**Penggunaan ARIMA untuk Memprediksi Angka Penumpang Kereta di Pulau Jawa**

Dimas Wahyu Saputro1

Dhea Sukma Agustiana1

Fadia Dilla Sabine1

Lis Nurani1Rezki P. Manullang1Taj Shavira H1

1Sains Data, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

| **ABSTRACT** | **KEYWORDS** |
| --- | --- |
| Kereta api pada saat ini telah menjadi salah satu transportasi umum yang paling populer untuk perjalanan jarak menengah dan jauh. Saat musim liburan, jumlah penumpang kereta api sulit diprediksi. Kereta api sudah tersebar di Pulau Jawa dan Sumatra, bahkan menjadi salah satu transportasi yang paling diminati di pulau tersebut. Tujuan dari penelitian ini adalah untuk memprediksi jumlah penumpang kereta api dengan metode Autoregressive Integrated Moving Average (ARIMA). Tahapan yang digunakan dalam penelitian ini meliputi: (1) persiapan dataset, (2) data preprocessing, dan (3) pengujian eksperimental dan metode. Model ARIMA yang didapat adalah ARIMA(2,1,0)(0,1,2)[12] dengan nilai AIC 2379.265. Dilakukan *model diagnostic* dan didapatkan bahwa model cukup baik karena residual berpusat di sekitar nilai nol, sehingga metode ARIMA yang digunakan dalam memprediksi penumpang adalah akurat. | Penumpang; Kereta Api; ARIMA; Pulau Jawa. |

# Introduction

Kereta api merupakan salah satu sarana transportasi yang banyak digunakan oleh masyarakat untuk bepergian lintas kota ataupun lintas provinsi. Kereta api sudah tersebar di Pulau Jawa dan Sumatra, bahkan menjadi salah satu transportasi yang paling diminati di pulau tersebut. Di Pulau Jawa sendiri, berdasarkan laporan Kementerian Perhubungan (Kemenhub), pada tahun 2020 terdapat sekitar 463 stasiun yang melayani naik/turunnya penumpang kereta api (Rizaty, 2021). Seiring dengan banyaknya peminat transportasi kereta api terutama di Pulau Jawa, maka prediksi dari angka penumpang kereta api menjadi sangat penting. Salah satu metode statistik yang akan diterapkan untuk mengetahui prediksi dari angka penumpang kereta api di Pulau Jawa adalah Metode ARIMA atau *Autoregressive Integrated Moving Average.* Model Autoregressive Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. Model ARIMA merupakan salah satu teknik peramalan series yang hanya didasarkan pada perilaku yang diamati dari data variabel. ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang.Model Autoregressive Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent). Dengan mengacu pada permasalahan diatas, penelitian ini akan memprediksi jumlah penumpang kereta api di pulau Jawa dengan menggunakan metode ARIMA. Metode ini dapat memprediksi data yang besar, menghasilkan prediksi jangka pendek yang akurat, efisien dalam memprediksi time series, serta mampu menghadapi trend data musiman dimana jumlah penumpang selalu berubah setiap harinya.

# Methods

## **Stages of Study**

Penelitian ini menggunakan metode eksperimen, meliputi (1) pengumpulan data, (2) data preprocessing, dan (3) tahap menggunakan model prediksi.

## **Data Collecting**

Dalam penelitian ini, data yang digunakan merupakan data Jumlah Penumpang Kereta Api di Pulau Jawa. Data ini berasal dari data yang telah dikumpulkan dan di rekap oleh Badan Pusat Statistik Indonesia. Jumlah data yang digunakan sebanyak 168 data terhitung mulai bulan Januari 2006 sampai dengan Desember 2019. Dataset terdiri dari satu kolom, yaitu kolom jumlah penumpang dalam satuan ribu.

## **ARIMA**

Penelitian ini menggunakan metode ARIMA. ARIMA (Autoregressive Integrated Moving Average) merupakan metode yang mengabaikan independen variabel dalam pembuatan peramalan. Peramalan yang digunakan atau dipakai merupakan sebuah parameter untuk perkiraan atas suatu kasus tertentu yang terjadi. Metode ARIMA dibuat berdasarkan pada nilai-nilai suatu peubah yang telah terjadi pada masa lampau. ARIMA pada dasarnya yang menghasilkan suatu rapalan atau prediksi merupakan suatu pemodelan secara sintesis berdasarkan pada pola data yang telah didapatkan secara historis. Kemudian, data pada masa lampau tersebut digunakan untuk membuat/menentukan pola-pola historis data yang kemudian akan digunakan untuk mengekstrapolasikan pola tersebut di masa yang akan datang (Wulandari & Gernowo, 2019, 41-48).

ARIMA juga biasa disebut sebagai metode runtun waktu Box-Jenkins. Model ARIMA dibagi ke dalam 3 kelompok, di antaranya:

1) *Autoregressive Model* (AR), bentuk umum *autoregressive* dengan ordo p (AR(p)) atau model ARIMA (p,0,0) dinyatakan sebagai berikut:



2) *Moving Average* (MA), bentuk umum model *moving average* ordo q (MA(q)) atau ARIMA (0,0,q) dinyatakan sebagai berikut:



3) Model campuran, model umum untuk campuran proses AR(1) murni dan MA(1) murni, misal ARIMA (1,0,1) dinyatakan sebagai berikut:

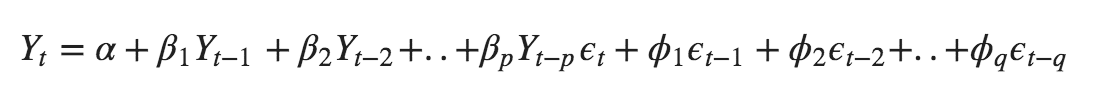


atau



## **Pemodelan ARIMA**

Pada pemodelan ARIMA ini metode yang digunakan pada ARIMA dengan derajat AR (*p)* , derajat selisihnya (*d),* dan derajat MA (q), maka permodelan yang ditulis dalam ARIMA dapat dituliskan dengan (p, d, q) yang memiliki bentuk umum sebagai berikut (William W. S. Wei, 2006) :



Pemodelan dari ARIMA pada dasarnya dibagi lagi menjadi beberapa tahapan diantaranya yaitu: Identifikasi model , penaksiran dari parameternya, uji untuk model dan diagnostik dari ARIMA tersebut.

**Identifikasi Pemodelan**

Proses identifikasi dari model musiman tergantung pada alat-alat statistik berupa autokorelasi dan parsial autokorelasi, serta pengetahuan terhadap sistem (atau proses) yang dipelajari.

Terdapat beberapa langkah tahapan untuk melakukan identifikasi pemodelan diantaranya sebagai berikut:

1. Pembuatan plot untuk melakukan pengecekan dari pola musiman tersebut masuk pada pola musiman atau tidak
2. Melakukan stasioneritas pada data. Fungsi dari adanya stasioneritas pada data yaitu untuk melihat apakah data tersebut stasioner dalam variasi nya atau stasioner dalam mean nya. Dalam hal ini jika stasioner terjadi dalam rata-rata nya maka harus dilakukan differencing dan sebaliknya jika tidak stasioner dalam variasi nya maka harus dilakukan transformasi.
3. Membuat plot pada ACF dan PACF. Hal ini dilakukan untuk melihat apakah data yang dibuat tersebut stasioner setelah dilakukannya proses differencing atau setelah dilakukannya proses transformasi. Dan kegunaan lainnya dari pembuatan plot pada ACF dan PACF ini yaitu untuk meramalkan model yang memungkinkan dan sesuai dengan data yang digunakan.

## **Musiman dan Model ARIMA**

Musiman didefinisikan sebagai suatu pola yang berulang-ulang dalam selang waktu yang tetap. Untuk data yang stasioner, faktor musiman dapat ditentukan dengan mengidentifikasi koefisien autokorelasi pada dua atau tiga time-lag yang berbeda nyata dari nol. Autokorelasi yang secara signifikan berbeda dari nol menyatakan adanya suatu pola dalam data. Untuk mengenali adanya faktor musiman, seseorang harus melihat pada autokorelasi yang tinggi. Untuk menangani musiman, notasi umum yang singkat adalah:

**ARIMA (p,d,q) (P,D,Q)s**

dimana:

(p,d,q) = bagian yang tidak musiman dari model

(P,D,Q) = bagian musiman dari model

S = jumlah periode per musim

**Penaksiran Parameter**

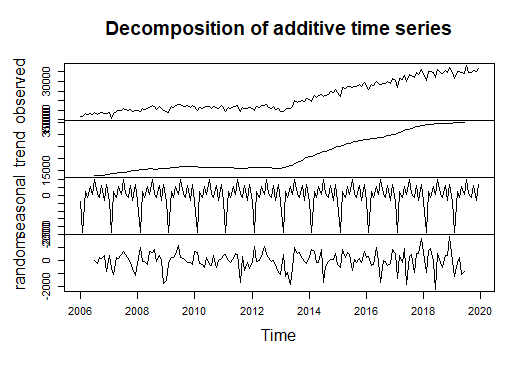
Ada dua cara yang mendasar untuk mendapatkan parameter-parameter tersebut:

1. Dengan cara mencoba-coba (trial and error), menguji beberapa nilai yang berbeda dan memilih satu nilai tersebut (atau sekumpulan nilai, apabila terdapat lebih dari satu parameter yang akan ditaksir) yang meminimumkan jumlah kuadrat nilai sisa (*sum of squared residual*).
2. Perbaikan secara iteratif, memilih taksiran awal dan kemudian membiarkan program komputer memperhalus penaksiran tersebut secara iteratif.

# Results and Discussion

1. Decomposition Time Series

*Decomposition Time Series* adalah mengurai deret waktu dan memisahkannya menjadi komponen penyusunnya, yang sering kali merupakan komponen tren, komponen acak, dan komponen musiman. Pada pemrograman R, digunakan perintah *plot(decompose(data\_ts)).* Setelah baris kode dijalankan, akan menghasilkan plot seperti pada Gambar 2. Jelas bahwa deret waktu tidak stasioner karena terdapat efek musiman dan trend.



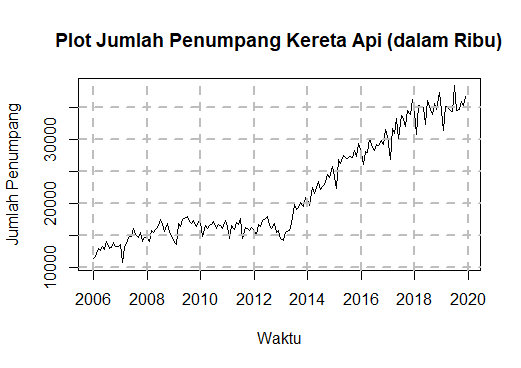
**Gambar 2**. Decomposition Time Series

1. Stasionary Test

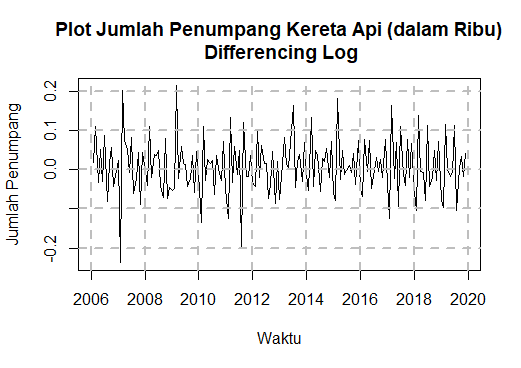
Stasioner adalah apabila suatu data deret waktu memiliki rata-rata dan memiliki kecenderungan bergerak menuju rata-rata. Pada dataset Jumlah Penumpang Kereta Api di Pulau Jawa awalnya tidak stasioner. Hal ini tercermin dari grafik pada Gambar 3. Oleh karena itu, perlu dilakukan *differencing.* Differencing yang digunakan adalah *differencing log*. Sebelum dilakukan differencing, dilakukan *ADF Test. Augmented Dickey-Fuller (ADF)* adalah pengujian yang dilakukan terhadap data deret waktu (time series) untuk mengetahui apakah data deret waktu tersebut stasioner atau tidak.

Saat dilakukan *ADF Test*, didapatkan nilai *p-value* = 0.7457. Karena *p-value* tidak kurang dari 0.05, artinya gagal menolak hipotesis nol. Ini berarti deret waktu tidak stasioner. Dengan kata lain, ia memiliki beberapa struktur yang bergantung pada waktu dan tidak memiliki varians yang konstan dari waktu ke waktu. Hal ini sesuai dengan Gambar 3.

Setelah dilakukan *differencing log,* didapatkan nilai p-value = 0.01. Keputusan dari hasil di atas adalah karena p-value lebih kecil dari 0.05 (tingkat signifikansi) maka tolak H0, artinya, data sudah stasioner. Hal ini sesuai dengan Gambar 4. Stasioner ditandai dengan rata-rata dan varians dari data Jumlah Penumpang Kereta Api di Pulau Jawa adalah konstan sepanjang waktu.



**Gambar 3**. Jumlah Penumpang Kereta Api

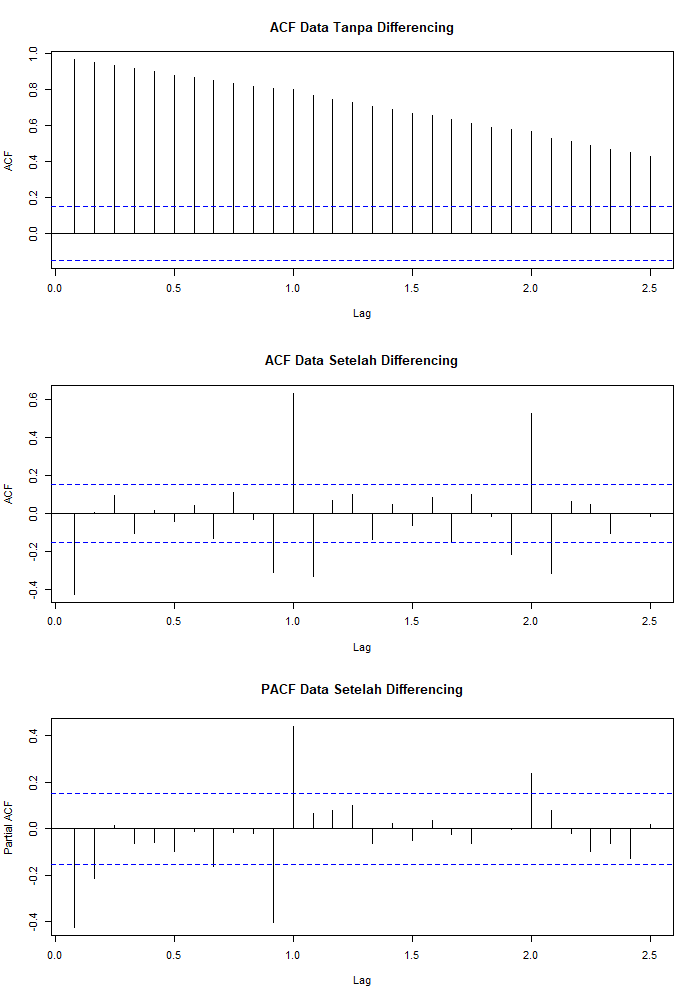


**Gambar 4**. Jumlah Penumpang Kereta Api Setelah Differencing

1. Determine of Order

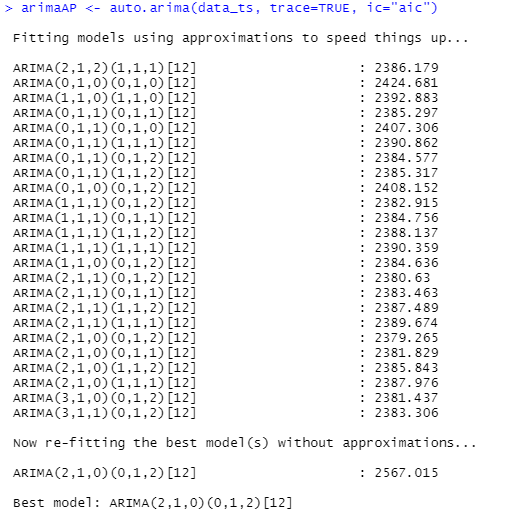
Setelah data dipastikan stasioner, akan dilakukan estimasi model Arima. Cara estimasinya adalah dengan cara mengamati ACF dan PACF. ACF merupakan singkatan dari *Autocorrelation Function*, yang mengindikasikan nilai *autoregressive*. PACF merupkan singkatan *Parcial Autocorrelation Function*, yang mengindikasikan nilai *Moving Averagenya*. Dalam menentukan data mengandung *Autoregressive (AR)* atau *Moving Average (MA)* adalah melihat pola atau perilaku ACF dan PACFnya. Hal yang perlu dipahami dalam bagian ini adalah harus mengerti mana pola yang dikatakan *Cut Off*, dan mana pola yang dikatakan *dying down.*

Penentuan metode ARIMA terbaik adalah dengan melihat plot ACF dan PACF setelah dilakukan *differencing log*. Hal ini ditunjukkan pada Gambar 5.



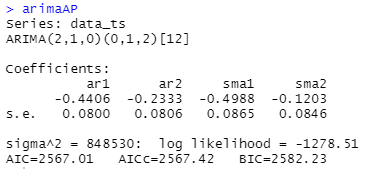
**Gambar 5**. ACF dan PACF

Nilai dari p, d, dan q dapat dilihat dari plot ACF dan PACF. Namun, pada bahasa R, terdapat library supaya dapat mendapatkan nilai tersebut yang akan digunakan untuk ARIMA. *Auto Arima* merupakan salah satu *package* atau *library* dalam R yang memudahkan pengguna sehingga tidak perlu secara manual mencari model ARIMA terbaik. Untuk menggunakan *Auto Arima,* pengguna diharuskan mengimport *library Forecast.*



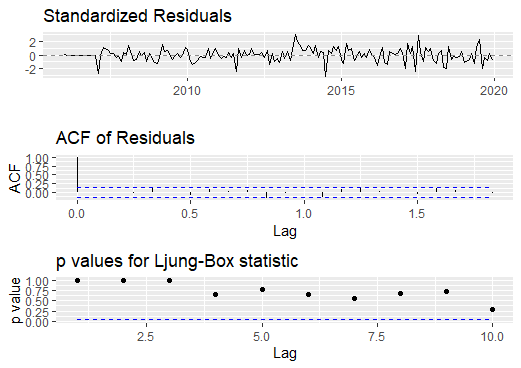
**Gambar 6**. Nilai AIC dari Setiap Model

Ketika menggunakan *Auto Arima*, akan ditampilkan nilai AIC. Setiap model yang digunakan dapat diukur atau dibandingkan dengan melihat *Akaike Information Criterion* (AIC). AIC adalah metode umum untuk menentukan seberapa baik suatu model cocok dengan data. Nilai AIC terkecil adalah model fitting terbaik. Jika dilihat dari data di atas, AIC terendah berada ketika ARIMA(2,1,0)(0,1,2)[12] dengan nilai AIC 2379.265. Oleh karena itu, model ARIMA di atas akan digunakan untuk peramalan sesuai dengan Gambar 6.



**Gambar 7**. Nilai AIC, AICC, dan BIC dari Model Terbaik

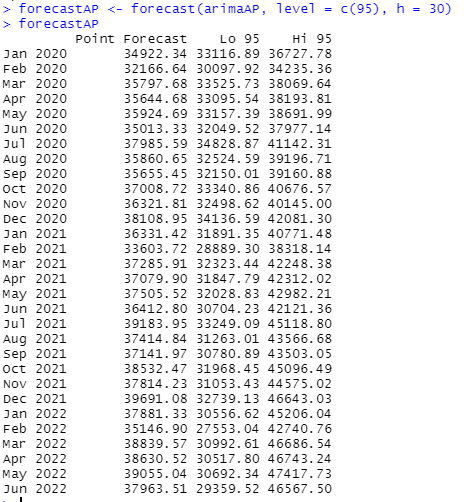
Setelah mendapatkan nilai model terbaik, dilakukan *model diagnostic.* Untuk melakukan *diagnostic* untuk residual dan acf, akan digunakan *library ggfortif.* Terlihat pada Gambar 8 model ARIMA di atas cukup baik karena residual berpusat di sekitar nilai nol.

**

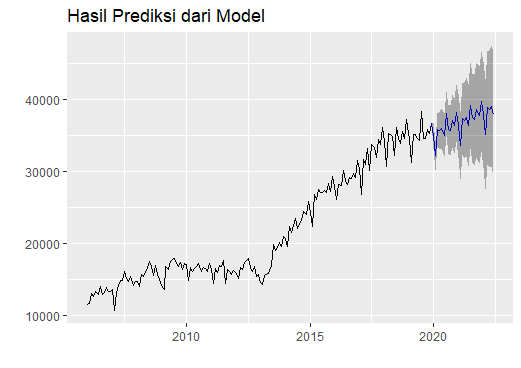
**Gambar 8.**  Model Diagnostic

1. Forecast

Tes prediksi dilakukan dengan menentukan jumlah hari yang akan diprediksi. Dalam hal ini, pengujian dilakukan untuk memprediksi jumlah penumpang kereta selama 30 bulan ke depan Hasil prediksi menggunakan ARIMA dapat dilihat pada Gambar 9.



**Gambar 9.**  Prediksi yang Didapatkan dari Model



**Gambar 10.**  Plot Prediksi

# Conclusion

Berdasarkan hasil dan pembahasan di atas, maka dapat diperoleh model terbaik untuk memprediksi jumlah penumpang kereta api di pulau jawa adalah dengan menggunakan model ARIMA (2,1,0)(0,1,2)[12] dengan nilai AIC 2379.265. Meskipun tidak menggunakan terlalu banyak data, 2006-2020, namun didapatkan hasil yang bagus.

# References

Chen, C.-F., Chang, Y.-H., & Chang, Y.-W. (2009). Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. *Transportmetrica*, *5*(2), 125–140. <https://doi.org/10.1080/18128600802591210>

Rahmani, F., & Fattahi, M. H. (2021). Investigation of denoising effects on forecasting models by statistical and nonlinear dynamic analysis. *Journal of Water and Climate Change*, *12*(5), 1614–1630. https://doi.org/10.2166/wcc.2020.014

Rizaty, M. A. (2021, November 4). *Berapa Jumlah Stasiun Kereta Api Penumpang di Jawa dan Sumatera? Databoks*. Databoks. Retrieved May 9, 2022, from <https://databoks.katadata.co.id/datapublish/2021/11/04/berapa-jumlah-stasiun-kereta-api-penumpang-di-jawa-dan-sumatera>

Widyaningtyas, T., Muladi, & Qonita, A. (2019). Use of ARIMA Method To Predict The Number of Train Passenger In Malang City. *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, 359–364. https://doi.org/10.1109/ICAIIT.2019.8834663

Wulandari, R. A., & Gernowo, R. (2019, Januari). Berkala Fisika. *METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DAN METODE ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) DALAM ANALISIS CURAH HUJAN*, *Vol. 22*, Hal. 41-48.

**PEMBAGIAN TUGAS**

| **No** | **Nama** | **Tugas** |
| --- | --- | --- |
| **1** | **Dimas Wahyu Saputro** | * **Mengerjakan pemrograman.** * **Mengerjakan laporan bagian Result and Discussion** * **Membuat PPT, bagian Methods dan Result.** * **Merapikan laporan.** |
| **2** | **Dhea Sukma Agustiana** | * **Mengerjakan laporan bagian Introduction, Methods.** * **Membuat PPT.** * **Merapikan laporan.** |
| **3** | **Fadia Dilla Sabine** | * **Mengerjakan laporan bagian Introduction.** * **Membuat PPT.** * **Merapikan laporan.** |
| **4** | **Lis Nurani** | * **Mengerjakan laporan bagian Methods.** * **Membuat PPT.** * **Merapikan laporan.** |
| **5** | **Rezki Pahala Manullang** | * **Mengerjakan laporan bagian Introduction, dan Referensi.** * **Membuat PPT.** * **Merapikan laporan.** |
| **6** | **Taj Shavira H** | * **Mengerjakan laporan bagian Introduction.** * **Membuat PPT.** * **Merapikan laporan.** |