanyak 6384 data. *Dataset* tersebut merupakan data ISPU harian dari periode Januari 2017 – Juli 2020. Pada Tabel 1 terdapat rincian dari dataset ISPU.

**Tabel 1. Rincian Dataset ISPU**

Tahun	Jumlah Data
2017	1825
2018	1825
2019	1825
2020	909
<b>Total</b>	<b>6384</b>

Dalam *dataset* tersebut terdapat beberapa variabel yaitu tanggal, wilayah, PM10, SO2, CO, O3, NO2, max, critical, dan kategori. Pada Tabel 2 terdapat penjelasan dari atribut *dataset* ISPU.

### Preprocessing Data ISPU

Tahap preprocessing merupakan tahap awal dalam proses penelitian ini. Tahap preprocessing dilakukan pada *dataset* dengan tujuan agar datasiap untuk dipakai dan sesuai dengan model yang dibangun. Berikut tahapan-tahapan dari *pre-processing* data ISPU:

**Tabel 2. Atribut Dataset ISPU**

Atribut	Penjelasan
Tanggal	Tanggal pengukuran kualitas udara
Wilayah	Lokasi SPKU di DKI Jakarta
PM <sub>10</sub>	Parameter Partikulat yang diukur
SO <sub>2</sub>	Parameter Sulfida (dalam bentuk SO <sub>2</sub> ) yang diukur
CO	Parameter Karbon Monoksida yang diukur
O <sub>3</sub>	Parameter Ozon yang diukur
NO <sub>2</sub>	Parameter Nitrit yang diukur
Max	Nilai ukur parameter paling tinggi dari seluruh parameter yang diukur dalam waktu bersamaan
Critical	Paramater yang memiliki nilai ukur paling tinggi
Kategori	Kategori hasil perhitungan ISPU

### Pembuatan Dataframe

Tahapan pertama pada *preprocessing* yaitu membuat *dataframe* baru yang berisikan atribut tanggal dan rata-rata dari ISPU DKI Jakarta. Atribut tanggal didapat dari *dataset* ISPU, sedangkan atribut rata-rata merupakan variabel baru yang didapat dari rata-rata zat pencemar udara yaitu PM<sub>10</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub> pada setiap SPKU, kemudian hasil rata-rata setiap SPKU di rata-rata kembali sehingga menghasilkan nilai rata-rata dari lima SPKU di DKI Jakarta. Pada Tabel 3 terdapat rincian *df1* yang merupakan *dataframe* yang baru saja dibuat.

### Visualisasi Data

*Dataframe* *df1* selanjutnya divisualisasi dengan tujuan untuk memberi gambaran

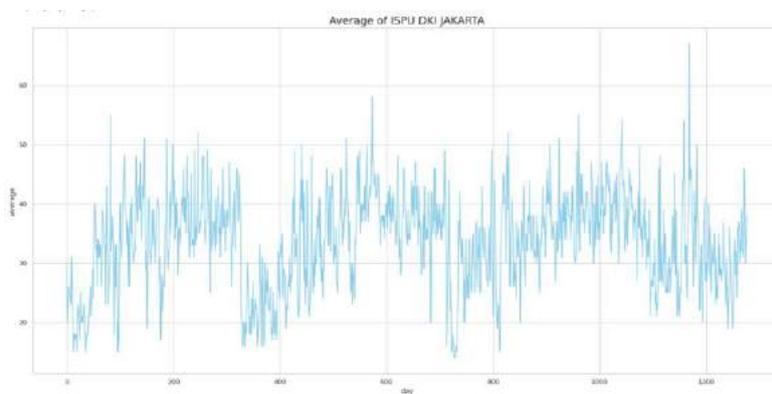
berbentuk grafik. Visualisasi yang ditampilkan yaitu grafik rata-rata ISPU DKI Jakarta dari periode Januari 2017 – Juli 2020. Keluaran yang dihasilkan berupa grafik rata-rata ISPU yang terlampir pada Gambar 2.

### Pembagian Dataset

*Dataframe* *df1* yang telah di visualilasi selanjutnya dibagi menjadi dua bagian yaitu *training* dan *testing*. *Data training* adalah data yang digunakan untuk melatih algoritma, agar mencapai model yang diinginkan. *Data testing* adalah data yang digunakan untuk mengetahui performa dan kebenaran dari algoritma yang telah dilatih dalam model. Pembagian *dataset* diatur dengan persentase 80% untuk *training* dan 20% untuk *testing* yang tertera pada Tabel 4.

**Tabel 3. Dataframe**

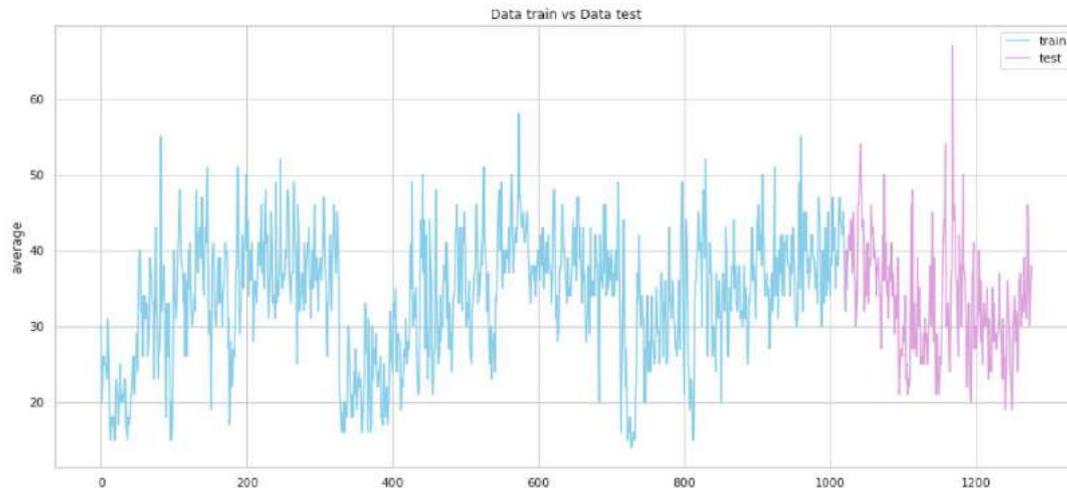
	Tanggal	Average
0	2017-1-1	30
1	2017-1-2	20
2	2017-1-3	22
....	....	....
1274	2020-6-28	32
1276	2020-6-29	36
1277	2020-6-30	38



**Gambar 2. Visualisasi Nilai Rata-rata ISPU DKI Jakarta**

**Tabel 4. Pembagian Dataset**

Dataframe	Data Training (80%)	Data Testing (20%)
Rata-rata ISPU DKI Jakarta	1022	256



**Gambar 3. Visualisasi Pembagian Dataframe**

*Dataframe* df1 yang sebelumnya telah diatur persentase pembagiannya, menghasilkan data yang terdiri dari 1022 untuk *data training* dan 256 untuk *data testing*. Pembagian data yang telah berhasil diimplementasikan, kemudian akan divisualisasi dalam bentuk grafik. Visualisasi dalam bentuk grafik bertujuan untuk mengkomunikasikan data secara jelas dan efisien. Visualisasi *dataframe* df1 terdapat pada Gambar 3.

### Normalisasi Data

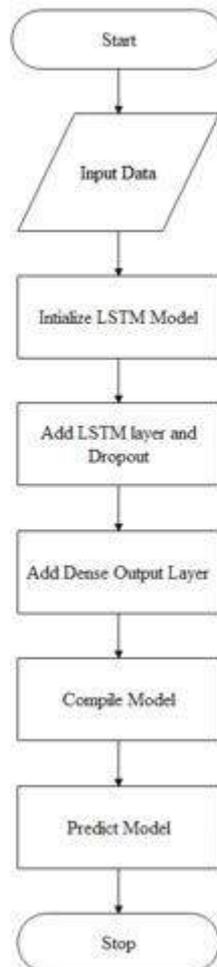
Pembagian dataframe menghasilkan data yang terdiri dari 1022 untuk data training dan 256 untuk data testing. Tahapan selanjutnya menyamaratakan perbedaan

rentang nilai pada numerical data yang dapat menimbulkan masalah dalam model machine learning.

Oleh karena itu, *feature scaling* bertujuan untuk menyamaratakan numerical data pada dataset sehingga memiliki rentang nilai (scale) yang sama. Pada Gambar 4 terdapat hasil dari normalisasi data. Setelah melakukan normalisasi data, tahap selanjutnya yaitu membuat struktur data untuk memvalidasi. Struktur data menghasilkan variabel baru yaitu *X\_train* dan *y\_train* dengan data yang digunakan untuk validasi sebanyak 60 data. Setelah pembuatan struktur data, selanjutnya melakukan perubahan urutan data atau reshape data dari 0 – 1.

```
array([[0.31472081],
       [0.15736041],
       [0.15736041],
       ...,
       [0.36040609],
       [0.34010152],
       [0.37563452]])
```

**Gambar 4. Hasil Normalisasi Data**



**Gambar 5. Alur Model LSTM**

### **Pembuatan Model LSTM**

Pembuatan model merupakan representasi dari bentuk matematis algoritma LSTM. Proses pembuatan model diawali dengan memasukkan data ISPU yang didapat dari website Satu Data Indonesia, data

tersebut telah dilakukan proses pre-processing yang terdiri dari pembuatan dataframe, pembagian dataframe menjadi data testing dan data training, normalisasi data, dan membangun struktur data. Setelah melakukan pre-processing data akan masuk ke lapisan

LSTM, lapisan pada LSTM dapat terdiri dari beberapa layer dengan tujuan untuk menghasilkan model yang akurat. Tahap terakhir adalah lapisan Dense Output yang bertujuan untuk menghasilkan prediksi. Pada Gambar 5 terdapat gambaran dari alur pembuatan model LSTM. Pada tahap awal dalam membuat model LSTM adalah pembuatan class sequential sebagai container agar lapisan LSTM dapat ditambah secara berurutan dan saling terhubung. Model yang dibangun terdiri dari 5 lapisan yaitu 4 lapisan LSTM dan 1 lapisan Dense Output. Lapisan LSTM menggunakan 64 unit neuron dan dropout sebesar 0.2. Jumlah lapisan LSTM tersebut dipilih karena menghasilkan nilai evaluasi paling optimal. Berikut pada tabel 4.1 terdapat rincian dari jumlah layer LSTM dan hasil evaluasi yang diperoleh. Pada Tabel 5 terdapat rincian angka yang digunakan.

Dari tabel 5 dapat diketahui bahwa layer LSTM dengan 64 neuron dan 0.1 dropout menghasilkan nilai MAPE paling optimal. Oleh karena itu, layer tersebut diterapkan pada pembuatan model LSTM prediksi rata-rata indeks pencemar udara.

Lapisan LSTM disertai dengan argumen `return_sequences` bernilai `true` yang dapat diartikan sebagai argumen yang mengeluarkan urutan agar lapisan LSTM berikutnya memiliki input 3D. Lapisan Dense Output memiliki 1 unit neuron. Complie model menggunakan fungsi optimasi RMSprop dengan learning rate 0.001 dan loss yang digunakan adalah mean squared error.

### **Pelatihan Model LSTM**

Model LSTM yang telah dibangun, selanjutnya akan masuk ke tahap pelatihan. Pelatihan model bertujuan untuk mengetahui hasil dari nilai loss dari data. Model pelatihan menggunakan *layer epoch* dan *batch size*. Jumlah *epoch* yang digunakan sebanyak 8 dan jumlah *batch\_size* sebanyak 4.

Implementasi pada tahap pelatihan model LSTM dengan *layer epoch* yang digunakan sebanyak 8 dan jumlah *batch\_size* sebanyak 4 menghasilkan nilai loss sebesar 0.0109. Pada Gambar 6 terdapat tampilan hasil dari pelatihan model.

**Tabel 5. Percobaan Jumlah Layer LSTM**

Neuron	Dropout	MAPE
64	0.1	12.49%
164	0.2	12.28%
128	0.1	12.35%
128	0.2	12.76%

```

Epoch 6/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
Epoch 7/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
Epoch 8/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
<tensorflow.python.keras.callbacks.History at 0x7f343bf366d8>

```

**Gambar 6. Hasil Pelatihan Model**

	Predict	Real
0	[32.28756]	[34]
1	[34.432434]	[31]
2	[32.498215]	[33]
3	[33.75685]	[33]
4	[33.728245]	[26]

**Gambar 7. Hasil Prediksi Model**

**Tabel 6. Tingkat Akurasi MAPE**

Nilai MAPE	Akurasi
< 10%	Sangat Akurat
11% - 20%	Akurat
21% - 50%	Cukup Akurat
>50%	Tidak Akurat

(Sumber : Lewis (1982, p. 40 [16]))

### Prediksi Model LSTM

Tahap yang dilakukan setelah melakukan proses training pada model yaitu prediksi model LSTM. Prediksi dilakukan pada *data testing*. Hasil dari prediksi tersebut dapat dilihat pada Gambar 7.

### Evaluasi Model LSTM

Hasil prediksi yang dihasilkan, selanjutnya akan di evaluasi. Evaluasi model menggunakan LSTM dengan metode *Mean Absolute Percentage Error* (MAPE) dengan

mencari nilai kesalahan absolut disetiap periode dibagi dengan nilai observasi yang aktual pada periode tersebut dan dibuat rata-rata dari *absolute percentage error*. Pada Tabel 6 terdapat penjelasan mengenai nilai MAPE untuk data industri dan bisnis serta interpretasinya[15]. Perhitungan MAPE dapat ditulis dengan rumus seperti persamaan 1.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (1)$$

Keterangan :

$Y_t$  = Nilai aktual pada periode  $t$

$\hat{Y}_t$  = Nilai *forecast* pada periode  $t$

Evaluasi diterapkan pada perbandingan antara data prediksi dengan *data test*. Pada Gambar 8 terdapat nilai evaluasi kesalahan dari prediksi *data training*. Nilai MAPE yang dihasilkan pada prediksi *data training* yaitu 12.28%.

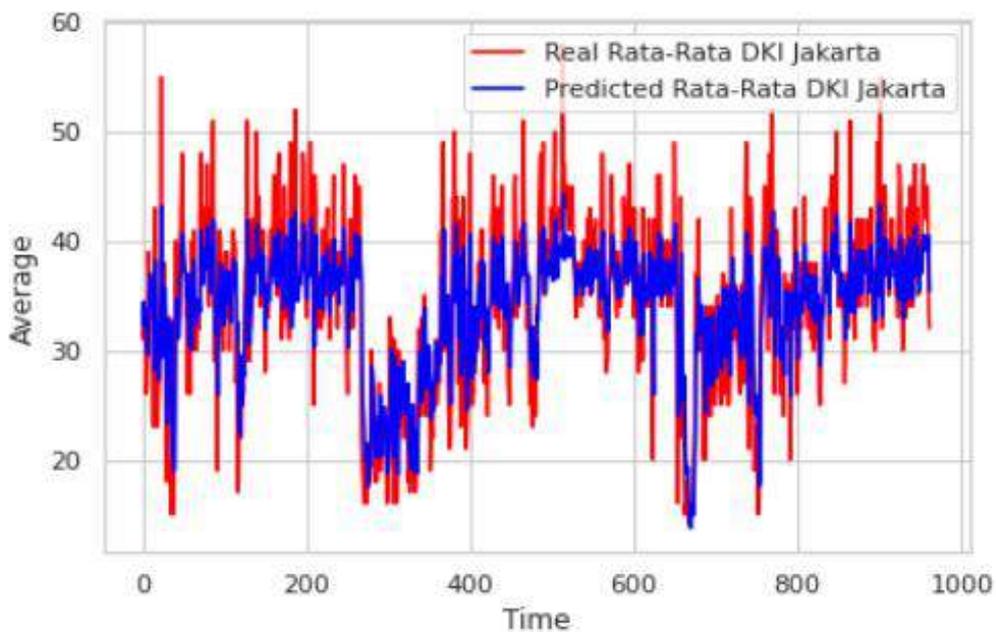
Hasil prediksi tersebut masuk ke dalam kategori MAPE tingkatan akurat. Visualisasi Model LSTM Evaluasi model LSTM yang dibangun mendapat kategori MAPE tingkatan akurat. Tahapan selanjutnya setelah

melakukan evaluasi model yaitu melakukan visualisasi.

Hasil dari prediksi rata-rata ISPU DKI Jakarta menggunakan pemodelan LSTM akan di visualisasi dalam bentuk *plot diagram*. Diagram tersebut membandingkan data rata-rata asli yang diwakilkan garis berwarna merah dengan data rata-rata hasil prediksi dari model yang diwakilkan dengan garis berwarna biru. Visualisasi tersebut dapat dilihat pada Gambar 9.

0.06183527458466975  
MAPE : 12.28%

**Gambar 8. Nilai MAPE pada Prediksi Data Training**



**Gambar 9. Hasil Visualiasi Prediksi Data**

## KESIMPULAN DAN SARAN

Prediksi rata-rata ISPU (Indeks Standar Pencemar Udara) dilakukan pada lima SPKU (Stasiun Pemantau Kualitas Udara) di DKI Jakarta dari periode Januari 2017 – Juni 2020. Data yang digunakan dalam membuat model prediksi adalah dari Dinas Lingkungan Hidup Provinsi DKI Jakarta yang didapat dari website [data.jakarta.go.id](http://data.jakarta.go.id). Prediksi rata-rata ISPU menggunakan metode *long short-term memory*. Pemodelan LSTM menggunakan *layer neuron* sebanyak 64 dan *dropout* 0.2. Hasil evaluasi model prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM dilakukan pada data test, menghasilkan nilai MAPE 12.28% dan masuk kedalam kategori model yang akurat. Berdasarkan hasil penelitian prediksi rata-rata ISPU di DKI Jakarta yang dilakukan maka untuk evaluasi lebih lanjut sebaiknya dilakukan penambahan data ISPU untuk dapat menghasilkan performa model yang semakin baik. Selain itu, hasil penelitian dapat dikembangkan lebih lanjut oleh pihak Dinas Lingkungan Hidup DKI Jakarta dalam meningkatkan informasi yang terdapat dalam website resmi instansi tersebut, sehingga dapat mempermudah pengguna yang ingin melihat informasi mengenai ISPU beserta rata-ratanya di setiap SPKU wilayah DKI Jakarta.

## DAFTAR PUSTAKA

- [1] A. Ahmad, “Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning,” in *Academia*, 2017. [Online] Available: <https://www.academia.edu> [Accessed: Nov 26, 2020].
- [2] IQAir, “World Air Quality,” 2019 *World Air Qual. Rep.*, pp. 1–22, 2019, [Online]. Available: <https://www.iqair.com/world-most-polluted-cities/world-air-quality-report-2019-en.pdf>. [Accessed: Nov 26, 2020].
- [3] AQIAir, “AQIAir Website.” [Online] Available: <https://www.iqair.com>. [Accessed: Nov 26, 2020].
- [4] L. Arumingtyas, “Jakarta Air Quality,” *mongabay.com*, para. 8, Apr. 25, 2020. [Online] <https://www.mongabay.co.id/2020/04/25/polusi-udara-pembunuh-senyap-di-jabodetabek/>. [Accessed: Nov 26, 2020].
- [5] C. T. Yang, L. Y. Lin, Y. T. Tsan, P. Y. Liu, and W. C. Chan, “The Implementation of a Real-time Monitoring and Prediction System of PM2.5 and Influenza-Like Illness Using Deep Learning,” *J. Internet Technol.*, vol. 20, no. 7, pp. 2237–2245, 2019, doi: 10.3966/160792642019122007020. [Online Serial]. Available: <https://jit.ndhu.edu.tw/article/view/2210>. [Accessed: Dec. 1, 2020]
- [6] V. Chaudhary, A. Deshbhratar, V.

- Kumar, and D. Paul, "Time Series Based LSTM Model to Predict Air Pollutant 's Concentration for Prominent Cities in India," *Udm'18*, 2018. [Online Serial]. Available: [http://philippe-fourmier-viger.com/utility\\_mining\\_workshop\\_2018/PAPER1.pdf](http://philippe-fourmier-viger.com/utility_mining_workshop_2018/PAPER1.pdf). [Accessed: Dec. 1, 2020].
- [7] Y. Sudriani, I. Ridwansyah, and H. A Rustini, "Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 299, no. 1, 2019, doi: 10.1088/1755-1315/299/1/012037. [Online Serial]. Available: <https://iopscience.iop.org/article/10.1088/1755-1315/299/1/012037>. [Accessed Dec. 5, 2020].
- [8] T. Li, M. Hua, and X. U. Wu, "A Hybrid CNN-LSTM Model for Forecasting," *IEEE Access*, vol. 8, Feb, 2020. [Online serial]. Available: <https://ieeexplore.ieee.org/abstract/document/8979420>. [Accessed Dec. 2, 2020].
- [9] V. Reddy, P. Yedavalli, S. Mohanty, and U. Nakhat, "Deep Air: Forecasting Air Pollution in Beijing, China," 2017. [Online Serial] Available: <https://www.semanticscholar.org/paper/Deep-Air-%3A-Forecasting-Air-Pollution-in-Beijing-%2C-Reddy-Mohanty/271279939e545be172f7b754b67d6811c11c8fae>. [Accessed Dec. 7, 2020].
- [10] K. S. Rao, G. L. Devi, and N. Ramesh, "Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 2, pp. 18–24, 2019, doi: 10.5815/ijisa.2019.02.03. [Online Serial] Available: <https://www.semanticscholar.org/paper/Air-Quality-Prediction-in-Visakhapatnam-with-LSTM-Rao-Devi/f34d77e3375bd512d14abe3cbacc29e31a2f83aa>. [Accessed Dec. 7, 2020].
- [11] Y. A. Ayturan *et al.*, "Short-term prediction of pm2.5 pollution with deep learning methods," *Glob. Nest J.*, vol. 22, no. 1, pp. 126–131, 2020, doi: 10.30955/gnj.003208. [Online Serial] Available: [https://journal.gnest.org/publication/gnest\\_03208](https://journal.gnest.org/publication/gnest_03208) [Accessed Dec. 10, 2020].
- [12] J. Lu, P. Bu, X. Xia, L. Yao, Z. Zhang, and Y. Tan, "A New Deep Learning Algorithm for Detecting the Lag Effect of Fine Particles on Hospital Emergency Visits for Respiratory Diseases," *IEEE Access*, vol. 8, pp. 145593–145600, 2020, doi: 10.1109/ACCESS.2020.3013543. [Online Serial] Available: <https://ieeexplore.ieee.org/document/91>

- 53761 [Accessed Dec. 10, 2020].
- [13] T. Xayasouk, H. M. Lee, and G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models," *Sustain.*, vol. 12, no. 6, 2020, doi: 10.3390/su12062570. [Online Serial] Available: <https://www.mdpi.com/2071-1050/12/6/2570> [Accessed Dec. 10, 2020].
- [14] Y. Zhu and X. Zhou, "Prediction of Air Quality Index Based on Wavelet Transform Combination Model," *Proc. - 2019 11th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2019*, vol. 1, no. Itaic, pp. 157–160, 2019, doi: 10.1109/IHMSC.2019.00044. [Online Serial] Available: <https://ieeexplore.ieee.org/document/8941558> [Accessed Dec. 13, 2020].
- [15] J. J. Montaña Moreno, A. Palmer Pol, A. Sesé Abad, and B. Cajal Blasco, "El índice R-MAPE como medida resistente del ajuste en la previsioón," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013, doi: 10.7334/psicothema2013.23. [Online Serial] Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=4655890> [Accessed Dec. 13, 2020].
- [16] C. Lewis, *Industrial and Business Forecasting Methods*. London: Butterworth-Heinemann, 1982.

# KLASIFIKASI AREA GEMPA BUMI MENGGUNAKAN ALGORITMA RANDOM FOREST

Ismail

Jurusan Sistem Komputer, Universitas Gunadarma,  
Jl Margonda Raya No 100, Pondok Cina, Depok  
ismail\_muchsin@staff.gunadarma.ac.id

## Abstrak

Salah satu informasi yang dibutuhkan oleh warga negara di dunia adalah informasi mengenai kejadian bencana alam khususnya gempa bumi. Kejadian gempa bumi yang telah terjadi dapat diklasifikasikan dengan menganalisis data gempa bumi pada masa lampau. Pada penelitian ini diimplementasikan penggunaan algoritma Random Forest untuk mengklasifikasikan suatu area apakah termasuk ke dalam kelas gempa bumi atau kelas bukan gempa bumi yang terjadi di dunia. Penelitian ini menghasilkan peta area di seluruh dunia yang terjadi gempa bumi berdasarkan data di masa lampau tahun 1965-2016 dari earthquake dataset kaggle. Penelitian ini menggunakan 7 atribut untuk melakukan klasifikasi antara lain date, time, latitude, longitude, depth, magnitude, dan type. Penelitian ini juga menghitung tabel Confusion Matrix yang dihasilkan dari data aktual dan data prediksi yang telah di proses dalam Random Forest Classifier. Hasil Pengujian Testing Dataset klasifikasian wilayah atau area yang terjadi gempa bumi menghasilkan akurasi sejumlah 99.97%. Hasil penelitian ini diharapkan dapat membantu pihak terkait yang menangani kejadian bencana alam khususnya gempa bumi dengan mengklasifikasikan suatu area termasuk gempa bumi atau bukan gempa bumi berdasarkan atribut yang telah ditentukan.

**Kata Kunci:** Gempa Bumi, Kaggle Dataset, Klasifikasi, Matriks Confusion , Random Forest

## Abstract

One of the information needed by citizens of the world is information about natural disasters, especially earthquakes. Earthquake events that have occurred can be classified by analyzing earthquake data in the past. In this study, the use of the Random Forest algorithm was implemented to classify an area whether it belongs to an earthquake class or a non-earthquake class that occurs in the world. This research produces a map of areas around the world where earthquakes occurred based on data in the past 1965-2016 from earthquake dataset cluster. This study uses 7 attributes to classify including date, time, latitude, longitude, depth, magnitude, and type. This study also calculates the Confusion Matrix table generated from actual data and predictive data that has been processed in the Random Forest Classifier. The Testing Process Results The classification dataset for areas or areas where an earthquake occurred resulted in an accuracy of 99.97%. The results of this study are expected to help related parties who handle natural disasters, especially earthquakes by classifying an area including earthquakes or non-earthquakes based on predetermined attributes.

**Keywords:** Earthquake, Kaggle Dataset, Classification, Confusion Matrix , Random Forest

## PENDAHULUAN

Informasi mengenai gempa bumi yang tersaji saat ini masih bersifat acak [1], sulit

dipahami, dan belum terbukti keabsahannya.

Gempa bumi merupakan suatu kejadian yang

tidak bisa dihindari, namun dampak bencana

alam ini dapat dikurangi atau dapat

diminimalisir dengan mengenali penyebab gempa bumi dan mempelajari kejadian gempa bumi yang telah terjadi dengan menganalisis data yang ada. Pengolahan data bencana alam gempa bumi yang umum dilakukan untuk menghasilkan sebuah informasi [2] yaitu menggunakan teknik *data mining* [3]. Salah satu teknik dari *data mining* adalah klasifikasi [4][5]. Klasifikasi digunakan untuk menemukan model fungsi [6] dan mendeskripsikan data ke kelas-kelas berdasarkan data di masa lampau. Data yang telah dikumpulkan akan dipelajari dan dianalisis hubungannya sesuai dengan label atau target yang telah ditentukan. Beberapa penelitian terkait implementasi algoritma klasifikasi *random forest* dilakukan peneliti terdahulu. Penelitian [7] membuat sistem pengklasifikasian untuk penilaian kredit dengan menggunakan *dataset German Credit*. Metode yang diterapkan pada sistem tersebut adalah *Random Over-under sampling Random Forest* yang dapat meningkatkan kinerja akurasi sebesar 14,1% dengan nilai akurasi sebesar 0,901 atau 90,1%. Penelitian [8] membuat suatu sistem yang dapat mendiagnosis kanker payudara dengan menggunakan 2 metode, yaitu *Support Vector Machine (SVM)* dan *Random Forest*. Hasil dari klasifikasi tersebut dibandingkan seberapa akurat hasil dari akurasi. Penelitian [9] membuat suatu model pengklasifikasian untuk memprediksi curah hujan dengan menggunakan metode *Random Forest* yang menghasilkan akurasi sebesar

99,45%. Penelitian [10] membuat suatu model pengklasifikasian terhadap faktor-faktor yang mempengaruhi tingkat penerimaan konsumen terhadap mobil menggunakan metode *Random Forest*. Penelitian [11] membuat suatu sistem yang dapat memprediksi waktu memperbaiki *bug* dari laporan *bug* dengan menggunakan praproses penyaringan *dataset* dan algoritma *Random Forest* untuk pembangunan pendekatan prediksi. Pada penelitian ini, diimplementasikan penggunaan algoritma *Random Forest* untuk mengklasifikasikan suatu area apakah termasuk ke dalam kelas gempa bumi atau kelas bukan gempa bumi yang terjadi di dunia. Penelitian ini juga menghasilkan peta area di seluruh dunia yang terjadi gempa bumi berdasarkan data di masa lampau tahun 1965-2016 dari *earthquake dataset kaggle*. Penelitian ini menghitung *Confusion Matrix* yang dihasilkan dari data aktual dan data prediksi yang telah di proses dalam *Random Forest Classifier*. Hasil penelitian ini diharapkan dapat membantu lembaga-lembaga yang menaungi kejadian bencana alam khususnya gempa bumi yang ada di dunia dalam mengklasifikasikan suatu area termasuk gempa bumi atau bukan gempa bumi berdasarkan parameter yang telah ditentukan.

## **METODE PENELITIAN**

Tahapan proses pada penelitian ini dimulai dengan data *input* dari data gempa

bumi pada tahun 1965-2016 di dunia, melakukan pemilihan parameter, mengkonversi waktu, memvisualisasikan peta area yang terjadi gempa bumi, melakukan pembentukan *classifier* yang terdiri dari *training set* dan *testing set*, menghitung akurasi dan tahap akhir melakukan pemodelan klasifikasi

### Data Gempa Bumi

Dalam penelitian ini akan menggunakan data gempa bumi pada Tabel 1 yang terjadi pada tahun 1965-2016 di dunia [12]. Data penelitian terdiri dari 21 *field*, yaitu *Date*, *Time*, *Latitude*, *Longitude*, *Type*, *Depth*, *Depth Error*, *Depth Seismic Stations*, *Magnitude*, *Magnitude Type*, *Magnitude Error*, *Magnitude Seismic Stations*, *Azimuthal Gap*, *Horizontal Distance*, *Horizontal Error*, *Root Mean Square*, *ID*, *Source*, *Location Source*, *Magnitude Source*, dan *Status*, serta jumlah *record* sebanyak 23412. Pada Tabel 1 baris yang diberi warna kuning adalah data sampel gempa bumi tahun 1965-2016 di dunia pada tanggal 5 Januari 1965 pada pukul 18.05.58 dengan *latitude* sebesar -20.579 derajat, *longitude* sebesar -173.972 derajat dengan *type Earthquake* pada kedalaman (*depth*) sebesar 20 km dan *depth error* tidak ada. Atribut penelitian ini menggunakan 7 *field* antara lain *date*, *time*, *latitude*, *longitude*, *depth*, *magnitude*, dan *type* dari total dataset awal sejumlah 21 *field*. Parameter *type* ini digunakan sebagai variabel target atau

*labeling* pada proses pengklasifikasian gempa bumi.

### Tahap *Preprocessing* Klasifikasi Gempa

Tahap awal (*preprocessing*) dalam melakukan klasifikasi terdiri atas beberapa tahapan proses antara lain :

1. Melakukan konversi waktu. Konversi waktu yang digunakan pada penelitian ini adalah *unix time* atau *unix epoch time*. *Unix time* atau *unix epoch time* adalah sistem untuk menggambarkan suatu titik waktu yang merupakan jumlah detik yang telah berlalu sejak 00:00:00 Kamis, 1 Januari 1970, *Universal Time Coordinated* (UTC). Setiap hari diperlakukan seolah-olah mengandung tepat 86400 detik. Pada sistem di mana waktu Unix disimpan sebagai bilangan bulat 32 bit yang telah ditandatangani, nilai terbesar yang dapat direkam adalah 2147483647 (231 - 1), yaitu 03:14:07 Selasa, 19 Januari 2038 UTC. Detik berikutnya, jam akan membungkus ke negatif 2147483648 (-231), yaitu 20:45:52 Jumat, 13 Desember 1901 UTC
2. Melakukan proses visualisasi Peta Area Gempa. Setiap wilayah memiliki titik koordinatnya masing-masing yang terdapat pada peta dunia. Titik koordinat tersebut dilewati dengan 2 garis yang saling berkesinambungan, yaitu garis lintang yang disebut dengan *latitude* dan garis bujur yang disebut dengan *longitude*. Peneliti menggunakan garis lintang (*latitude*) dan garis bujur (*longitude*) untuk melakukan

proses visualisasi area gempa pada peta dunia.

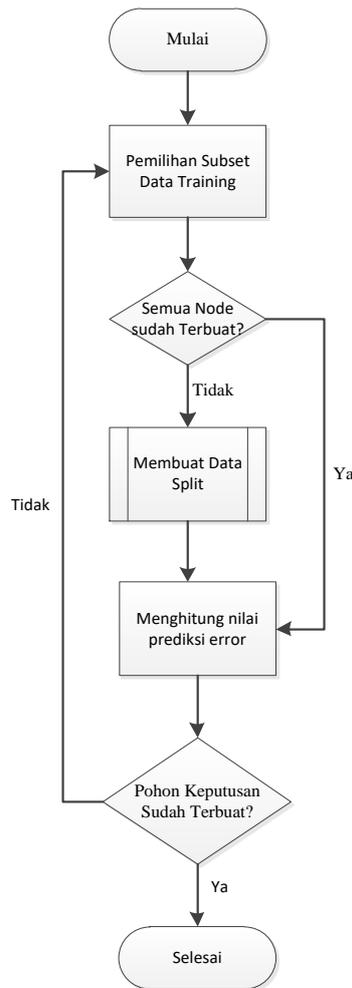
yang dijadikan sebagai parameter dalam proses pengklasifikasian dan variabel target (*label*) dengan *base learner*, yaitu *decision tree* seperti dapat dilihat pada Gambar 1.

**Tahap Pembentukan Classifier**

Tahapan dari pembentukan *Classifier* adalah menentukan variabel dari data *input*

**Tabel 1. Contoh Lima Sampel Data Gempa Bumi Tahun 1965-2016 di Dunia[12]**

Date	Time	Latitude	Longitude	Type	Depth	Depth Error
01/02/1965	13:44:18	19.246	145.616	Earthquake	131.06	-
01/04/1965	11:29:49	1.863	127.352	Earthquake	80	-
01/05/1965	18:05:58	-20.579	-173.972	Earthquake	20	-
01/08/1965	18:49:43	-59.076	-23.557	Earthquake	15	-
01/09/1965	13:32:50	11.938	126.427	Earthquake	15	-



**Gambar 1. Alur Pembentukan Classifier dengan Random Forest**

Tahap pembentukan *Classifier* penelitian ini terdiri dari :

1. Pemilihan *subset data training* untuk membentuk *node-node* pada *tree*. Jika variabel subset tersebut belum terbuat, maka harus terlebih dahulu melakukan *sampling data* dengan teknik *bootstrapping* dengan melakukan pengacakan pada dataset kemudian menghitung nilai *Gini Index* dan selanjutnya mengurutkan variabel subset tersebut dan memilih *split* terbaik. Pada pembentukan *Random Forest* menggunakan nilai *Gini Index* untuk menentukan pemilah (*split*) yang dijadikan *root/node* berdasarkan tingkat kehomogenan nilai peubah respon (variabel dependen). Setiap *node* akan memilah dan membentuk *node* baru lagi sehingga membentuk *tree* sampai semua variabel yang ada terpenuhi. *Gini index (S)* dapat dihitung menggunakan persamaan 1.

$$\text{Gini}(S) = 1 - \sum_{i=1}^k P_i^2 \quad (1)$$

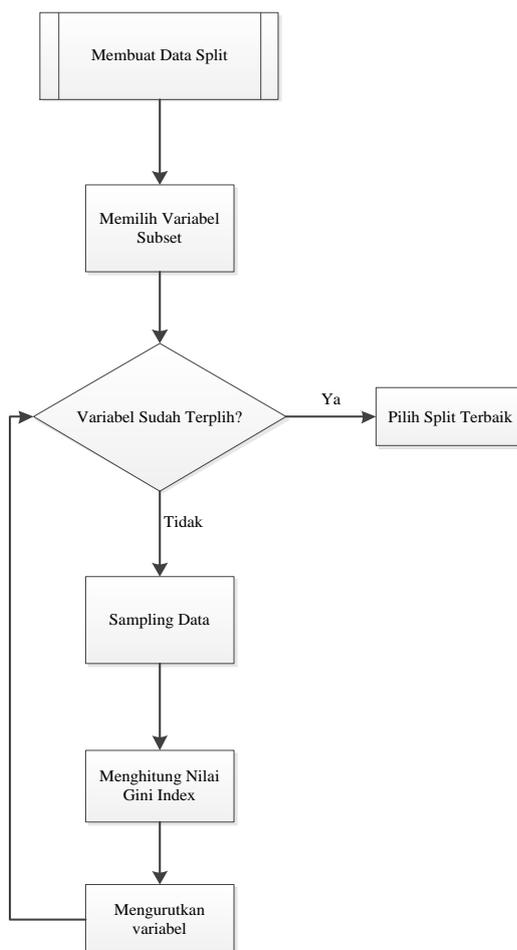
Nilai  $P_i$  adalah probabilitas dari *Gini Index (S)* yang dimiliki kelas  $I$  dan  $k$  adalah banyaknya nilai atribut yang

termasuk ke dalam suatu kelas berdasarkan atribut data. Hasil nilai *Gini index (S)* dapat dilihat pada Tabel 2. Tabel 2 merupakan tabel dari hasil perhitungan nilai *Gini Index (S)* dimana *feature Depth* memiliki nilai *Gini Index (S)* sebesar 0.758921, *feature Latitude* memiliki nilai *Gini Index (S)* sebesar 0.159995, *feature Longitude* memiliki nilai *Gini Index (S)* sebesar 0.080344, dan *feature Magnitude* memiliki nilai *Gini Index (S)* sebesar 0.000740. *Feature name* yang dijadikan *root* adalah *feature Depth*.

2. Membuat *split* antara *data training* dan *data testing*. Pada tahap *split data* ini *data training* dipilah dan dipisahkan dengan *data testing*. Pencarian *node* akan terus dilakukan sampai semua *node* pada setiap *tree* terbentuk. Jika semua *node* pada *tree* telah terbuat kemudian dihitung nilai *error prediction*. Pohon-pohon keputusan ini kemudian dikombinasikan setiap pohon atau kelasnya untuk dilakukan *voting*. Pemilihan *voting* terbaik akan dijadikan sebagai hasil dari klasifikasi. Gambar 2 merupakan *flowchart* dalam pembuatan *data split*.

**Tabel 2. Hasil Nilai Gini Index (S)**

<i>Feature Name</i>	Nilai <i>Gini Index (S)</i>
Depth	0.758921
Latitude	0.159995
Longitude	0.080344
Magnitude	0.000740



**Gambar 2. Flowchart Pembagian Data *Training* dan Data *Testing***

Tahapan pembentukan classifier menggunakan *random forest* dilakukan dengan tahapan :

### 1. Pembentukan *Training Set*

Proses pertama yang dilakukan oleh *Classifier* adalah fase pembelajaran (*learning*), dimana algoritma *Random Forest Classifier* ini dibuat untuk menganalisa data *training* sebagai data masukan lalu direpresentasikan dalam bentuk model klasifikasi dan data *test* yang digunakan untuk memperkirakan akurasi dari model klasifikasi. Data latih yang digunakan pada penelitian ini adalah

18727 atau sekitar 80% dari *dataset* yang ada. Data latih ini digunakan untuk melatih model klasifikasi menghasilkan kemampuan generalisasi yang dapat dipercaya dan tingkat kesalahan yang kecil.

### 2. Pembentukan *Testing Set*

Proses kedua yang dilakukan oleh *Classifier* setelah membagi data latih adalah fase pengujian atau mengevaluasi *output* yang dihasilkan dari algoritma *Random Forest Classifier* untuk mendapatkan hasil akurasi dari proses pengklasifikasian. Data *testing* yang

digunakan pada penelitian ini adalah 4682 atau sekitar 20% dari *dataset* yang ada. Data *testing* ini digunakan untuk membandingkan nilai aktual dan nilai prediksi dalam perhitungan tingkat akurasi.

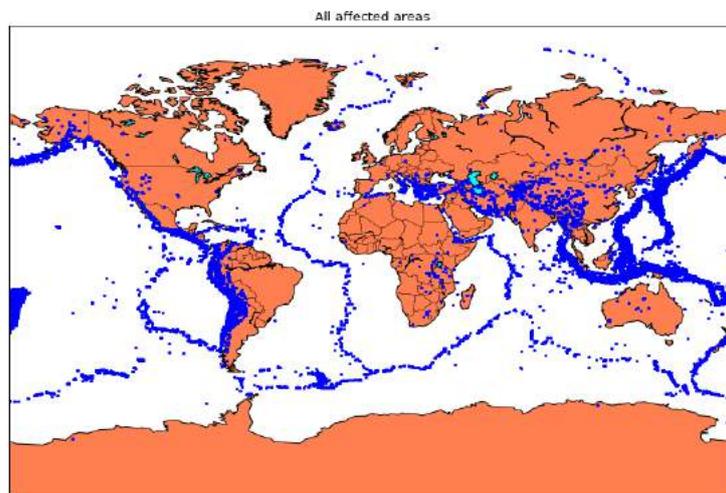
## HASIL DAN PEMBAHASAN

Gambar 3 menampilkan hasil dari visualisasi area gempa bumi di dunia. Daerah yang berwarna biru merupakan daerah yang terjadi gempa bumi, daerah yang berwarna merah bata (coral) merupakan daratan, daerah yang berwarna biru muda merupakan danau,

dan daerah yang berwarna putih merupakan laut.

### Hasil Pengujian Sistem dengan *Confusion Matrix*

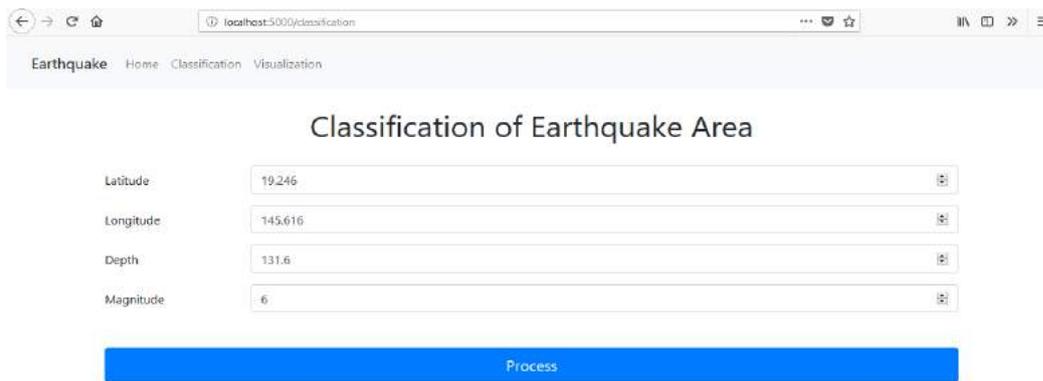
Pada Tabel 3 menampilkan hasil dari pengujian dengan data tes yang berjumlah 5 record. Jenis prediksi (*Predict Type*) memiliki 2 nilai, yaitu 0 atau 1 dimana mendeskripsikan jika jenis prediksi memiliki nilai 0 maka pada *field Description* akan menampilkan “Gempa Bumi” dan sebaliknya jika jenis prediksi memiliki nilai 1 maka pada *field Description* akan menampilkan “Bukan Gempa Bumi”.



Gambar 3. Hasil Visualisasi Area Gempa

Tabel 3. Hasil Pengujian dengan *Confusion Matrix*

	Latitude	Longitude	Depth	Magnitude	Predict Type	Description
0	19.246000	145.616000	131.6	6.00	1	Gempa Bumi
1	37.302167	-116.408333	1.2	5.62	0	Bukan Gempa Bumi
2	37.295333	-116.455667	1.2	5.63	0	Bukan Gempa Bumi
3	37.231500	-116.473667	1.4	5.52	0	Bukan Gempa Bumi
4	7.420000	106.030000	10.0	5.20	1	Gempa Bumi



**Gambar 4. Hasil Pengujian *Testing Dataset***

### Hasil Pengujian *Testing Dataset*

Pengujian dilakukan dengan memasukkan data-data yang ada pada data tes. Data yang dimasukkan antara lain *latitude*, *longitude*, *depth*, dan *magnitude*. Pada pengujian pertama menggunakan *latitude* sebesar 19.246, *longitude* sebesar 145.616, *depth* sebesar 131.6, dan *magnitude* sebesar 6 seperti dapat dilihat pada Gambar 4.

menghasilkan nilai persentase akurasi sebesar 99.97%.

Hasil dari penelitian ini diharapkan dapat membantu lembaga-lembaga yang menaungi kejadian bencana alam khususnya gempa bumi. Pengembangan lebih lanjut dapat dilakukan untuk penelitian selanjutnya dengan penambahan atribut dalam melakukan klasifikasi sehingga akurasi klasifikasi dapat meningkat.

### KESIMPULAN DAN SARAN

Implementasi algoritma *Random Forest* pada pengklasifikasian terhadap area atau wilayah yang terjadi gempa bumi di dunia dengan menggunakan data lampau pada tahun 1965-2016 telah berhasil dilakukan. Pemvisualisasian peta area yang terjadi gempa bumi di dunia dari data lampau pada tahun 1965-2016 berhasil dilakukan dalam proses pengklasifikasian dengan menggunakan algoritma *Random Forest*.

Perhitungan *Confusion Matrix* telah berhasil dihitung pada proses pengklasifikasian dengan algoritma *Random Forest*

### DAFTAR PUSTAKA

- [1] G. Otari and R. Kulkarni, "A review of application of data mining in earthquake prediction," *International Journal of Computer Science and Information Technologies*, vol.3, no.2, pp.3570-3574, 2012.
- [2] A.S.N. Alarifi, N.S.N. Alarifi, S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern Red Sea area", *Journal of King Saud University-Science*, 24(4), 301-313, 2012.

- [3] Han, J., dan Kamber, M, “*Data Mining Concept and Tehniques*”, San Fransisco: Morgan Kauffman, 2016.
- [4] Iswari, N.M.S, “*Penggunaan Teknik Data Mining untuk Manajemen Resiko Sistem Informasi Rumah Sakit*”, ULTIMATICS. Vol. 3, No. 2, pp. 16–22, 2015.
- [5] E. Buulolo, N. Silalahi, Fadlina and R. Rahim, "C4.5 Algorithm To Predict the Impact of the", *International Journal of Engineering Research & Technology (IJERT)*, vol. 6, no. 2, pp.10-15, 2017.
- [6] S. Mangalathu, H.V. Burton, "Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions", *International Journal of Disaster Risk Reduction*, IJDRR 101111, 6 March 2019.
- [7] A. Syukron & A.Subekti, "Penerapan Metode Random Over-Under Sampling dan Random Forest untuk Klasifikasi Penilaian Kredit", *Jurnal Informatika*, 5(2), 175–185, 2018.
- [8] H. Aliady, N.J. Tuasikal, & E. Widodo, “Implementasi Support Vector Machine (SVM) dan Random Forest”, *Sentika*, 23–24, 2018.
- [9] A. Primajaya & B.N. Sari, “Random Forest Algorithm for Prediction of Precipitation”, *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIM)*, 1(1), 27–31, 2018.
- [10] Y.S. Nugroho dan N. Emiliyawati, “Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest”, *Jurnal Teknik Elektro*, 9(1), 24–29, 2016.
- [11] N.F. Azhar, & Rochimah, “Memprediksi Waktu Memperbaiki Bug dari Laporan Bug Menggunakan Klasifikasi Random Forest”, *Jurnal Sistem Dan Informatika*, Vol. 11(No. 1), 156–164, 2016.
- [12] Earthquake Kaggle Dataset, U. G, “Significant Earthquakes, 1965-2016” Available on: <https://www.kaggle.com/usgs/earthquake-database#database.csv>. Tanggal akses: 05 Juli 2019.

# KLASIFIKASI PENGADUAN LARAS ONLINE BERBASIS TEXT MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN NAÏVE BAYES

Muhammad Azis Suprayogi  
Universitas Gunadarma  
Jalan Margonda Raya No. 100, Depok, Jawa Barat  
azissuprayogi.gunadarma@gmail.com

## Abstrak

Laras Online adalah fasilitas pada Pemkab Bogor yang diperuntukkan bagi masyarakat terutama warga Kabupaten Bogor sebagai wadah aspirasi dan pengaduan. Seiring dengan jumlah pengaduan yang masuk sangat banyak sehingga mengakibatkan waktu yang lebih lama yang digunakan oleh petugas admin dalam memilah kemudian menentukan unit tujuan pengaduan. Penelitian ini bertujuan untuk membandingkan performansi klasifikasi dokumen pengaduan pada situs Laras Online menggunakan algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes Classifier (NBC). Penelitian dilakukan dengan cara mengumpulkan dokumen pengaduan, melakukan preprocessing, pembobotan kata, klasifikasi, dan pengujian. Pengujian menggunakan cross validation dengan parameter  $k\text{-fold}=10$  dan confusion matrix berdasarkan nilai accuracy, precision, recall, dan score-f1. Hasil pengujian terhadap 360 dataset menunjukkan bahwa algoritma NBC lebih baik dari algoritma KNN dengan nilai  $k=3$ ,  $k=5$ ,  $k=7$ , dan  $k=9$  untuk mengklasifikasikan dokumen pengaduan ke dalam 6 kategori. Hasil klasifikasi menggunakan algoritma NBC memberikan nilai accuracy sebesar 79,16% dengan nilai precision tertinggi pada 2 kategori yaitu Dinsos 91,30% dan SatpolPP 66,80%, nilai recall tertinggi pada 4 kategori yaitu Disdukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50%, serta nilai score-f1 tertinggi pada 4 kategori yaitu Disdukcapil sebesar 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%.

**Kata Kunci:** k-nearest neighbor, klasifikasi, naïve bayes classifier, pengaduan masyarakat

## Abstract

Laras Online is a facility at the Bogor Regency Government that is intended for the community, especially residents of Bogor Regency for aspirations and complaints. Along with the large number of incoming complaints, it results in a longer time used by admin in sorting and determining the complaint unit. This study aims to compare the performance of the classification of public complaints on the Laras Online site using the K-Nearest Neighbor (KNN) and Naïve Bayes Classifier (NBC). The research is carried out by collecting documents, preprocessing, terms weighting, classifying and testing. The test used the  $k\text{-fold}$  cross validation ( $k\text{-fold}=10$ ) and confusion matrix based on the accuracy, precision, and f1-score values. The test results of 360 datasets show that the NBC algorithm is better than the KNN ( $k=3$ ,  $k=5$ ,  $k=7$ , and  $k=9$ ) to classify documents into 6 categories. The NBC classification provide accuracy rate of 79,16% with highest score of precision in 2 categories, namely Dinsos 91,30% and SatpolPP 66,80%, the highest score of recall in 4 categories namely Dukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, and Dishub 76,50%, also the highest score of score-f1 in 4 categories namely Disdukcapil 82,10%, Dislinghidup 82,90%, Dinsos 88,90% and Dishub 81,20%.

**Keywords:** classification, complaints, k-nearest neighbor, naïve bayes classifier

## PENDAHULUAN

Pemkab Bogor memiliki visi mewujudkan Kabupaten Bogor yang termaju, aman, nyaman, dan berkeadaban. Untuk mencapai visi tersebut, salah satu upaya yang dilakukan Pemerintah Kabupaten Bogor adalah menyediakan layanan aspirasi dan pengaduan masyarakat dalam rangka untuk memenuhi hak bertanya masyarakat serta untuk menyampaikan aspirasi atau pengaduan masyarakat terkait penyelenggaraan pemerintahan Kabupaten Bogor.

Pemkab Bogor telah menyediakan fasilitas pengaduan masyarakat melalui aplikasi Layanan Resmi Aspirasi Online (Laras Online). Fasilitas pengaduan tersebut dapat diakses oleh masyarakat melalui situs Laras Online <http://pengaduan.bogorkab.go.id>, melalui SMS yang dikirim ke no. **08111184399** dengan format pesan **teman#nama#alamat#pesan**, melalui aplikasi *mobile android* Laras Online, melalui *email* Satuan Kerja Perangkat Daerah, melalui surat kabar serta dengan cara datang langsung ke kantor yang dituju. Seluruh sumber aspirasi dan pengaduan tersebut terintegrasi dalam satu sistem dan dapat diakses pada situs Laras Online. Dalam hal ini, Dinas Komunikasi dan Informatika memiliki peran sebagai *admin* yang bertugas memonitor, mengelola, serta mengingatkan Satuan Kerja Perangkat Daerah (SKPD) atau Kecamatan untuk segera merespon dan menindaklanjuti aspirasi pengaduan masyarakat tersebut[1]. Jumlah

petugas *admin* yang terbatas serta jumlah pengaduan masuk yang cukup banyak sehingga menyebabkan proses penyampaian aduan masyarakat oleh petugas ke unit tujuan menjadi lambat. Oleh karena itu, perlu adanya solusi berupa sistem yang dapat melakukan klasifikasi penentuan unit tujuan pengaduan masyarakat. Aplikasi pengaduan Laras Online dapat digunakan sebagai objek penelitian dengan tema klasifikasi dokumen secara otomatis.

Penelitian serupa yang pernah dilakukan oleh beberapa peneliti tentang klasifikasi dokumen teks, yaitu Kurnia Iswardani dan Dyah Ariyanti[2] meneliti tentang klasifikasi keluhan masyarakat menggunakan *Naïve Bayes Classifier* untuk mengelompokkan keluhan masyarakat yang berasal dari SMS, telepon, datang ke kantor secara langsung maupun melalui media sosial ke Satuan Kerja Perangkat Desa (SKPD) tujuan. Dari hasil penelitian menunjukkan rata-rata akurasi yang diperoleh sebesar 95%. Penelitian selanjutnya oleh Titien S. Sukanto dan Heru Pramono Hadi[3] tentang klasifikasi jenis laporan masyarakat menggunakan algoritma *K-Nearest Neighbour*. Hasil penelitian ini menunjukkan algoritma *K-Nearest Neighbour* memberikan akurasi yang baik untuk klasifikasi laporan masyarakat menjadi tiga kategori yaitu permintaan informasi, pengaduan, dan aspirasi. Berdasarkan evaluasi dan validasi dengan *Confusion Matrix* diperoleh akurasi tertinggi sebesar 82% menggunakan  $k=11$ . Penelitian

Selanjutnya oleh Aida Indriani[4] yaitu melakukan analisa perbandingan algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier* untuk klasifikasi data forum diskusi mahasiswa menjadi kategori topik berdasarkan isi materi forum. Penelitian tersebut menunjukkan bahwa algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier* dapat digunakan untuk klasifikasi data forum, hasil pengukuran terhadap efektifitas klasifikasi terhadap 15 data uji menunjukkan nilai 73% untuk *Naïve Bayes Classifier* dan nilai 80% untuk *K-Nearest Neighbour* dengan menggunakan *Confusion Matrix*, sehingga disimpulkan bahwa *K-Nearest Neighbour* lebih baik dari *Naïve Bayes Classifier*. Penelitian berikutnya dilakukan oleh Hamdani Asril, Insanul Kamila, dan Mustakim[5] terhadap klasifikasi dokumen tugas akhir menggunakan algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Penelitian tersebut menghasilkan akurasi terbaik pada percobaan tiga kelas sebesar 91,67% untuk *K-Nearest Neighbor* nilai  $k=6$ ,  $k=7$ ,  $k=8$  dan  $k=9$  dan sebesar 86,11% untuk *Naïve Bayes Classifier*. Berdasarkan penelitian-penelitian sebelumnya tersebut, kami mencoba melakukan perbandingan akurasi klasifikasi antara *K-Nearest Neighbor (KNN)* dengan *Naïve Bayes Classifier (NBC)* menggunakan pengujian *K-fold Cross Validation* pada objek penelitian yang berbeda yaitu pengaduan masyarakat

Laras Online pada Pemerintah Kabupaten Bogor.

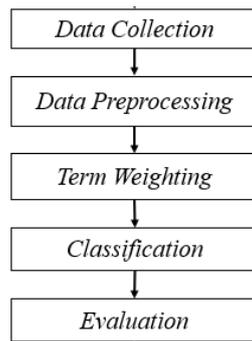
Tujuan penelitian ini untuk membandingkan dan mendapatkan model klasifikasi pengaduan dengan akurasi terbaik antara algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Selain itu dapat memberi saran kepada admin Laras Online untuk peningkatan fungsi sistem Laras Online agar dapat memberikan pelayanan yang lebih cepat.

## METODE PENELITIAN

Kerangka kerja secara umum dapat dilihat pada Gambar 1, proses terdiri dari 5 tahapan yaitu proses pengumpulan data (*Data Collection*), praproses teks (*Data Preprocessing*), proses pembobotan kata (*Term Weighting*), proses klasifikasi (*Classification*) dan proses evaluasi (*Evaluation*).

### Pengumpulan Data

Dataset yang digunakan pada penelitian adalah data aduan masyarakat pada sistem Laras Online yang bersumber dari SMS, *website* Laras Online <http://pengaduan.bogorkab.go.id> dan aplikasi *mobile* mulai bulan Januari 2019 sampai dengan Desember 2020. Data-data tersebut dapat diakses pada situs pengaduan Kabupaten Bogor secara bebas.



**Gambar 1. Kerangka Kerja Umum Penelitian**

Hasil *webscrapping* data *website* Laras Online seluruhnya sebanyak 829 data yang terbagi menjadi 53 instansi tujuan aduan. Proses pengumpulan data dilakukan dengan cara *webscrapping* halaman *website* Laras Online menggunakan *library* berbasis Python yaitu *BeautifulSoup*. Selanjutnya data teks tersebut diubah ke dalam *format CSV* dengan tujuan supaya lebih mudah diolah menggunakan *tools* pemrograman *Jupyter Notebook*.

Data yang digunakan dalam penelitian perlu dibatasi jumlahnya supaya waktu yang digunakan lebih efisien sehingga proses pengujian dan evaluasi dapat dilakukan secara efektif. Dengan demikian, perlu ditentukan kriteria data yang digunakan untuk penelitian yaitu: 1) Data aduan yang masuk mulai Januari 2019 s.d. Desember 2020, 2) Data aduan masuk melalui *website*, SMS, serta aplikasi *mobile*, 3) Data aduan sudah mendapat respon petugas *admin* Laras Online, 4) Kategori atau kelas ditentukan berjumlah 6 kategori berdasarkan unit yang memiliki

jumlah aduan terbanyak yaitu Dinas Kependudukan dan Pencatatan Sipil, Dinas Lingkungan Hidup, Dinas Sosial, Dinas Pekerjaan Umum dan Penataan Ruang, Dinas Perhubungan, dan Satuan Polisi Pamong Praja. Berdasarkan kriteria data tersebut, ditentukan dataset penelitian berjumlah 360 aduan. Tabel 1 memperlihatkan contoh data yang digunakan untuk penelitian.

Selanjutnya dipilih 6 kategori dari data tersebut yang merupakan unit tujuan terpopuler berdasarkan jumlah data aduan terbanyak, selanjutnya diambil 60 data dari masing-masing unit tersebut. Tujuan pengambilan data seimbang pada masing-masing unit dalam rangka mengeliminasi faktor data tidak seimbang yang dapat mempengaruhi tingkat akurasi klasifikasi kategori tertentu. Kami membuat kategori/label **0,1,2,3,4,5** yang mencerminkan unit/kategori klasifikasi untuk mempermudah proses penelitian. Jumlah data masing-masing kategori yang digunakan pada penelitian dapat dilihat pada Tabel 2.

**Tabel 1. Contoh Data Aduan pada Laras Online**

<b>Aduan</b>	<b>Kategori</b>
Lampu merah perempatan pemda mati sejak pagi hari tadi. Mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. Terimakasih	Dinas Perhubungan
Mohon dibantu perbaikan jalan raya bojonggede menuju citayam terutama depan rumah sakit citama dan suzuki citayam, saya lihat ini masih bagian dari kabupaten bogor, dan setiap kali hujan pasti jalan tergenang ber hari hari	Dinas Pekerjaan Umum dan Penataan Ruang
Saya urus surat pindah dari ciomas ke jakarta dari tanggal 27 oktober samapai skrg belum Selesai dan belum ada jawaban dari dukcapil	Dinas Kependudukan dan Pencatatan Sipil
Truck pengangkut sampah menjatuhkan sampah didepan restoran kami sehingga berserakan dan berantakan, mengganggu dan beraroma busuk, TOLONG SEGERA DIBERSIHKAN	Dinas Lingkungan Hidup
Assalamu'alaikum maaf bapak/ibu, keluarga saya belum mendapat bantuan selama masa pandemi ini padahal tetangga-tetangga sudah mendapat bantuan lebih dari satu kali, bagaimana agar keluarga saya bisa dapat bantuan? Lokasi desa ciomas, kecamatan ciomas	Dinas Sosial
Kepada pemerintahan kabupaten bogor mohon kira nya dapat ditertibkan atau ditutup warung minum beralkohol dan karaokean ditempat terbuka dijalan utama narogong kp cikerewis desa limusnunggal sehingga mengganggu waktu istirahat warga sekitar...mohon kira nya ditindak demi kenyamanan warga sekitar ...terima kasih untuk perhatiannya	Satuan Polisi Pamong Praja

**Tabel 2. Jumlah Dataset Penelitian**

<b>Kategori</b>	<b>Label</b>	<b>Jumlah Data</b>
Dinas Kependudukan dan Pencatatan Sipil	0	60
Dinas Lingkungan Hidup	1	60
Dinas Sosial	2	60
Dinas Pekerjaan Umum dan Penataan Ruang	3	60
Dinas Perhubungan	4	60
Satuan Polisi Pamong Praja	5	60
<b>Total</b>		<b>360</b>

### **Preprocessing**

*Preprocessing* data dilakukan melalui beberapa tahapan, dimulai dari tahap *case folding* yaitu mengubah teks dokumen menjadi huruf kecil seluruhnya, kemudian tahap *tokenization* yaitu membersihkan teks dari tanda baca, spasi berulang, mengubah baris baru menjadi spasi, dan memisahkan kata per kata dari kalimatnya, *stopword* yaitu menghapus kata yang tidak dibutuhkan, dan

tahap *stemming* yaitu menghilangkan imbuhan kata pada teks dokumen. Adapun prosesnya menggunakan *tools* berbasis Python berupa *library Numpy, Pandas, Sastrawi, dan NLTK*.

### **Pembobotan, Klasifikasi dan Evaluasi**

Proses *preprocessing* dokumen hanya menghasilkan dokumen *bag of words* yang perlu diubah menjadi *term vector* yang bisa dilakukan dengan cara proses pembobotan[6].

Proses pembobotan dilakukan setelah *preprocessing* dengan menggunakan metode *Term Frequency – Inverse Document Frequency (TF-IDF)* yaitu menghitung bobot *term* pada sebuah dokumen berdasarkan seringnya kata tersebut muncul dimana bobot tersebut mengindikasikan pentingnya sebuah *term* terhadap dokumen, semakin banyak *term* tersebut muncul pada dokumen maka semakin tinggi nilai *term* tersebut[7]. Teknik TF-IDF juga mengeliminasi *term* yang bersifat sangat umum dan mengekstrak *term* yang memiliki relevansi yang tinggi dari *corpus*[8]. Tahap pertama menentukan nilai *Term Frequency (TF)* yaitu jumlah *term* yang terdapat pada setiap dokumen, tahap selanjutnya menentukan nilai *Inverse Document Frequency (IDF)* yang berfungsi untuk mengurangi bobot *term* yang jumlah kemunculannya banyak di seluruh dokumen menggunakan Persamaan (1). Selanjutnya menentukan bobot *term* dengan cara mengalikan nilai TF dengan IDF menggunakan Persamaan (2).

$$idf_i = \log\left(\frac{N}{df_i}\right) \quad (1)$$

$$w_{i,j} = tf_{i,j} \times idf_i \quad (2)$$

Dimana  $w_{i,j}$  adalah bobot *term*  $i$  pada dokumen  $j$ ,  $tf_{i,j}$  adalah jumlah kemunculan *term*  $i$  pada dokumen  $j$ ,  $idf_i$  adalah *inverse*  $df_i$ ,  $df_i$  adalah banyaknya dokumen yang memuat *term* ke- $i$ .  $N$  adalah jumlah dokumen. Proses klasifikasi menggunakan algoritma Naïve bayes Classifier dan *K-Nearest*

*Neighbor* serta memanfaatkan modul klasifikasi pada *scikit-learn python* yaitu *KNeighborsClassifier* dan *MultinomialNB*. Model algoritma *K-Nearest Neighbor* dibuat menggunakan ukuran jarak *Euclidean Distance* dan nilai  $k$  yang umum digunakan yaitu  $k=3,5,7$ , dan 9 untuk kemudian dipilih nilai  $k$  dengan performansi terbaik. Adapun model algoritma Naïve Bayes merupakan algoritma klasifikasi sederhana yang tidak mempunyai banyak pengaturan.

Selanjutnya dilakukan uji *K-fold Cross Validation* dengan nilai  $k=10$ . Pemilihan nilai  $k=10$  tersebut karena pengujian *Cross Validation* pada nilai  $k=10$  memiliki bias terkecil pada estimasi *error* prediksi serta memiliki nilai *mean square error* terendah[9][10]. Berdasarkan nilai  $k=10$ , maka jumlah keseluruhan 360 data pada penelitian ini akan dibagi ke dalam 10 bagian data yang masing-masing bagian memiliki 36 data dan pada setiap bagiannya terdiri dari 6 data dari masing-masing kategori. Proses pengujian dilakukan sebanyak  $k$  kali, jika salah satu bagian digunakan sebagai data *testing*, maka sembilan bagian sisanya akan menjadi data *training*. Tahapan selanjutnya adalah evaluasi klasifikasi menggunakan pengukuran nilai akurasi, recall, precision dan score-f1 dalam satuan persen berdasarkan rumusan menggunakan *Confusion Matrix* yang ditampilkan pada Tabel 3. True Positive (TP) adalah data yang diklasifikasikan berkaitan dengan kategori yang benar, False Positive (FP) adalah data yang

diklasifikasikan berkaitan dengan kategori yang salah, True Negative (TN) adalah data yang diklasifikasikan tidak berkaitan dengan kategori yang benar, False Negative (FN) adalah data yang diklasifikasikan tidak berkaitan dengan kategori yang salah[11]. Terdapat beberapa pengukuran untuk menghitung performansi klasifikasi yaitu *accuracy*, *precision*, *recall*, dan *score-f1*. Accuracy adalah jumlah proporsi prediksi yang benar[12], didefinisikan dalam Persamaan (3)

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

*Precision* adalah proporsi jumlah dokumen yang relevan di antara seluruh dokumen yang terpilih oleh sistem[12], didefinisikan dalam Persamaan (4)

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

*Recall* adalah proporsi jumlah dokumen yang relevan di antara seluruh dokumen relevan yang terdapat pada koleksi dokumen[12], didefinisikan dalam Persamaan (5)

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Adapun *score-f1* (*f-measure*) didefinisikan dalam Persamaan (6)

$$score - f1 = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

## HASIL DAN PEMBAHASAN

### Hasil Preprocessing

Beberapa tahapan *preprocessing* yang dilakukan, yaitu *case folding*, *tokenization*, *stopword*, dan *stemming*. Contoh tahapan *preprocessing* pada teks aduan dapat dilihat seperti Tabel 4.

**Tabel 3. Confusion Matrix**

Terprediksi	Aktual	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

**Tabel 4. Contoh Teks Aduan Hasil Preprocessing**

Tahapan	Hasil
<i>Raw Data</i>	Lampu merah perempatan pemda mati sejak pagi hari tadi. Mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. Terimakasih
<i>Case Folding</i>	lampu merah perempatan pemda mati sejak pagi hari tadi. mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. terimakasih
<i>Tokenization</i>	lampu merah perempatan pemda mati sejak pagi hari tadi mohon ditindak lanjuti karena sangat berbahaya bagi pengguna jalan terimakasih
<i>Stopword</i>	lampu merah perempatan pemda mati ditindak lanjuti berbahaya pengguna jalan
<i>Stemming</i>	lampu merah empat mati bahaya guna jalan

### Hasil Pembobotan

Proses ini dilakukan dengan menghitung bobot tiap kata (*term weighting*) pada sebuah dokumen berdasarkan seringnya kata tersebut muncul pada dokumen dengan cara mengalikan nilai TF dengan nilai IDF. Hasil dari pembobotan berupa nilai TF-IDF yang akan digunakan untuk perhitungan klasifikasi KNN dan NBC. Contoh nilai TF-IDF dapat dilihat pada Tabel 5.

### Hasil Klasifikasi KNN dan NBC

Klasifikasi menggunakan data sebanyak 360 dokumen yang merupakan hasil

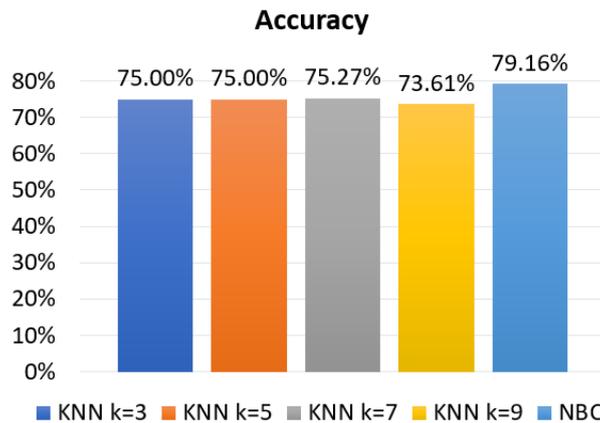
tahap *preprocessing* dan tahap pembobotan TF-IDF. Klasifikasi dilakukan menggunakan algoritma KNN dengan variasi nilai  $k$   $k=3$ ,  $k=5$ ,  $k=7$ , dan  $k=9$  dan algoritma NBC kemudian dilakukan uji *Cross Validation* dengan nilai  $k\text{-fold}=10$  terhadap seluruh data dan hasilnya diukur dengan menghitung nilai rata-rata *accuracy*. Pengukuran juga dilakukan terhadap nilai *precision*, *recall*, dan *score-f1* untuk masing-masing kategori yaitu Disdukcapil, Dislinghidup, Dinsos, DisPupr, Dishub, dan SatpolPP. Perbandingan nilai *accuracy* hasil klasifikasi dapat dilihat pada Tabel 6 dan grafik pada Gambar 2.

**Tabel 5. Contoh Perhitungan Nilai TF dan IDF**

Term	tf pada d1	df untuk N=360	idf = $\text{Log}(N/\text{df})$	w = tf x idf
lampu	1	32	1,0511	1,0511
merah	1	8	1,6532	1,6532
empat	1	2	2,2552	2,2552
mati	1	23	1,1945	1,1945
bahaya	1	22	1,2138	1,2138
guna	1	18	1,3010	1,3010
jalan	1	118	0,4844	0,4844

**Tabel 6. Perbandingan Nilai Accuracy**

Algoritma	Accuracy
KNN $k=3$	75,00 %
KNN $k=5$	75,00 %
KNN $k=7$	75,27 %
KNN $k=9$	73,61 %
NBC	79,16 %



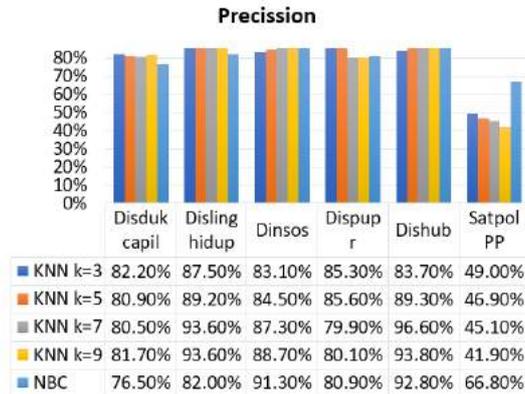
Gambar 2. Perbandingan Accuracy Klasifikasi dengan Cross Validation

Tabel 7. Perbandingan Nilai Precision, Recall, Dan Score-F1

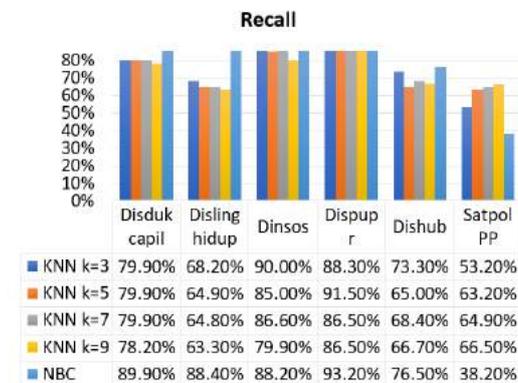
Algoritma	Pengukuran	Disduk capil	Disling hidup	Dinsos	Dis pupr	Dishub	Satpol PP
KNN k=3	precision	82,20%	87,50%	83,10%	85,30%	83,70%	49,00%
KNN k=3	recall	79,90%	68,20%	90,00%	88,30%	73,30%	53,20%
KNN k=3	score-f1	80,20%	73,20%	85,50%	82,10%	74,50%	49,30%
KNN k=5	precision	80,90%	89,20%	84,50%	85,60%	89,30%	46,90%
KNN k=5	recall	79,90%	64,90%	85,00%	91,50%	65,00%	63,20%
KNN k=5	score-f1	79,60%	71,60%	83,40%	87,90%	71,50%	53,00%
KNN k=7	precision	80,50%	93,60%	87,30%	79,90%	96,60%	45,10%
KNN k=7	recall	79,90%	64,80%	86,60%	86,50%	68,40%	64,90%
KNN k=7	score-f1	79,40%	72,90%	85,90%	82,50%	76,20%	52,10%
KNN k=9	precision	81,70%	93,60%	88,70%	80,10%	93,80%	41,90%
KNN k=9	recall	78,20%	63,30%	79,90%	86,50%	66,70%	66,50%
KNN k=9	score-f1	78,70%	72,40%	83,20%	82,90%	74,60%	50,30%
NBC	precision	76,50%	82,00%	91,30%	80,90%	92,80%	66,80%
NBC	recall	89,90%	88,40%	88,20%	93,20%	76,50%	38,20%
NBC	score-f1	82,10%	82,90%	88,90%	85,80%	81,20%	44,60%

Berdasarkan hasil perbandingan pada Tabel 6 dapat disimpulkan bahwa hasil accuracy tertinggi klasifikasi dokumen pengaduan diperoleh algoritma NBC sebesar 79,16 %. Adapun hasil akurasi algoritma KNN k=3 sebesar 75,00 %, k=5 sebesar 75,00

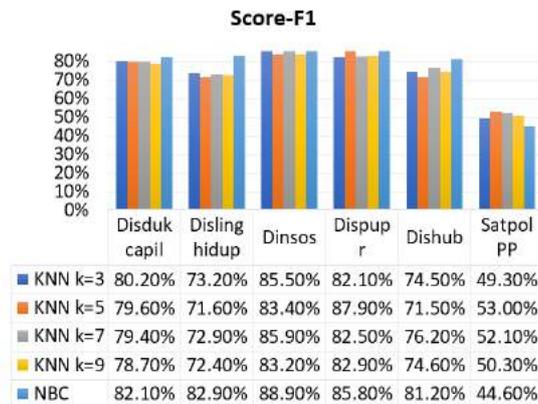
%, k=7 sebesar 75,27 %, dan k=9 sebesar 73,61%. Adapun untuk perbandingan pengukuran nilai precision, recall, dan score-f1 masing-masing kategori pada 6 kategori yang tersedia dapat dilihat pada Tabel 7 serta grafik perbandingan pada Gambar 3,4, dan 5.



Gambar 3. Perbandingan Nilai *Precision* terhadap Kategori Klasifikasi



Gambar 4. Perbandingan Nilai *Recall* terhadap Kategori Klasifikasi



Gambar 5. Perbandingan Nilai *Score-F1* terhadap Kategori Klasifikasi

Hasil pengukuran berdasarkan Tabel 7 dan grafik pada Gambar 3,4, dan 5 menunjukkan bahwa algoritma NBC memberikan nilai tertinggi untuk pengukuran *precision* pada 2 kategori yaitu Dinsos sebesar 91,30% dan SatpolPP 66,80%, sedangkan untuk 4 kategori lainnya yaitu Disdukcapil, Dislinghidup,

Dispupr dan Dishub nilai tertinggi diperoleh algoritma KNN dengan nilai  $k$  yang bervariasi yaitu secara berurutan sebesar 82,20% (KNN  $k=3$ ), 93,60% (KNN  $k=7, k=9$ ), 85,60% (KNN  $k=5$ ), dan 96,60% (KNN  $k=7$ ). Algoritma NBC memberikan nilai tertinggi untuk pengukuran *recall* pada 4 kategori yaitu

Disdukcapil sebesar 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50%, sedangkan 2 kategori lainnya diperoleh algoritma KNN dengan nilai  $k$  yang bervariasi yaitu Dinsos 90,00% (KNN  $k=3$ ), dan SatpolPP 66,50% (KNN  $k=9$ ).

Adapun untuk pengukuran nilai *score-f1*, nilai tertinggi diperoleh algoritma NBC untuk 4 kategori yaitu Disdukcapil sebesar 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%. Untuk 2 kategori lainnya diperoleh algoritma KNN dengan nilai  $k$  yang bervariasi yaitu Dispupr 87,90% (KNN  $k=5$ ) dan SatpolPP 53,00% (KNN  $k=5$ ). Hasil perbandingan pengukuran nilai *precision*, *recall*, dan *score-f1* tersebut menunjukkan bahwa algoritma NBC mengungguli algoritma KNN dengan nilai  $k$  yang bervariasi ( $k=3, k=5, k=7, k=9$ ) pada sebagian besar kategori klasifikasi.

Nilai *precision*, *recall* dan *score-f1* pada kategori SatpolPP memiliki nilai yang lebih rendah dibandingkan dengan kategori yang lain. Hal ini disebabkan karena pada kategori tersebut banyak terdapat kata yang memiliki nilai pembobotan kata yang rendah sehingga berpengaruh pada klasifikasi.

## KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian dapat disimpulkan bahwa klasifikasi pengaduan masyarakat pada situs Laras Online dapat dilakukan menggunakan

algoritma KNN dan NBC didahului dengan proses *preprocessing* serta pembobotan kata.

Hasil pengujian *k-fold cross validation* dengan parameter *k-fold=10* terhadap klasifikasi menggunakan algoritma KNN dengan nilai  $k=3, k=5, k=7$ , dan  $k=9$  dan algoritma NBC menunjukkan bahwa nilai *accuracy* tertinggi diperoleh algoritma NBC sebesar 79,16%.

Algoritma NBC memperoleh nilai *precision* tertinggi pada 2 kategori dari total 6 kategori yaitu kategori Dinsos 91,30% dan SatpolPP 66,80%, nilai *recall* tertinggi pada 4 kategori dari 6 kategori yaitu Disdukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50% serta nilai *score-f1* tertinggi pada 4 kategori dari 6 kategori yaitu Disdukcapil 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%. Dengan demikian dapat disimpulkan bahwa algoritma NBC lebih baik dari algoritma KNN dalam mengklasifikasikan pengaduan masyarakat pada situs Laras Online.

Saran untuk hasil klasifikasi yang lebih akurat agar dilakukan proses *preprocessing* yang lebih baik yaitu dengan menambahkan fungsi perbaikan penulisan kata yang salah (*spelling correction*) maupun fungsi perbaikan penulisan spasi antar kata pada data aduan karena dalam data aduan ditemukan penulisan beberapa kata yang digabung tanpa menggunakan spasi. Diharapkan dengan melakukan proses *preprocessing* yang lebih baik akan meningkatkan akurasi klasifikasi.

## DAFTAR PUSTAKA

- [1] Dinas Komunikasi dan Informatika Kabupaten Bogor, "Layanan Aspirasi Online." <https://diskominfo.bogorkab.go.id/laras-online/> (accessed Jan. 02, 2021).
- [2] D. Ariyanti, K. Iswardani, U. Panca, and M. Probolinggo, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes," *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [3] H. P. Hadi and T. S. Sukamto, "Klasifikasi Jenis Laporan Masyarakat Dengan K-Nearest Neighbor Algorithm," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 77–85, 2020, doi: 10.33633/joins.v5i1.3355.
- [4] A. Indriani, "Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data," *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020, doi: 10.46984/sebatik.v24i1.909.
- [5] I. K. Hamdani Asril, Mustakim, "Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 2579–5406, 2019, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/7995>.
- [6] V. Kalra and R. Aggarwal, "Importance of Text Data Preprocessing & Implementation in RapidMiner," *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.*, vol. 14, pp. 71–75, 2018, doi: 10.15439/2017km46.
- [7] N. G. Yudiarta, M. Sudarma, and W. G. Ariastina, "Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 3, p. 339, 2018, doi: 10.24843/mite.2018.v17i03.p06.
- [8] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," *Int. Conf. Electr. Electron. Optim. Tech. ICEEOT 2016*, no. November, pp. 61–66, 2016, doi: 10.1109/ICEEOT.2016.7754750.
- [9] D. Berrar, "Cross-validation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. April, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [10] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: A comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005, doi: 10.1093/bioinformatics/bti499.
- [11] D. Yuliana and C. Supriyanto, "Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network," *UPI*

- YPTK J. KomTekInfo*, vol. 5, no. 3, pp. 92–116, 2019.
- [12] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.

# MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DALAM PERAMALAN NILAI HARGA SAHAM PENUTUP INDEKS LQ45

<sup>1</sup>Devita Priyadi, <sup>2</sup>Iffatul Mardhiyah

<sup>1,2</sup>Manajemen Sistem Informasi, Program Pasca Sarjana, Universitas Gunadarma  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

<sup>1</sup>priyadi.devita@gmail.com, <sup>2</sup>iffatul@staff.gunadarma.ac.id

## Abstrak

Data indeks LQ45 dapat digunakan membantu manajer investasi, investor ataupun calon investor terkait dalam proses perencanaan dan proses pengambilan keputusan dalam membeli ataupun menjual saham. Data indeks LQ45 memiliki peran penting dalam melakukan peramalan untuk mencapai tujuan terkait proses perencanaan dan proses pengambilan keputusan dalam membeli ataupun menjual saham. Peramalan deret waktu (time series) menggunakan penerapan model Autoregressive Integrated Moving Average (ARIMA) untuk meramalkan nilai harga saham penutup dalam indeks LQ45 pada data mingguan. Data yang digunakan merupakan data dari 25 November 2019 sampai dengan 30 November 2020. Hasil pengujian model terbaik adalah ARIMA(1,1,1). Model ARIMA(1,1,1) terpilih karena memenuhi asumsi dan didukung oleh nilai Adjusted R-squared, nilai S.E. of regression, dan Akaike Info Criterion. Hasil peramalan jangka pendek selama 2 bulan (7 Desember 2020 sampai 25 Januari 2021) yang didapat dari model ARIMA(1,1,1) mendekati data aktual dengan nilai Mean Absolute Percentage Error (MAPE) yang paling kecil yaitu 18.41269.

**Kata Kunci:** ARIMA, Deret Waktu, Indeks LQ45, Mean Absolute Percentage Error, Peramalan

## Abstract

The LQ45 can be used to help investment manager, investors or potential investors involved in the planning and decision-making process for buying or selling stocks. The LQ45 has an important role as a forecast to achieve the purpose in the planning and decision-making process for buying or selling stocks. Time series forecasting uses the practice of the Autoregressive Integrated Moving Average (ARIMA) model to forecast the value of closing stock price in the LQ45 index on weekly data. The data used are from 25 November 2019 to 30 November 2020. As result, the best model test is ARIMA (1,1,1). This model was chosen because it meets the presumption and is supported by the Adjusted R-squared value, the S.E. of regression, and Akaike Info Criterion. Moreover, the result of short-term forecasting over the next 2 months (7 December 2020 to 25 January 2021) is obtained from the ARIMA model (1,1,1) which is close to the actual data with the smallest Mean Absolute Percentage Error (MAPE) value, 18.41269.

**Keywords:** ARIMA, Forecasting, LQ45 index, Mean Absolute Percentage Error, Time Series

## PENDAHULUAN

Harga saham merupakan salah satu indikator minat dari calon investor untuk memiliki saham suatu perusahaan. Informasi pergerakan indeks harga saham, kinerja harga saham, laporan keuangan perusahaan, dan sebagainya. Data informasi mengenai pergerakan indeks harga saham dapat diperoleh melalui Bursa Efek Indonesia (BEI). Indeks LQ45 merupakan indikator indeks saham di Bursa Efek Indonesia (BEI) yang merupakan 45 emiten termasuk dalam kategori saham *blue chip* dengan saham-saham yang memiliki kapitalisasi besar dan memiliki pengaruh besar dalam perekonomian. Salah satu keuntungan dalam berinvestasi pada perusahaan-perusahaan dengan label LQ45 adalah perusahaan tersebut merupakan perusahaan yang teruji tidak hanya dalam segi kapitalisasi pasar (umumnya di atas 4 Triliun) melainkan secara ketahanan bisnis dan pengaruhnya terhadap ekonomi Indonesia [1].

Data harga saham merupakan deret waktu (*time series*) yang memiliki aktifitas yang tinggi, sehingga membutuhkan analisis untuk meminimalkan resiko dalam membeli dan menjual saham. Peramalan diperlukan untuk membantu dalam proses perencanaan dan proses pengambilan keputusan, namun hasil peramalan tidaklah selalu akurat. Keakuratan hasil peramalan tergantung dengan ketepatan data dan metode yang digunakan. Model peramalan deret waktu

(*time series*) terus berkembang seiring waktu, salah satu model yang banyak digunakan untuk meramalkan data ekonomi adalah *Autoregressive Integrated Moving Average* (ARIMA). Penelitian ini mengenai analisis peramalan deret waktu (*time series*) pada data penutupan harga saham kategori LQ45 menggunakan model *Autoregressive Integrated Moving Average* (ARIMA). Penelitian dengan memaparkan keunggulan akurasi peramalan model *Auto Regressive Integrated Moving Average* (ARIMA) pada data perbankan dari pasar saham Amman (ASE) di Yordania [2]. Penelitian menggunakan model ARIMA pada data Indonesia *Composite Index* (ICI). Data yang digunakan adalah data bulanan Indonesia *Composite Index* (ICI) di Bursa Efek Indonesia (BEI) periode Januari 2000 sampai dengan Desember 2017 [3]. Penelitian dengan menggunakan model ARIMA untuk meramalkan harga saham penutupan harian sektor Bank *Nifty* di India [4]. Data yang digunakan pada periode Juli 2016 sampai dengan Juni 2017. Penelitian dengan mengimplementasikan model ARIMA untuk meramalkan deret waktu (*time series*) mengenai data harian *Nifty* dari indeks *Nifty50* di Bursa Efek India [5]. Penelitian yang menggunakan model *Auto Regressive Integrated Moving Average* (ARIMA) yang bertujuan untuk mengembangkan model ARIMA sesuai analisis dan peramalan harga saham Sektor Otomotif [6]. Penelitian menggunakan model ARIMA dalam

meramalkan tren empat harga saham perusahaan minyak sektor publik HPCL, IOCL, BPCL, dan ONGC di Bombay *Stock Exchange* [7]. Penelitian selanjutnya dilakukan dengan mengimplementasikan model ARIMA untuk memprediksi tren saham fluktuasi pasar di Shanghai *Composite Stock Price Index* (SCSPI).

Data yang digunakan untuk penelitian dimulai dari periode Januari 2014 sampai dengan Desember 2014 [8]. Penelitian dilakukan dengan mengkombinasikan model ARIMA dengan *Wavelet de-noising*. Data yang digunakan adalah indeks harga saham yang termasuk ke dalam saham Asia, yaitu Hong Kong Hang Seng *Index* (HSI), Taiwan *Weighted Index* (TAIEX), Shanghai *Composite Index* (SSE) dan Shenzhen *Component Index* (SZSE) [9]. Kemudian penelitian dengan menggunakan model ARIMA dalam memprediksi data runtun waktu harga penutup yang telah dikumpulkan dari Bursa Efek Amman (ASE). Data yang dikumpulkan dari Bursa Efek Amman (ASE) periode Januari 2010 sampai dengan Januari 2018 [10].

Penelitian dilakukan dengan mengimplementasikan penggunaan model *Auto Regressive Integrated Moving Average* (ARIMA) dengan menggunakan model *selection criteria* yaitu *Akaike Information Criterion* (AIC) untuk memilih model regresi terbaik. Data yang digunakan merupakan data harga penutupan harian pada Indeks Harga Saham Gabungan (IHSG) periode 4 Januari

2010 sampai dengan 5 Desember 2014 [11]. Penelitian ini mengenai peramalan deret waktu (*time series*) dengan penerapan model *Autoregressive Integrated Moving Average* (ARIMA) untuk meramalkan nilai harga saham penutup dalam indeks LQ45 dengan menggunakan data mingguan penutupan harga saham.

Data yang digunakan merupakan data sekunder yang diperoleh dari situs *Yahoo Finance* dengan total data observasi sebanyak 242 data harian, dari hari Senin sampai Jumat tidak termasuk hari libur [12].

Data diubah menjadi data mingguan untuk mendapatkan data deret waktu yang sesuai dengan penelitian, sehingga diperoleh data observasi sebanyak 54 data. Data observasi dalam penelitian ini merupakan data dari 25 November 2019 sampai dengan 30 November 2020 untuk meramalkan nilai harga saham penutup dalam indeks LQ45 pada 2 bulan ke depan.

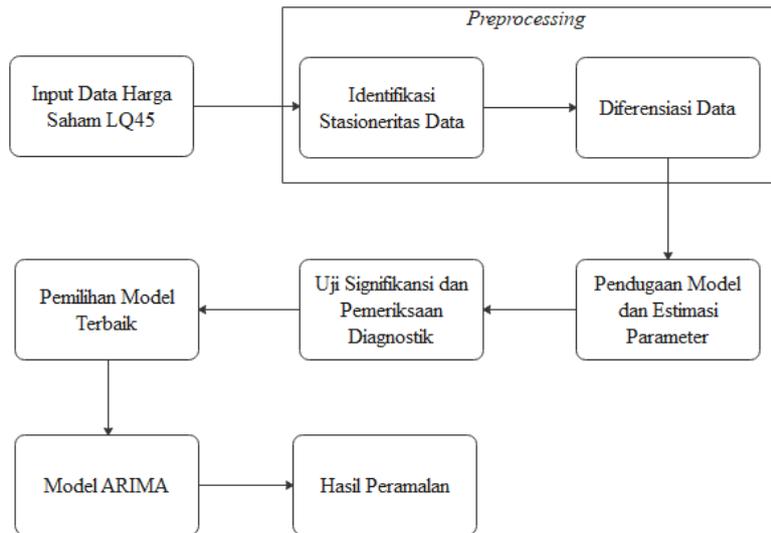
Penerapan model *Autoregressive Integrated Moving Average* (ARIMA) dalam penelitian menggunakan nilai *Mean Absolute Percentage Error* (MAPE) dalam mengevaluasi akurasi peramalan. Hasil penelitian diharapkan dapat menghasilkan model yang tepat dalam meramalkan nilai harga saham penutup dalam indeks LQ45, sehingga membantu manajer investasi, investor ataupun calon investor terkait dalam proses perencanaan dan proses pengambilan keputusan dalam membeli ataupun menjual saham.

## METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan proses dalam menentukan model yang tepat untuk meramalkan nilai harga saham penutup dalam indeks LQ45 menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA). Bagan umum metode penelitian ini dapat dilihat pada Gambar 1. Secara garis besar, metode pada penelitian terdiri atas beberapa tahapan proses seperti dapat dilihat pada Gambar 1. Tahapan awal dimulai dari input data harga saham LQ45. Tahap selanjutnya *preprocessing* yang terdiri dari 2 tahap yaitu identifikasi stasioneritas data dan diferensiasi data [13]. Setelah itu tahapan yang perlu diperlukan yaitu analisis regresi model-model sementara untuk mengestimasi suatu parameter yang nilainya tidak diketahui, untuk mengestimasi suatu parameter digunakan suatu metode pendugaan parameter melalui metode Kuadrat Terkecil (*Least Square*) [14]. Tahap selanjutnya model-model sementara yang didapat perlu adanya uji signifikansi dan pemeriksaan diagnostik

meliputi pengujian hipotesis (asumsi) berdasarkan estimasi model dan pemeriksaan asumsi *white noise* melalui uji statistik *Ljung-Box* menggunakan *Correlogram Residual*. Model-model yang telah melalui uji signifikansi dan pemeriksaan model memerlukan proses pemilihan model untuk menghasilkan model-model sementara, sehingga dari model-model tersebut dapat ditentukan model ARIMA terbaik menggunakan *Akaike's Information Criterion* (AIC).

Tahap terakhir adalah menghasilkan model ARIMA terbaik untuk meramalkan nilai harga saham penutup dalam indeks LQ45 selama 2 bulan periode tanggal 7 Desember 2020 sampai 25 Januari 2021. Hasil peramalan yang diperoleh dapat dilanjutkan dengan tahap Akurasi Model ARIMA dengan pemeriksaan *error* berdasarkan evaluasi *Mean Absolute Percentage Error* (MAPE). Semakin kecil nilai MAPE menunjukkan data hasil peramalan semakin mendekati nilai aktual, sehingga model yang dipilih merupakan model dengan nilai MAPE terkecil.



**Gambar 1. Tahapan Metode Penelitian ARIMA**

**Tabel 1. Sampel Data Mingguan Nilai Harga Saham Penutup dalam Indeks LQ45**

Periode	Harga Penutup (Rp)
11/25/2019	5.865.860.047
12/2/2019	5.911.105.011
12/9/2019	5.956.349.974
12/16/2019	6.098.220.032
12/23/2019	6.095.169.922
12/30/2019	6.109.074.951
1/6/2020	6.122.979.980
1/13/2020	6.145.570.070
1/20/2020	5.979.459.960
1/27/2020	5.897.585.022

[Sumber: *Yahoo Finance*, 2020]

### Input Data Harga Saham LQ45

Pada tahapan ini, data input yang digunakan yaitu nilai harga saham penutup dalam indeks LQ45 merupakan data sekunder yang diperoleh dari situs *Yahoo Finance* dengan total data observasi sebanyak 242 data harian, dari hari Senin sampai Jumat tidak termasuk hari libur [12]. Data diubah menjadi data mingguan untuk mendapatkan data deret waktu yang sesuai dengan penelitian, sehingga diperoleh data observasi sebanyak

54 data. Data observasi dalam penelitian merupakan data dari 25 November 2019 sampai dengan 30 November 2020. Contoh sampel data mingguan nilai harga saham penutup dalam indeks LQ45 yang digunakan untuk data input dapat dilihat pada Tabel 1.

### ARIMA

Model ARIMA mempunyai beberapa tahapan, yaitu tahapan *preprocessing* yang memiliki 2 tahapan yaitu identifikasi

stasioneritas data dan diferensiasi data. Tahapan setelah *preprocessing* yaitu pendugaan model dan estimasi parameter, uji signifikansi dan pemeriksaan diagnostik, pemilihan model, menentukan model ARIMA terbaik, kemudian melakukan peramalan [15].

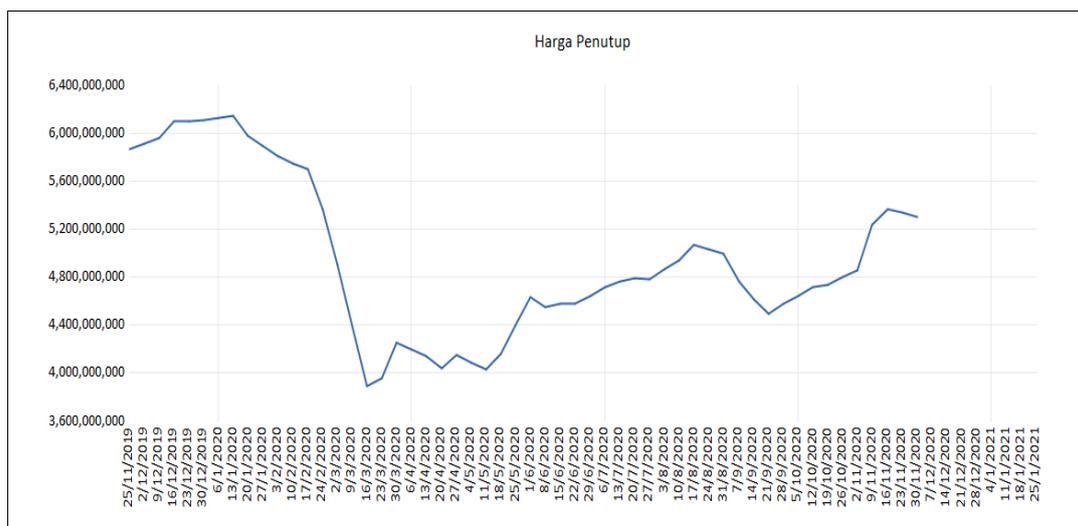
### Identifikasi Stasioneritas Data

Tahapan identifikasi stasioneritas data merupakan *preprocessing* yang pertama yaitu identifikasi stasioneritas data, data input yang digunakan berupa data mingguan nilai harga saham penutup dalam indeks LQ45. Data input berupa nilai harga saham penutup dalam indeks LQ45 perlu dilakukan pengecekan kestasioneran melalui uji *unit root* menggunakan *Augmented Dickey-Fuller* (ADF), serta melalui plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) menggunakan

*Correlogram*. Gambar 2 menunjukkan grafik nilai harga saham penutup indeks LQ45 periode tanggal 25 Desember 2019 sampai dengan 30 November 2020.

### Preprocessing Diferensiasi Data

Tahapan yang terdapat dalam *preprocessing* kedua yaitu diferensiasi data, proses diferensiasi data dilakukan setelah uji *unit root* menggunakan uji *Augmented Dickey-Fuller* (ADF). Diferensiasi data dilakukan ketika data deret waktu masih menunjukkan non-stasioner. Kriteria yang perlu dipenuhi adalah parameter yang digunakan pada pengujian ADF adalah *Akaike Info Criterion* (AIC) dengan nilai probabilitas (*p-value*) lebih kecil dari 0,05 dan nilai statistik (*t-statistic*) lebih kecil dari nilai kritis pengujian (*t-critical values*) 5% - 1% [16].



**Gambar 2. Grafik Nilai Harga Saham Penutup Indeks LQ45**

Pengujian terdiri dari 3 pilihan diferensiasi dimulai dari *level 0 (Level)*, *level 1 (1st difference)* dan *level 2 (2nd difference)*. Pengecekan dimulai dari level paling awal, yaitu *level 0 (Level)*. Apabila data dianggap sudah stasioner maka proses diferensiasi dapat dihentikan pada level yang sedang diuji.

### **Pendugaan Model dan Estimasi Parameter**

Data deret waktu yang sudah stasioner, seperti pada penelitian bahwa plot *Autocorrelation Function (ACF)* dan *Partial Autocorrelation Function (PACF)* pada *Correlogram* dapat dijadikan sebagai pembentukan model (p,d,q) [17]. Model-model yang ditentukan untuk sementara untuk peramalan model ARIMA dapat diperoleh dari aturan pola ACF dan PACF. Tabel 2 menunjukkan aturan pola ACF dan PACF. Setelah mendapatkan plot *Autocorrelation Function (ACF)* dan *Partial Autocorrelation Function (PACF)* dari data deret waktu yang sudah stasioner dalam pendugaan model, tahap selanjutnya diperlukan analisis regresi untuk mengestimasi suatu parameter yang nilainya tidak diketahui.

Estimasi parameter terbaik untuk model ARIMA didapatkan melalui metode Kuadrat Terkecil (*Least Square*). Metode Kuadrat Terkecil (*Least Square*) adalah

metode yang digunakan untuk mendapatkan nilai-nilai penduga dalam pemodelan regresi yang meminimumkan jumlah kuadrat *error*. Kuadrat Terkecil (*Least Square*) yaitu untuk menganalisis data menggunakan analisis regresi linier.

Uji regresi linier dapat menghasilkan Jumlah Observasi, Identifikasi Variabel, *Adjusted R Square* dan *Standard Error of the regression (S.E. of regression)*.

### **Uji Signifikansi dan Pemeriksaan Diagnostik**

Uji signifikansi yaitu untuk mengetahui parameter yang didapat dari model-model ARIMA sementara apakah signifikan atau tidak. Model signifikan pada plot ACF dan PACF apabila nilai signifikansi (*p-value*) < *alpha* ( $\alpha$ ) dengan nilai  $\alpha$  adalah 0.05 [14]. Pemeriksaan diagnostik yang dilakukan yaitu uji *white noise*, cara melihat proses *white noise* pada model yaitu dengan menggunakan uji statistik *Ljung-Box*. Uji *white noise* suatu model dikatakan baik dengan melihat residual. Residual sudah *white noise* pada plot ACF dan PACF apabila nilai signifikansi (*p-value*) pada *Ljung-Box* > *alpha* ( $\alpha$ ) dengan nilai  $\alpha$  adalah 0,05. Uji *white noise* dengan melihat residual dapat dilihat melalui *Correlogram Diagnostic* [14].

**Tabel 2. Aturan Pola ACF dan PACF**

ACF	PACF	Model Tentatif
<i>Tails off</i>	<i>Cuts off</i> setelah lag q	AR(p)
<i>Cuts off</i> setelah lag p	<i>Tails off</i>	MA(q)
<i>Cuts off</i> setelah lag p	<i>Cuts off</i> setelah lag p	AR(p) atau MA(q), Pilih model terbaik.

## Pemilihan Model

Pemilihan model merupakan tahapan dalam pemilihan model ARIMA terbaik. Tahapan pemilihan model dilakukan setelah melakukan uji *white noise* dalam pemeriksaan diagnostik. Untuk memilih model sebagai model ARIMA terbaik dibutuhkan kriteria pemilihan model. Kriteria pemilihan model terbaik dilihat dari nilai *Adjusted R-squared* yang paling besar, nilai *S.E. of regression*, dan *Akaike Info Criterion* yang paling kecil. Jika model yang diperoleh tidak memenuhi kriteria model terbaik maka perlu kembali ke tahap pendugaan model.

## Model ARIMA, Peramalan, dan Akurasi Model ARIMA

Model ARIMA yang terpilih merupakan satu model yang telah melalui proses-proses pengujian model sampai dengan menentukan model terbaik untuk mengetahui ketepatan sebelum melakukan peramalan. Tahap awal melakukan stasioneritas data dan diferensiasi data, mengidentifikasi model melalui pola ACF dan PACF *Correlogram*, sehingga memperoleh 3 model sementara. 3 model sementara yang diperoleh bertujuan untuk melakukan estimasi parameter terbaik untuk model ARIMA melalui metode Kuadrat Terkecil (*Least Square*) dengan uji regresi linier. Tahap selanjutnya adalah melakukan uji Signifikansi Parameter dan Pemeriksaan Diagnostik, sehingga didapatkan model terbaik. Peramalan (*Forecasting*) adalah

meramalkan beberapa peristiwa atau kegiatan di masa yang akan datang. Peramalan diklasifikasikan menjadi jangka pendek, menengah dan panjang. Peramalan jangka pendek digunakan untuk memprediksi peristiwa hanya beberapa periode waktu (hari, minggu dan bulan) ke masa depan. Peramalan jangka menengah digunakan untuk periode waktu 1 hingga 2 tahun ke depan [13]. Peramalan jangka panjang digunakan untuk periode beberapa tahun. Pada penelitian, pemodelan dan peramalan nilai harga saham penutup indeks LQ45 bersifat jangka pendek selama 2 bulan periode tanggal 7 Desember 2020 sampai 25 Januari 2021. Akurasi model ARIMA merupakan tahapan untuk menganalisis nilai *error* terkecil dari setiap parameter evaluasi akurasi. Parameter yang digunakan untuk menentukan model ARIMA terbaik dalam penelitian adalah nilai *Mean Absolute Percentage Error* (MAPE) terkecil. Dalam menetapkan model dengan nilai MAPE kurang dari 10% dianggap hasil peramalan sangat baik, nilai MAPE kurang dari 20% dianggap hasil peramalan baik, dan nilai MAPE diantara 20% sampai dengan 50% dianggap hasil peramalan cukup baik [13]. Semakin kecil nilai MAPE menunjukkan data hasil peramalan semakin mendekati nilai aktual, sehingga model yang dipilih merupakan model dengan nilai MAPE terkecil [14]. MAPE merupakan perhitungan yang digunakan untuk menghitung rata-rata persentase kesalahan mutlak dalam peramalan, dengan persamaan di bawah ini:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{e_t - Y_t}{e_t} \right| 100}{n} \quad (1)$$

Keterangan :

- n = jumlah data
- $e_t$  = nilai hasil aktual
- $Y_t$  = nilai hasil pendugaan

## HASIL DAN PEMBAHASAN

Pengujian model peramalan dalam penelitian menggunakan harga saham penutup dalam indeks LQ45 yang merupakan data sekunder yang diperoleh dari situs *Yahoo Finance* dengan total data observasi sebanyak 242 data harian, dari hari Senin sampai Jumat tidak termasuk hari libur. Data diubah menjadi data mingguan untuk mendapatkan data deret waktu yang sesuai dengan penelitian, sehingga diperoleh data observasi sebanyak 54 data. Data observasi dalam penelitian merupakan data dari 25 November 2019 sampai dengan 30 November 2020.

## ARIMA

Model ARIMA digunakan untuk menganalisis data deret yang memiliki *trend* dengan jangka waktu pendek untuk menghasilkan nilai prediksi di masa depan [18]. Model ARIMA akan menghasilkan nilai prediksi harga saham penutup dalam indeks LQ45 berdasarkan model terbaik dengan nilai *Mean Absolute Percentage Error* (MAPE) terkecil.

### Hasil Identifikasi Stasioneritas Data dan Diferensiasi Data

Uji *unit root* menggunakan uji *Augmented Dickey-Fuller* (ADF) untuk mengidentifikasi stasioneritas data dan melakukan diferensiasi data.

Identifikasi stasioneritas data dan melakukan diferensiasi data dapat menggunakan *p-value* dan *t-statistic* yaitu membandingkan *p-value* dan *t-statistic* dengan tingkat signifikansi atau  $\alpha$  seperti pada Gambar 3.

Augmented Dickey-Fuller Unit Root Test on HARGA_PENUTUP		
Null Hypothesis: HARGA_PENUTUP has a unit root		
Exogenous: Constant		
Lag Length: 1 (Automatic - based on AIC, maxlag=15)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.149160	0.2270
Test critical values:	1% level	-3.562669
	5% level	-2.918778
	10% level	-2.597285
*Mackinnon (1996) one-sided p-values.		

(a)

Augmented Dickey-Fuller Unit Root Test on D(HARGA_PENUTUP)		
Null Hypothesis: D(HARGA_PENUTUP) has a unit root		
Exogenous: Constant		
Lag Length: 1 (Automatic - based on AIC, maxlag=15)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.836310	0.0047
Test critical values:	1% level	-3.565430
	5% level	-2.919952
	10% level	-2.597905
*Mackinnon (1996) one-sided p-values.		

(b)

Gambar 3. Hasil Uji Unit Root dengan ADF (a) Hasil Uji Level 0 (b) Hasil Uji 1st Difference

Dilihat dari Gambar 3. (a) bahwa proses uji unit *root* pada *Level 0* menghasilkan persamaan  $p\text{-value } 0.2270 > 0.05$  dengan nilai  $\alpha = 0.05$  dan persamaan  $t\text{-statistic } -2.149160 > -2.918778$  maka  $H_0$  ditolak artinya variabel harga\_penutup tidak signifikan terhadap indeks LQ45 atau dapat dikatakan data belum stasioner. Parameter pengujian unit *root* meliputi tipe pengujian menggunakan *Augmented Dickey-Fuller*, pengujian untuk unit *root* pada *Level* ( $d=0$ ), dan pengujian model deret waktu dengan tipe intersep untuk menguji *trend* yang ada. Pengujian unit *root* berdasarkan parameter *Akaike Info Criterion* dengan maksimum *lag* sebanyak 15 data Gambar 3. (b) menunjukkan proses uji unit *root* pada *1st Difference* ( $d=1$ ) menghasilkan persamaan  $p\text{-value } 0.0047 < 0.05$  dengan nilai  $\alpha = 0.05$  dan persamaan  $t\text{-statistic } -3.836310 < -2.919952$  maka variabel harga\_penutup signifikan terhadap indeks LQ45 atau dapat dikatakan data sudah stasioner. Parameter pengujian unit *root* meliputi tipe pengujian menggunakan *Augmented Dickey-Fuller*, pengujian untuk unit *root* pada *1st Difference* ( $d=1$ ), dan pengujian model deret waktu dengan tipe intersep untuk menguji *trend* yang ada. Pengujian unit *root* berdasarkan parameter *Akaike Info Criterion* dengan maksimum *lag* sebanyak 15 data. Berdasarkan hasil uji *unit root* menggunakan *Augmented Dickey-Fuller* (ADF), untuk data nilai harga saham penutup dalam indeks LQ45 telah memenuhi uji stasioner adalah *1st Difference* ( $d=1$ ) dengan

$p\text{-value } 0.0047 < 0.05$  dan  $t\text{-statistic } -3.836310 < -2.919952$ . Tujuan mengetahui asumsi yang terpenuhi dalam uji *unit root* menggunakan *Augmented Dickey-Fuller* (ADF) pada data nilai harga saham penutup dalam indeks LQ45 yaitu untuk mengidentifikasi stasioneritas data. Tahapan selanjutnya adalah pengujian *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) dengan melihat *lag* yang melewati batas 0. Pengujian ACF dan PACF dilakukan melalui *correlogram* dengan diferensiasi data deret yang sudah stasioner yaitu diferensiasi level 1 seperti terlihat pada Gambar 4. Gambar 4 merupakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) yang mengalami *cuts off* (turun drastis) pada baris pertama seperti ditunjukkan pada kotak merah yaitu pada *lag 1*. Jika *cuts off* pada ACF dan PACF maka kemungkinan pertama  $p=1$  dan  $q=0$  kemudian kemungkinan kedua  $p=0$  dan  $q=1$ . Pada kondisi *Correlogram* Gambar 4 sudah menunjukkan data stasioner, sehingga jika digabung dengan  $d=1$  (*1st Difference*) yang sudah diketahui nilainya. Kemungkinan model ARIMA( $p,d,f$ ) terbentuk 3 model yaitu ARIMA(1,1,0) dengan parameter AR(1), model ARIMA(0,1,1) dengan parameter MA(1) dan ARIMA(1,1,1) dengan parameter ARMA(1,1).

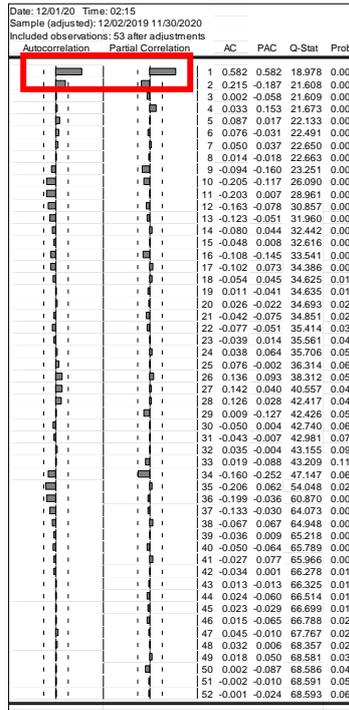
### Hasil Estimasi Model

Tahapan untuk menghasilkan estimasi model dengan menguji ketiga model dugaan

yaitu ARIMA(1,1,0), ARIMA(0,1,1) dan ARIMA(1,1,1).

Hasil dari pengujian ketiga model dugaan dapat terlihat pada Gambar 5. Gambar 5 menunjukkan hasil estimasi 3 model

sementara yaitu ARIMA(1,1,0), ARIMA(0,1,1), dan ARIMA(1,1,1) yang dijelaskan dengan 3 parameter penentu yaitu *Adjusted R-squared*, nilai *S.E. of regression* dan *Akaike Info Criterion*.



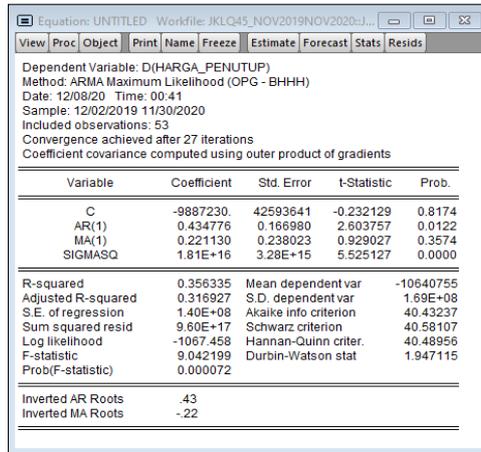
Gambar 4. Hasil Correlogram Harga Saham Penutup Indeks LQ45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10196526	31453609	-0.324177	0.7472
MA(1)	0.515129	0.134735	3.823272	0.0004
SIGMASQ	1.97E+16	3.78E+15	5.202067	0.0000
R-squared	0.301144	Mean dependent var		-10640755
Adjusted R-squared	0.273190	S.D. dependent var		1.69E+08
S.E. of regression	1.44E+08	Akaike info criterion		40.47436
Sum squared resid	1.04E+18	Schwarz criterion		40.58558
Log likelihood	-1059.570	Hannan-Quinn criter.		40.51725
F-statistic	10.77275	Durbin-Watson stat		1.676554
Prob(F-statistic)	0.000129			
Inverted MA Roots	-0.52			

(a)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9867266	46419499	-0.212567	0.8325
AR(1)	0.572640	0.072782	7.867845	0.0000
SIGMASQ	1.86E+16	3.20E+15	5.805588	0.0000
R-squared	0.339267	Mean dependent var		-10640755
Adjusted R-squared	0.312837	S.D. dependent var		1.69E+08
S.E. of regression	1.40E+08	Akaike info criterion		40.41994
Sum squared resid	9.85E+17	Schwarz criterion		40.53147
Log likelihood	-1068.128	Hannan-Quinn criter.		40.46283
F-statistic	12.83675	Durbin-Watson stat		1.764483
Prob(F-statistic)	0.000032			
Inverted AR Roots	.57			

(b)



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9887230	42593641	-0.232129	0.8174
AR(1)	0.434776	0.166980	2.603757	0.0122
MA(1)	0.221130	0.238023	0.929027	0.3574
SIGMASQ	1.81E+16	3.28E+15	5.525127	0.0000

R-squared	0.358335	Mean dependent var	-10640755
Adjusted R-squared	0.316927	S.D. dependent var	1.59E+08
S.E. of regression	1.40E+08	Akaike info criterion	40.43237
Sum squared resid	9.60E+17	Schwarz criterion	40.58107
Log likelihood	-1067.458	Hannan-Quinn criter.	40.48956
F-statistic	9.042199	Durbin-Watson stat	1.947115
Prob(F-statistic)	0.000072		

Inverted AR Roots	43
Inverted MA Roots	-22

(c)

Gambar 5. Hasil Estimasi Model (a) ARIMA(1,1,0) (b) ARIMA(0,1,1) (c) ARIMA(1,1,1)

Tabel 3. Estimasi Model ARIMA Data Nilai Harga Saham Penutup Indeks LQ45

Estimasi Model	Adjusted R-squared	S.E. of regression	Akaike Info Criterion
ARIMA(1,1,0)	0.312837	1.40E+08	<b>40.41994</b>
ARIMA(0,1,1)	0.273190	1.44E+08	40.47436
ARIMA(1,1,1)	<b>0.316927</b>	1.40E+08	40.43237

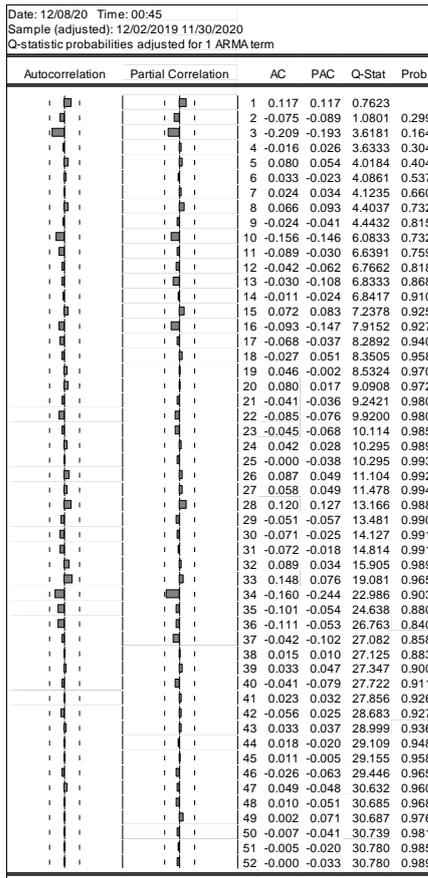
Tabel 3 menunjukkan 3 jenis nilai parameter yang dihasilkan dari setiap estimasi model ARIMA data nilai harga saham penutup indeks LQ45. Penentuan estimasi model terbaik dilihat dari nilai *Adjusted R-squared* yang paling besar, nilai *S.E. of regression* dan *Akaike Info Criterion* yang paling kecil. Berdasarkan ketentuan pada tabel 3, model ARIMA(1,1,1) memiliki nilai *Adjusted R-squared* paling besar diantara 3 model ARIMA yang terbentuk dan model ARIMA(1,1,0) memiliki nilai *Akaike Info*

*Criterion* paling kecil diantara 3 model ARIMA yang terbentuk.

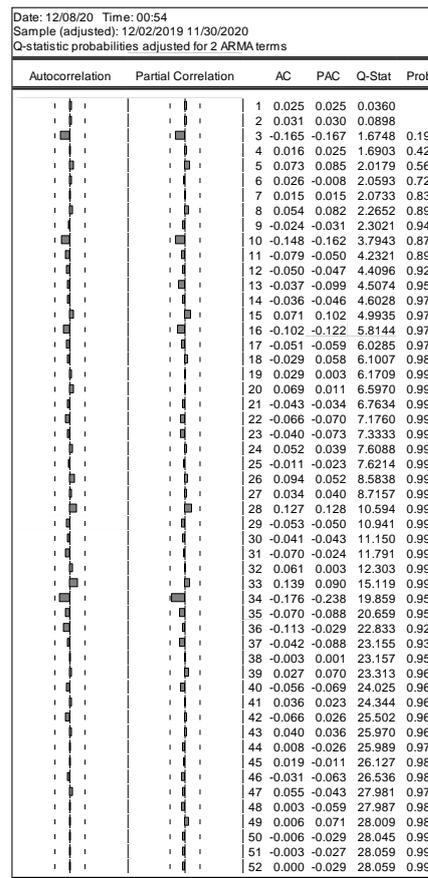
Berdasarkan ketentuan tersebut, pada tabel 4 terdapat 2 model yaitu model ARIMA(1,1,0) dan ARIMA(1,1,1). Hasil tersebut didapatkan dari nilai *Adjusted R-squared* yang paling besar, nilai *S.E. of regression* dan *Akaike Info Criterion* yang paling kecil. Pada tahap penentuan estimasi model ARIMA, sehingga dapat digunakan pada proses pemeriksaan diagnostik residual untuk mendapatkan model ARIMA terbaik.

**Tabel 4. Penentuan Estimasi Model ARIMA Data Nilai Harga Saham Penutup Indeks LQ45**

Estimasi Model	Adjusted R-squared	S.E. of regression	Akaike Info Criterion
ARIMA(1,1,0)	0.312837	1.40E+08	<b>40.41994</b>
ARIMA(1,1,1)	<b>0.316927</b>	1.40E+08	40.43237



(a)



(b)

**Gambar 6. Hasil Correlogram Residual (a) ARIMA(1,1,0) (b) ARIMA(1,1,1)**

**Hasil Uji Signifikansi dan Pemeriksaan Diagnostik**

Tujuan dari uji signifikansi yaitu untuk mengetahui parameter yang didapat dari model-model ARIMA sementara apakah signifikan atau tidak. Model dianggap signifikan apabila nilai signifikansi (*p-value*)

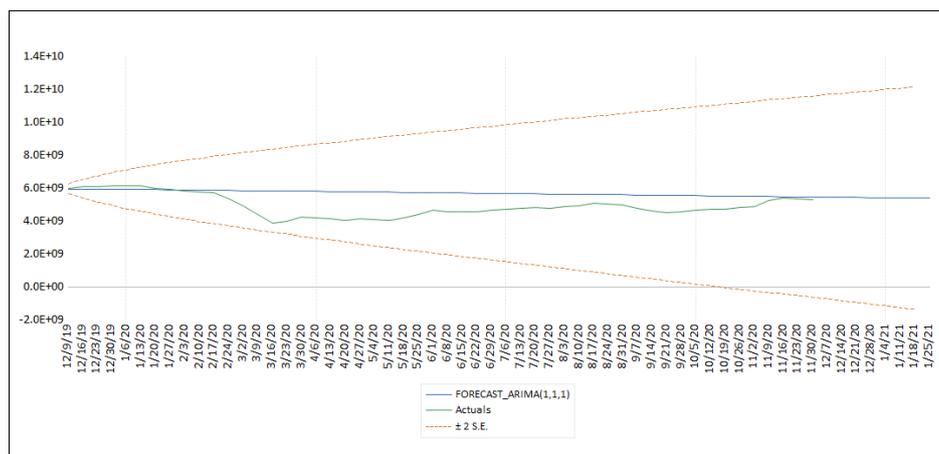
<  $\alpha$  ( $\alpha$ ) dengan nilai  $\alpha$  adalah 0.05 [14].

Gambar 6 menunjukkan model ARIMA(1,1,0) dan ARIMA(1,1,1) di lag ke-1 sampai dengan lag ke-52 tidak memiliki (*p-value*) yang kurang dari 0.05 artinya tidak ada lag yang signifikan. Bila ACF dan PACF tidak signifikan pada uji signifikansi, hal ini

mengindikasikan bahwa adanya residual *white noise* yang berarti hipotesis sudah terpenuhi dan model-model tersebut dapat dilakukan ke pengujian selanjutnya. Residual *white noise* terjadi jika nilai signifikansi (*p-value*) pada Ljung-Box  $>$   $\alpha$  ( $\alpha$ ) dengan nilai  $\alpha$  adalah 0.05. Berdasarkan Gambar 6 pada model ARIMA(1,1,0) dan ARIMA(1,1,1) di *lag* ke-1 sampai dengan *lag* ke-52 memiliki (*p-value*) yang melebihi dari 0.05, (*p-value*). Hal ini menunjukkan bahwa residual model ARIMA(1,1,0) dan ARIMA(1,1,1) sudah *white noise* dan terdistribusi dengan baik, sehingga kedua model tersebut tepat untuk melakukan peramalan. Semua model sudah memenuhi hipotesis uji signifikansi dan melewati pemeriksaan diagnostik, kemudian model ARIMA(1,1,0) dan ARIMA(1,1,1) sebagai model yang tepat untuk melakukan peramalan.

## Hasil Peramalan

Berdasarkan tahapan pengujian yang sudah dilakukan, diperoleh hanya ada 1 model yang dianggap sebagai model ARIMA terbaik untuk meramalkan nilai harga saham penutup dalam indeks LQ45 yaitu model ARIMA(1,1,1). Gambar 7 menampilkan grafik nilai peramalan harga penutup indeks saham LQ45 dengan peramalan jangka pendek 2 bulan periode 7 Desember 2020 sampai 25 Januari 2021 menghasilkan model ARIMA(1,1,1) berdasarkan data harian yang diakumulasi menjadi data mingguan dari 25 Desember 2019 sampai dengan 30 November 2020. Garis merah merupakan batas atas dan batas bawah dalam peramalan. Garis biru merupakan garis peramalan dengan membentuk *trend* yang lurus cenderung mendekati ke arah yang sama sejalan dengan garis aktual yaitu pada garis hijau.



**Gambar 7. Grafik Forecast dan Evaluasi Data Nilai Peramalan Harga Penutup Indeks LQ45**

**Tabel 5. Hasil Pemeriksaan Error Model ARIMA**

Model ARIMA	MAPE
ARIMA(1,1,0)	18.71530
ARIMA(1,1,1)	<b>18.41269</b>

**Tabel 6. Hasil Peramalan pada Data Nilai Harga Penutup Indeks Saham LQ45**

Periode	Harga Penutup (Rp)
12/7/2020	5.443.582.995
12/14/2020	5.433.695.764
12/21/2020	5.423.808.534
12/28/2020	5.413.921.304
1/4/2021	5.404.034.074
1/11/2021	5.394.146.844
1/18/2021	5.384.259.614
1/25/2021	5.384.259.614

Tabel 5 menunjukkan nilai peramalan data harga penutup indeks saham LQ45 dengan peramalan jangka pendek 2 bulan periode 7 Desember 2020 sampai 25 Januari 2021 menghasilkan pemeriksaan *error* pada akurasi model ARIMA menggunakan *Mean Absolute Percentage Error* (MAPE). Nilai *Mean Absolute Percentage Error* (MAPE) digunakan untuk mengetahui ketepatan peramalan pada data uji. Hasil evaluasi model ARIMA(1,1,1) menggunakan nilai *Mean Absolute Percentage Error* (MAPE) membuktikan bahwa model yang dipilih memiliki tingkat akurasi peramalan yang menghasilkan data aktual yang mendekati nilai *Mean Absolute Percentage Error* (MAPE) yang paling kecil yaitu 18.41269. Model dengan nilai MAPE kurang dari 10% dianggap hasil peramalan sangat baik, nilai MAPE kurang dari 20% dianggap hasil peramalan baik, dan nilai MAPE diantara

20% sampai dengan 50% dianggap hasil peramalan cukup baik [13]. Tabel 6 menunjukkan hasil peramalan pada data nilai harga penutup indeks saham LQ45, peramalan ARIMA yang dilakukan adalah peramalan jangka pendek 2 bulan periode 7 Desember 2020 sampai 25 Januari 2021.

## KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian pada data nilai harga saham penutup dalam indeks saham LQ45 menunjukkan model peramalan terbaik adalah model ARIMA(1,1,1). Model ARIMA(1,1,1) terpilih karena memenuhi asumsi dan didukung oleh nilai *Adjusted R-squared* yang paling besar, nilai *S.E. of regression*, dan *Akaike Info Criterion* yang paling kecil. Hasil peramalan jangka pendek selama 2 bulan ke depan yaitu pada periode 7 Desember 2020

sampai 25 Januari 2021 dengan model ARIMA(1,1,1) berdasarkan data harian yang diakumulasi menjadi data mingguan dari 25 Desember 2019 sampai dengan 30 November 2020. Hasil evaluasi model ARIMA(1,1,1) menggunakan nilai *Mean Absolute Percentage Error* (MAPE) membuktikan bahwa model yang dipilih memiliki tingkat akurasi peramalan yang menghasilkan data aktual yang mendekati nilai *Mean Absolute Percentage Error* (MAPE) yang paling kecil yaitu 18.41269.

Beberapa saran yang diharapkan untuk penelitian selanjutnya menjadi lebih baik, antara lain pemodelan dan peramalan pada nilai harga saham penutup dalam indeks saham LQ45 bersifat jangka pendek, sehingga perlu dikembangkan lebih lanjut dalam hal pemodelan jangka menengah ataupun jangka panjang. Penelitian dapat dikembangkan dengan model lain yang menghasilkan nilai kesalahan prediksi yang lebih kecil serta peramalan yang lebih baik.

## DAFTAR PUSTAKA

- [1] M. Abubakar, Belajar Investasi Saham: Low Risk For Maximum Profit In Indonesia Stock Exchange, Jakarta: BSK Capital, 2020.
- [2] Kasmir, Studi Kelayakan Bisnis, Jakarta: Prenada Media, 2015.
- [3] S. Luo, F. Yan, D. Lai, W. Wui and F. Lu, "Using ARIMA Model to Fit and Predict Index of Stock Price Based on Wavelet De-Noising," *International Journal of Science and Technology*, vol. 9, pp. 317-326, 2016.
- [4] M. Almasarweh and S. A. Wadi, "ARIMA Model in Predicting Banking Stock Market Data," *Modern Applied Science*, vol. 12, 2018.
- [5] Z. Amry and B. H. Siregar, "ARIMA Model Selection for Composite Stock Price Index in," *International Journal of Accounting and Finance Studies*, vol. 2, 2019.
- [6] M. Ashik and S. Kannan., "Forecasting Nifty Bank Sectors Stock Price Using ARIMA Model," *International Journal of Creative Research Thoughts*, vol. 5, no. 4, 2017.
- [7] B. Dhyani, M. Kumar, P. Verma and A. Jain, "Stock Market Forecasting Technique using Arima Model," *International Journal of Recent Technology and Engineering*, vol. 8, no. 6, 2020.
- [8] A. Edward and JyothiManoj, "Forecast Model Using ARIMA for Stock Prices of Automobile Sector," *International Journal of Research in Finance and Marketing*, vol. 6, no. 4, 2016.
- [9] A. Jadhav and K.B.Kamble, "Prediction of Stock prices in Oil Sectors using ARIMA Model," *International Journal of Mathematics Trends and Technology*, vol. 51, no. 4, 2017.

- [10] R. Jin, S. Wang, F. Yan and J. Zhu, "The Application of ARIMA Model in 2014 Shanghai," *Science Journal of Applied Mathematics and Statistics*, vol. 3, pp. 199-203, 2015.
- [11] S. A. Wadi, M. Almasarweh and A. A. Alsarairh, "Predicting Closed Price Time Series Data Using ARIMA Model," *Modern Applied Science*, vol. 12, 2018.
- [12] S. T. Wahyudi, "The ARIMA Model for the Indonesia Stock Price," *International Journal of Economics and Management*, vol. 11, pp. 223 – 236 , 2017.
- [13] D. C. Montgomery, C. L. Jennings and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, 2nd Edition ed., New Jersey: John Wiley & Sons, Inc, 2015.
- [14] P. Ramos, N. Santos and R. Rebelo, "Performance of State Space and ARIMA Models for Consumer Retail," *Robotics and Computer-Integrated Manufacturing*, vol. 34, pp. 151–163, 2015.
- [15] S. Makridakis, E. Spiliotis and V. Assimakopoulo, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *PLoS ONE*, vol. 13, no. 3, 2018.
- [16] "LQ45 Index (^JKLQ45)," 2020. [Online]. Available: <https://finance.yahoo.com/quote/%5EJKLQ45/history?p=%5EJKLQ45&guccounter=1>.