

BAB IV

HASIL PENGUJIAN DAN ANALISIS

4.1 Hasil dan Analisis

Pada bagian hasil, peneliti menjabarkan hasil penelitian yang didapat dari pengumpulan data, perancangan dan implementasi rekomendasi *trend keyword SEO* menggunakan metode regresi linear dan *ARIMA*.

4.1.1 Hasil *Preprocessing* Data

1. Pengumpulan data

Pada tahap ini dilakukan pengumpulan data dan informasi sebagai dasar *training* dalam model untuk rekomendasi *trend keyword SEO*. Data atau informasi didapatkan dengan cara mengunduh data dari <https://trends.google.co.id/trends/>. Data yang telah diunduh akan memiliki format .CSV yang siap untuk melalui tahap *pre-processing* data.

Kemudian pada tahap ini data dilakukan identifikasi untuk melihat bagaimana struktur data berbentuk (kolom dan baris), jika struktur data tidak sesuai dengan bentuk dataset *time series* maka akan dilakukan dengan penyesuaian dan perbaikan. Adapun data yang digunakan untuk penelitian ini mencakup:

1. Data *search volume* untuk kata kunci “perumahan” selama kurun waktu 5 tahun kebelakang yaitu semenjak tanggal 16 september 2018 sampai dengan tanggal 3 september 2023
2. Data jumlah *search volume* per-periode mingguan,
Data yang telah diunduh memiliki informasi jumlah *search volume* (intensitas volume pencarian kata kunci) dalam kurun waktu satu minggu, beberapa bulan memiliki 5 minggu dikarenakan agregasi tanggal yang dilakukan oleh *google trend*.
3. Data *search volume* adalah data yang memberikan informasi berapa banyak pencarian kata kunci yang dilakukan dalam

kurun waktu tertentu. Pada penelitian ini data *search volume* adalah informasi dari berapa banyak pencarian kata kunci “perumahan” dalam kurun waktu 1 minggu.

4. Data ini hanya memberikan informasi *search volume* untuk kata kunci “perumahan” saja dan tidak memberikan informasi untuk kata kunci selain “perumahan”.

Berikut hasil pengumpulan data yang digunakan penelitian ini dapat dilihat pada Tabel 4.1.

Tabel 4. 1 Dataset yang di peroleh

Tahun	Bulan	M1	M2	M3	M4	M5
2018	ke-9	71	75	68	52	
	ke-10	72	74	72	70	
	ke-11	75	77	78	74	
	ke-12	72	70	83	78	81
2019	ke- 1	84	84	85	87	
	ke- 2	92	78	78	74	
	ke- 3	73	76	77	76	78
	ke- 4	73	71	71	75	
	ke- 5	71	71	63	64	
	ke- 6	67	82	85	100	88
	ke- 7	86	83	80	74	
	ke- 8	68	76	73	70	
	ke- 9	71	75	70	69	68
	ke- 10	67	70	70	65	
	ke- 11	71	70	71	69	
	ke- 12	61	68	71	70	82
2020	ke-1	88	75	79	75	
	ke-2	73	76	75	78	
	ke-3	75	67	61	56	58
	ke-4	59	60	61	58	
	ke-5	60	68	65	84	95
	ke-6	93	93	93	84	
	ke-7	84	75	73	65	
	ke-8	68	69	76	71	74
	ke-9	67	68	65	64	
	ke-10	68	67	71	66	
	ke-11	63	69	66	67	70
	ke-12	64	65	65	71	
2021	ke-1	72	65	76	75	71

Tahun	Bulan	M1	M2	M3	M4	M5
	ke-2	66	77	72	70	
	ke-3	72	75	69	62	
	ke-4	72	64	57	63	
	ke-5	58	63	76	81	72
	ke-6	77	66	68	61	
	ke-7	61	56	55	60	
	ke-8	59	68	58	58	59
	ke-9	59	61	63	60	
	ke-10	63	61	64	57	58
	ke-11	60	60	60	57	
	ke-12	62	62	65	58	
	2022	ke-1	63	62	57	63
ke-2		65	69	65	68	
ke-3		65	64	64	63	
ke-4		61	61	64	56	
ke-5		58	69	69	66	65
ke-6		62	64	63	64	
ke-7		61	60	63	60	60
ke-8		60	61	58	59	
ke-9		56	65	63	61	
ke-10		57	58	55	58	59
ke-11		59	60	53	55	
ke-12		55	53	57	62	
2023	ke-1	67	62	63	65	61
	ke-2	66	62	60	61	
	ke-3	65	57	54	55	
	ke-4	53	50	46	61	63
	ke-5	62	61	65	65	
	ke-6	62	65	60	64	
	ke-7	70	66	62	58	56
	ke-8	53	57	63	54	

Data diatas telah dilakukan perubahan dan penyesuaian agar lebih mudah melakukan identifikasi dan penyajian data. Data yang akan diproses kedalam model regresi linear dan *ARIMA* adalah data yang tidak diproses untuk tujuan idenfikasi dan penyajian data.

Dataset diatas terdiri dari data *time series* untuk *search volume* keyword *SEO* “peumahan” selama kurang lebih 5 tahun yang lalu. Pada dataset diatas sudah dilakukan perubahan yaitu memisahkan element

waktu pada tanggal dataset yang pada awalnya dataset hanya menampilkan dataset *search volume* setiap minggu nya tanpa memisahkan element mingguan dan tahunan. Dataset yang sudah dilakukan pemisahan element akan terlihat seperti diatass.

2. EDA (*Exploratory Data Analysis*)

a. *Loading data*

Dalam tahap ini, proses pengambilan data dilakukan melalui pengunduhan file berformat .CSV dari *Google Trends*. Untuk mengakses data, peneliti memanfaatkan alamat situs web resmi *Google Trends*, yaitu <https://trends.google.co.id/trends/>. Pengunduhan dilakukan dengan cermat, melibatkan proses penyortiran kata kunci berdasarkan potensi kontribusinya terhadap industri real estate, yang dalam penelitian ini adalah kata kunci “perumahan”. Selain itu, peneliti juga mempertimbangkan jangka waktu data yang diperlukan, menjadikan aspek ini sebagai faktor krusial dalam pemilihan data yang relevan. Dengan melakukan sorting ini, peneliti berupaya untuk memastikan bahwa data yang diperoleh tidak hanya relevan namun juga mencerminkan tren dan dinamika terkini dalam industri properti. Proses ini menciptakan dasar yang kokoh untuk analisis lanjutan terkait dengan perkembangan industri real estate.

Proses pertama dalam tahap ini adalah melakukan *loading dataset* yang telah diunduh, untuk memastikan dataset dapat diproses. Kemudian, dataset yang sudah *di-load* akan ditampilkan

5 data teratas atau biasa disebut dengan *head* dari dataset.

	Minggu	perumahan: (Indonesia)
0	2018-09-16	71
1	2018-09-23	75
2	2018-09-30	68
3	2018-10-07	72
4	2018-10-14	74

Gambar 4. 1 Hasil Load Data

Dataset diatas memberikan informasi yaitu dataset terdiri dari 2 kolom, minggu dan perumahan: (Indonesia). Kolom minggu memberikan informasi tentang tanggal dan interval waktu dari setiap baris, yang dalam penelitian ini interval setiap baris data terjadi setiap satu minggu. Sedangkan untuk kolom perumahan: (Indonesia) memberikan informasi berupa *search volume*. Dengan kedua informasi diatas dapat disimpulkan bahwa dataset ini memberikan informasi data *search volume* dengan interval waktu 1 minggu.

b. Informasi dasar data

Setelah melakukan loading data dan melihat 5 data teratas (*head*) didapati informasi bahwa data dataset terdiri dari 2 kolom yaitu kolom mingguan dan perumahan: (Indonesia) yang masing-masing memberikan informasi tanggal atau interval waktu dan *search volume*. Selanjutnya adalah melihat informasi dasar data, seperti ada berapa banyak data dan terdiri dari berapa banyak kolom dan baris dan format apa yang digunakan.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 2 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Minggu                                260 non-null    object
1   perumahan: (Indonesia)                260 non-null    int64
dtypes: int64(1), object(1)
memory usage: 4.2+ KB

```

Gambar 4. 2 Hasil informasi dasar data

Hasil info dasar data memberikan informasi data bahwa terdapat 2 kolom data dengan 260 baris, kemudian data ini juga memiliki format untuk setiap kolom nya dengan format kolom *object* untuk kolom mingguan, dan format *int64* untuk kolom perumahan: (indonesia). Hasil ini informasi dasar ini sudah cukup baik memberikan informasi dan dapat dilanjutkan ke tahap selanjutnya

c. Mencari nilai kosong (*null*)

Setelah mendapatkan informasi dasar mengenai dataset, selanjutnya akan dilakukan pencarian data yang kosong (*null*) atau data yang tidak memiliki nilai pada kolom atau baris nya, hal ini penting karena data yang memiliki *null* maka akan menyebabkan dataset tidak seimbang dan membuat model memiliki akurasi yang buruk dalam melakukan prediksi

```

df.isnull().sum()

Minggu                                0
perumahan: (Indonesia)                0
dtype: int64

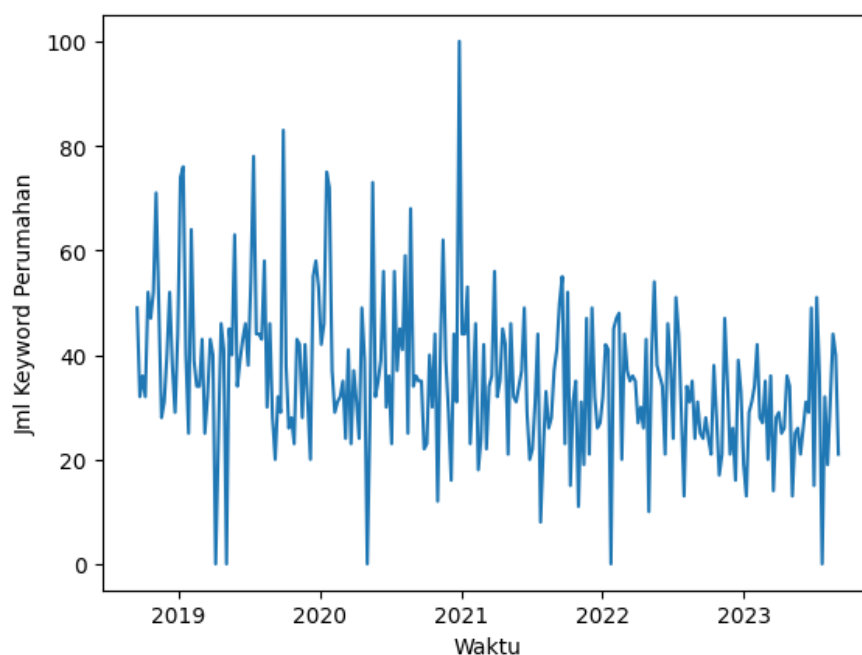
```

Gambar 4. 3 Hasil informasi data null

Pada informasi diatas didapatkan hasil bahwa tidak adanya *null* pada *dataset* hal ini memberikan hasil yang sangat baik dan dengan begitu dapat dilanjutkan ketahap selanjutnya.

d. Visualisasi data

Pada tahap ini visualisasi *dataset* dilakukan untuk dapat melihat karakteristik data, seperti apakah ada data yang nilainya terpaut sangat jauh dari rata-rata data yang ada, hal tersebut biasa disebut sebagai *outlier*. Selain itu visualisasi data juga memberikan informasi tentang tren dan arah data.



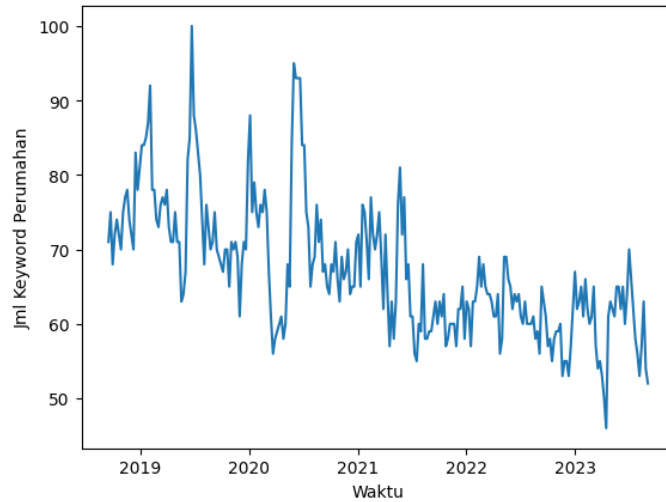
Gambar 4. 4 Hasil Visualisasi Data

Berdasarkan hasil visualisasi data diatas dapat disimpulkan bahwa ada data yang menjadi *outlier* hal ini tidaklah baik dan perlu dilakukan pembersihan, karena *outlier* dapat membuat data jadi tidak seimbang yang akan berdampak pada akurasi model dalam memprediksi. Hasil pengecekan dan analisa awal ini sudah cukup untuk mendapatkan informasi tentang *outlier* dan dapat dilanjutkan ketahap selanjutnya.

e. Interpolasi data

Pada tahap ini akan dilakukan interpolasi terhadap data yang memiliki *outlier* seperti yang telah ditunjukkan oleh tahap sebelumnya pada saat visualisasi data. Interpolasi akan bekerja

dengan memasukkan nilai rata-rata yang sesuai dengan dataset, sehingga membuat hasil interpolasi sangat baik.



Gambar 4. 5 Hasil visualisasi data setelah di interpolasi

Berdasarkan hasil interpolasi diatas, didapati bahwa grafik data sudah berubah dan menghasilkan visualiasi dan pergerakan tren yang lebih jelas. Data visual diatas juga tidak memperlihatkan adanya *outlier* seperti yang ditunjukkan pada tahap sebelumnya, hal ini menandakan bahwa interpolasi berhasil dan dapat dilanjutkan ketahap selanjutnya

f. Merubah format data pada kolom minggu

Pada tahap ini akan dilakukan perubahan format data pada kolom minggu. Seperti yang sudah diketahui dari proses pada tahap sebelumnya, bahwa info dasar pada dataset menunjukkan bahwa kolom minggu memiliki format *object* yang perlu dilakukan perubahan atau parse. Kolom perumahan (Indonesia) juga perlu dilakukan perunahan format menjadi *float64* yang pada dasarnya tidak perlu dilakukan, namun untuk mempermudah proses pemodelan maka perlu dilakukan penyesuaian.


```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 2 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Minggu                                260 non-null    datetime64[ns]
1   perumahan: (Indonesia)                260 non-null    float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 4.2 KB

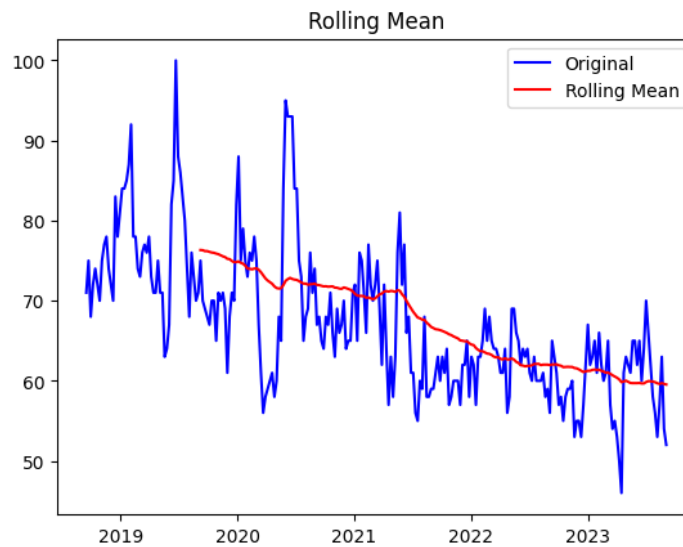
```

Gambar 4. 6 Hasil Merubah format *date*

Berdasarkan hasil perubahan format diatas, didapati informasi perubahan yang telah berjalan dengan baik. Hal ini dapat dibuktikan dengan melihat *dtype* yang sudah menunjukkan perubahan dari *object* menjadi *datetime64*, begitupun dengan perubahan untuk kolom perumahan: (Indonesia) yang telah berubah menjadi *float64*. Setelah melakukan perubahan format maka dapat dilanjutkan ke tahap selanjutnya.

g. Melihat arah tren dataset

Pada tahap ini, setelah melakukan berbagai macam proses *pre-processing dataset*, proses terakhir adalah melihat pergerakan tren dari *dataset* menggunakan *rolling-mean*. Proses melihat tren ini bekerja dengan cara membantu meratakan jika ada fluktuasi pada data sehingga memudahkan untuk melihat pergerakan data yang tersembunyi dalam pola grafik yang sudah diproses



Gambar 4. 7 Hasil visualisasi tren data

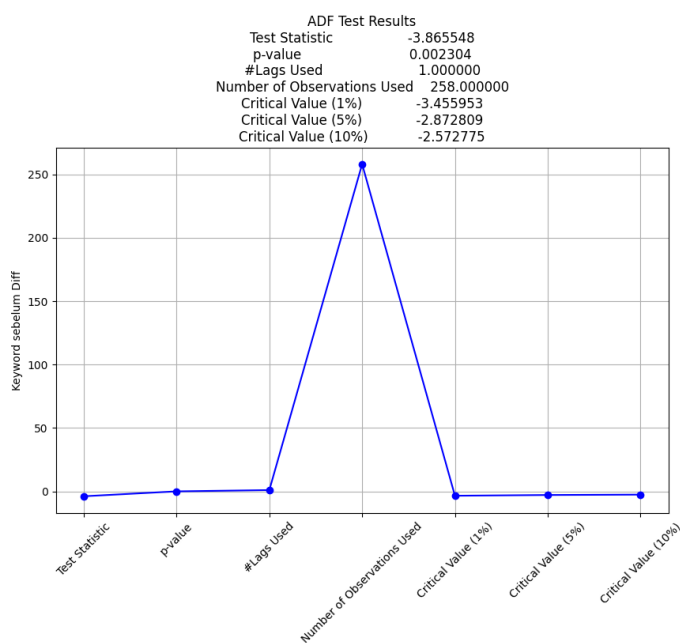
Berdasarkan hasil visualisasi tren data diatas menggunakan *rolling mean*, didapati hasil pergerakan data menurun. Hal ini memberikan informasi bahwa data *search volume* untuk *keyword* perumahan terus mengalami penurunan selama kurang lebih 5 tahun kebelakang. Selanjutnya, berdasarkan informasi yang sudah didapatkan ini, penelitian akan dilanjutkan dengan melakukan modeling dengan *dataset* yang telah melewati tahap *pre-processing*.

4.1.2 Tahap Perancangan dan Pengujian Model

Hasil dari tahap EDA selanjutnya dilakukan proses training menggunakan model Regresi Linear dan *ARIMA* agar dapat melakukan peramalan *time series* dengan baik. Namun sebelum itu akan dilakukan 2 test untuk membuat model *ARIMA* dapat bekerja dengan baik., Kemudian dilakukan pembagian dataset *train* dan *test* dibagi menjadi 80% *train* dan 20% *test*. Berikut adalah tahapan perancangan dan pengujian model

1. Analisa ADF (*augmented dickey fuller*)

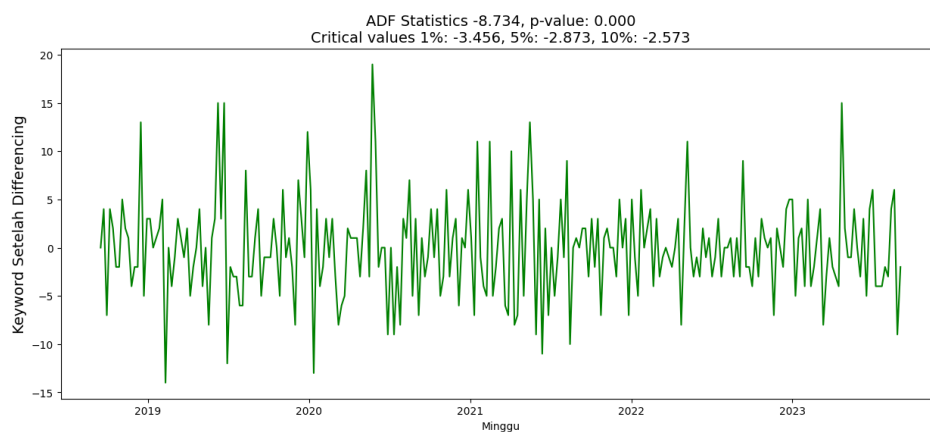
Pada proses ini akan dilakukan uji stationeritas dataset menggunakan ADF, karena model *ARIMA* dapat bekerja dengan baik jika dataset teruji stationer. Model *ARIMA* sangat bergantung pada dataset yang stationer karena asumsi dasar model ini yang mendasari bahwa dataset bersifat stasioner. Kestasioneran adalah kondisi di mana statistik deskriptif dari data, seperti rata-rata atau varians, tetap konstan sepanjang waktu. Pada proses ini pengujian dataset stationer menggunakan *automatic dickey fuller* (ADF) yang dapat dilihat pada Gambar 4.8



Gambar 4. 8 Hasil uji stationeritas menggunakan Automatic Dickey Fuller

Berdasarkan hasil yang dapat dilihat diatas, didapati bahwa hasil *p-value* berada pada angka 0.0023 hal ini sebenarnya sudah dapat dikatakan bahwa dataset stasioner karena dalam test ADF ini akan mengacu dengan uji hipotesa jika *p-value* bernilai lebih dari 0.005 maka dataset tidak stasioner namun, jika kurang dari 0.005 maka dataset dapat dikatakan stasioner. Hal ini menunjukkan bahwa sebenarnya dataset sudah stasioner namun, ada baiknya dilakukan *differencing* untuk mendapatkan hasil yang lebih stasioner

Pada proses ini dilakukan differencing untuk mendapatkan hasil pengujian dataset yang lebih baik dan sesuai dengan kebutuhan karena nilai p -value yang masih tinggi. Hasil pengujian dapat dilihat pada Gambar 4.9



Gambar 4. 9 Hasil uji stationeritas menggunakan *Automatic Dickey Fuller* dan *Differencing*

Berdasarkan hasil diatas dapat disimpulkan bahwa hasil test ADF pada data yang sudah dilakukan differencing memberikan p -value yang lebih kecil. Jadi, dapat disimpulkan hasil differencing dataset lebih baik dan akan digunakan sebagai dataset yang akan dilanjutkan pada proses berikut nya

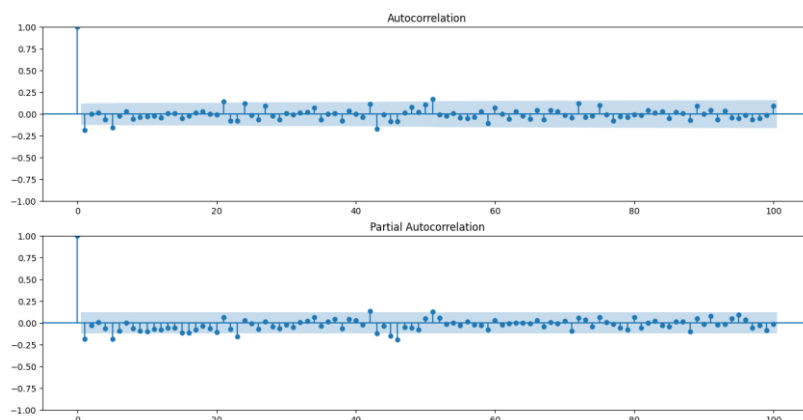
2. Analisa (ACF) *Autocorrelation Function* dan (PACF) *Partial Autocorrelation Function*

Pada proses ini akan dataset yang telah dilakukan uji ADF dan juga differencing akan dilakukan proses analisa dengan ACF dan PACF, proses ini berguna untuk mengidentifikasi keberadaan *serial correlation*, yaitu korelasi antara dataset *time series* dan *lags* (titik waktu pada timeseries) dan untuk mendapatkan nilai hyper parameter seperti *auto regression AR* dan *moving average MA* dengan menganalisa korelasi yang ada pada *lag* di dataset.

Pada analisa ACF, jika lag 1 bernilai tinggi maka itu berarti ada hubungan linear yang kuat antara nilai sekarang dengan nilai yang

ada di masa lalu. Analisa ACF juga dapat menunjukkan pergerakan yang berpola misalnya ada kenaikan setiap 12 *lag*, maka kemungkinan besar dataset yang dianalisa memiliki pola *seasonal*. Selain itu, jika ACF menunjukkan penurunan *lag* yang signifikan setelah beberapa *lag* maka kemungkinan sebagian besar informasi ada pada *lag* yang turun signifikan tersebut.

Pada analisa PACF, korelasi antara dataset dan nilai *lag* nya akan diukur setelah menghilangkan *linear dependence* pada dataset tersebut. PACF juga dapat membantu dalam mengidentifikasi urutan model AR. Model AR merepresentasikan nilai saat ini sebagai kombinasi linier dari nilai masa lalu pada *dataset*. PACF membantu mengidentifikasi jumlah nilai masa lalu pada *dataset* yang secara langsung mempengaruhi nilai saat ini, dan secara efektif menunjukkan korelasi yang jelas. Hasil Analisa ACF dan PACF dapat dilihat pada Gambar 4.10.



Gambar 4. 10 Hasil analisa ACF dan PACF

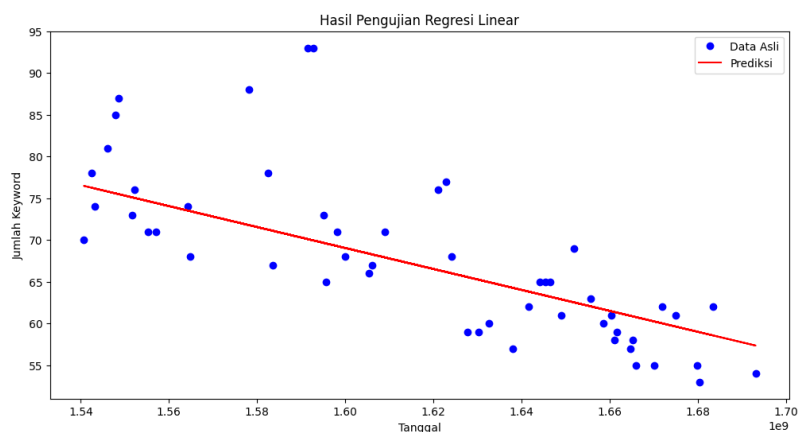
Berdasarkan hasil pengujian dan analisa diatas, didapati hasil yang cukup memuaskan dengan ACF berada pada kisaran 35, hasil ACF ini akan mewakili komponen MA yang berarti nilai untuk hyper parameter q pada model *ARIMA*. Selanjutnya hasil PACF menunjukkan nilai pada kisaran 15 yang mewakili komponen AR yang berarti nilai untuk hyper parameter p . Dengan ini seluruh

komponen *ARIMA* telah diketahui dan akan dilanjutkan ke tahap pemodelan regresi linear dan *ARIMA*.

3. Analisa model regresi linear

Pada tahap pengujian ini dilakukan *fitting* model regresi linear dengan dataset yang telah diproses. Tahap ini dilakukan untuk memastikan apakah model dapat berjalan dengan baik dan dapat digunakan untuk peramalan *trend keyword SEO*.

Pengujian yang dilakukan dengan cara melatih model data *train* dan mengujinya dengan data *test* yang didapat dari dataset, kemudian apakah model mampu berjalan dengan baik sehingga *output* yang dihasilkan sesuai dengan kebutuhan peramalan *trend keyword SEO*. Berikut hasil dari pengujian model regresi linear yang telah dilakukan.



Gambar 4. 11 Hasil pengujian regresi linear

Pada pengujian model regresi linear ini, dataset tanggal yang sudah diubah formatnya menjadi *datetime* pada *pre-processing* diubah kembali menjadi *unix timestamp* agar dataset dapat diolah oleh model regresi linear.

Hasil pengujian memberikan nilai RMSE sebesar 6.8 dan nilai MAPE sebesar 6.80%. Proses pengujian model regresi ini menggunakan library dari *numpy* dan model regresi linear menggunakan library dari *sklearn.linear_model* serta pemisahan data *train* dan *test* menggunakan library *sklearn.model_selection*.

Dalam tahapan pertama mengimplemntasikan rumus regresi linear sederhana menghitung manual kita dapat diketahui bahwa nilai:

$$n = 60$$

$$\Sigma x = 260$$

$$\Sigma y = 17492$$

$$\Sigma xy = 76684$$

$$\Sigma x^2 = 1140$$

$$(\Sigma x)^2 = 67600$$

Sehingga jika nilai tersebut diimplementasikan kedalam rumus mencari variabel a atau konstanta sebagai berikut:

$$a = \frac{(\Sigma y) (\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n \Sigma x^2 - (\Sigma x)^2}$$

$$a = \frac{(17492)(1140) - (260)(76684)}{60(1140) - 67600}$$

$$a = \frac{19940880 - 19937840}{68400 - 67600}$$

$$a = \frac{3040}{800}$$

$$a = 3,8$$

Sedangkan untuk mengimplementasikan variabel b atau koefisien kedalam rumus adalah sebagai berikut:

$$b = \frac{n \Sigma xy - (\Sigma x) (\Sigma y)}{n \Sigma x^2 - (\Sigma x)^2}$$

$$b = \frac{60 (76684) - 260(17492)}{60(1140) - 67600}$$

$$b = \frac{4601040 - 4547920}{68400 - 67600}$$

$$b = \frac{53120}{800}$$

$$b = 66,4$$

Nilai konstanta dan koefisien setelah didapatkan, untuk mencari nilai y (*search volume*) dapat diimplementasikan kedalam rumus dengan nilai $x = 5$ sebagai contohnya adalah sebagai berikut:

$$\hat{y} = a + b + e$$

$$\hat{y} = 3,8 + (66,4)5 + e$$

$$\hat{y} = 335,8$$

Dari persamaan diatas untuk menghitung y dengan $x = 5$ diperoleh hasil 335,8

Tabel prediksi masa depan berdasarkan perhitungan regresi linear tersebut dapat di lihat pada Tabel berikut:

Tabel 4. 2 Tabel Predeksi Regresi Linear

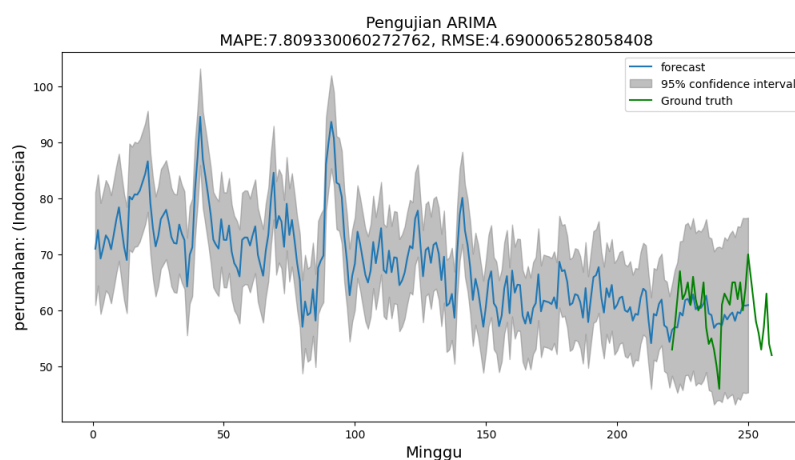
Tahun	Total Minggu (x)	Total Keyword(y)	x^2	xy
2018	17	1242	289	21114
2019	52	3891	2704	202332
2020	52	3700	2704	192400
2021	52	3364	2704	174928
2022	52	3181	2704	165412
2023	35	2114	1225	73990
Jumlah	260	17492	12330	830176
Rata-rata	43,33333333	2915,333333		

Berdasarkan dari pengujian diatas didapati hasil prediksi menggunakan regresi linear baik secara manual maupun menggunakan model regresi linear dari sklearn. Hasil analisa model regresi linear menunjukkan penurunan tren untuk kata kunci perumahan, serta hasil prediksi yang akan dibandingkan dengan model *ARIMA* pada tahap selanjutnya.

4. Analisa model *ARIMA*

Pada tahap pengujian ini dilakukan *fitting* model *ARIMA* seperti pengujian yang dilakukan dengan regresi linear, namun pada tahap

pengujian ini melibatkan order atau *hyper parameter* dari model *ARIMA*. Berdasarkan hasil yang didapatkan dari analisa ACF dan PACF, *order hyper parameter* yang akan di gunakan untuk membangun model *ARIMA* ini adalah 15, 1, dan 35 dengan masing-masing mewakili p , d , dan q . Setelah melakukan beberapa pengujian berulang didapatkan hasil pengujian seperti yang dapat dilihat pada Gambar 4.12



Gambar 4. 12 Hasil pengujian *ARIMA*

Pada pengujian model *ARIMA* yang pertama kali ini menggunakan dataset yang belum dilakukan differencing kemudian menggunakan dataset yang sudah dilakukan differencing, hasilnya pada hyper parameter *ARIMA* yaitu d akan diberikan nilai 1. Setelah melakukan uji coba model *ARIMA* tersebut didapati hasil akurasi dari pengujian ini adalah 4.69 untuk RMSE dan 7.8% untuk MAPE. Selain itu penelitian ini juga melakukan uji coba berulang laki dengan berbagai kombinasi hyper parameter untuk *ARIMA* yang menghasilkan akurasi lebih buruk dibandingkan menggunakan *hyper parameter* dengan kombinasi 15, 1, dan 35. Jadi dapat disimpulkan bahwa *order* paling optimal untuk model *ARIMA* ini adalah dengan nilai 15, 1, dan 35.

5. Uji Akurasi dan Hasil Prediksi

Pada tahap ini hasil dari prediksi *trend keyword SEO* menggunakan metode regresi linear dan *ARIMA* akan dibandingkan untuk melihat dan memahami lebih baik perbedaan akurasi yang ada dengan data aktual. Hasil prediksi dapat dilihat pada Tabel 4.4.

Tabel 4. 3 Hasil prediksi *search volume* regresi linear dan *ARIMA*

Minggu	ARIMA	Regresi Linear	Data Aktual
238	57.541549	59.633494	55
239	57.673088	66.008517	53
240	57.354014	74.432654	50
241	59.226232	62.062074	46
242	58.377601	69.271921	61
243	59.138895	60.847784	63
244	59.624986	76.178196	62
245	58.137899	68.361204	61
246	59.696104	69.651387	65
247	59.474187	69.954959	65
248	60.919386	65.553158	62
249	60.838552	75.039799	65
250	60.943964	64.946013	60
251	60.670764	75.571051	64
252	58.703009	58.950455	70
253	58.589857	66.387982	66
254	57.861900	71.776395	62
255	58.594624	69.575494	58
256	59.987400	64.262975	56
257	58.847882	70.106746	53
258	59.981429	61.379036	57

Berdasarkan Tabel diatas didapati hasil bahwa model *ARIMA* lebih baik karena mendekati nilai kebenaran (data aktual). Sedangkan untuk model regresi linear sedikit lebih jauh dari data aktual yang ada. Dengan melihat hasil perbandingan ini maka dapat disimpulkan bahwa model *ARIMA* jauh lebih baik daripada model regresi linear jika mengacu pada hasil prediksi dengan membandingkan data aktual.

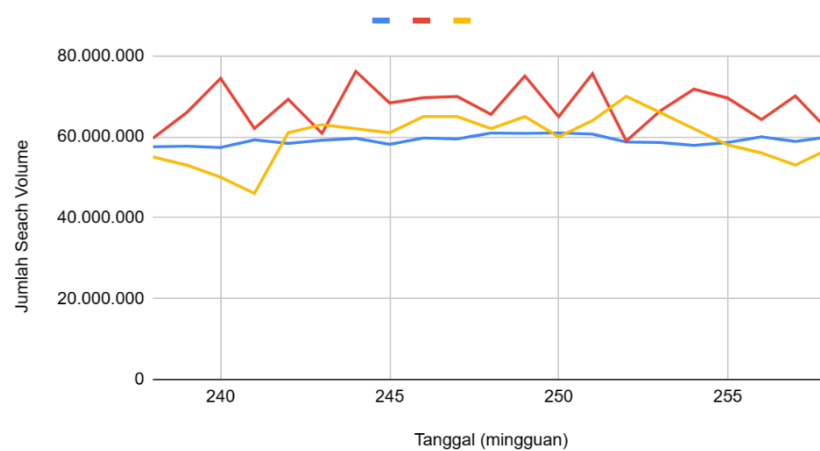
Dari hasil analisis diatas dapat dilihat untuk hasil perbandingan RMSE dan MAPE antara model Regresi Linear dengan *ARIMA* dapat dilihat pada Tabel 4.3.

Tabel 4. 4 Tabel Hasil Perbandingan RMSE dan MAPE

Model	RMSE	MAPE
Regresi Linear	68.04	6.80%
<i>ARIMA</i>	4.69	7.80%

Dari Tabel 4.3 dapat disimpulkan bahwa hasil RMSE dari model Regresi Linear di peroleh 68,048% dan MAPE sebesar 6,80%, sedangkan untuk model *ARIMA* memperoleh hasil RMSE sebesar 4,690% lebih kecil dibandingkan dengan hasil RMSE dari Regresi Linear. Untuk hasil MAPE dari model *ARIMA* sebesar 7,80% lebih besar dibandingkan hasil MAPE model Regresi Linear.

Perbandingan Model Dengan Data Aktual



Gambar 4. 13 Hasil perbandingan model

Dari Gambar 4. 13 dapat dilihat bahwa ada 3 garis yang masing-masing mewakili hasil prediksi model dan data aktual dengan keterangan plot garis berwarna biru adalah plot untuk hasil prediksi model *ARIMA*, plot garis berwarna merah adalah plot hasil prediksi regresi linear, dan plot garis berwarna kuning adalah data aktual dari hasil prediksi. Berdasarkan hasil perbandingan diatas kita mendapati hasil bahwa metode *ARIMA* memberikan hasil yang lebih mendekati nilai data aktual sehingga dapat dikatakan lebih baik dari model regresi linear. Hasil prediksi ini juga menunjukkan bahwa kedua model ini menunjukkan hasil penurunan tren untuk *keyword* perumahan.