**LAPORAN UTS TAKE HOME**

**“ DATA PREDIKSI KEBUTUHAN SEKOLAH DI IKN”**

****

**Disusun oleh:**

**Nama : Nahdliyah Zahrah**

**Nim : 22031554024**

**Kelas : Penambangan Data 2022B**

**PRODI SAINS DATA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SURABAYA**

**2024**

1. **Algoritma**

Algoritma yang digunakan dalam analisis kebutuhan sekolah dari tahun ketahun diwilayah IKN adalah algoritma ARIMA. ARIMA terdiri dari tiga komponen yaitu, Autoregressive (AR), Integrated (I), Moving Average (MA). Autoregressive merupakan komponen yang mengevaluasi hubungan antara nilai-nilai sebelumnya dalam deret waktu dan nilai saat ini. Derajat autoregresi juga menjadi penentu atau tolak ukur prediksi, autoregressive biasanya disebut p.

Integrated (I) mencerminkan terhadap proses diferensiasi pada deret waktu, yang bertujuan untuk menjadikan deret waktu atau data tersebut berupa stasioner. Sehingga memudahkan dalam pemodelan ARIMA, Integreted biasanya disebut q. Selanjutnya Moving Average (MA) merupakan komponen yang mengevaluasi eror atau kesalahan dalam forecast sebelumnya dalam memprediksi nilai saat ini, dan moving average biasanya disebut q. Penyebutan disetiap komponen ARIMA (p,d,q) merupakan penyesuaian parameter yang optimal untuk pemodelan.

Langkah dalam peneraopan ARIMA:

1. Analisis Data: Analisis awal terhadap data deret waktu untuk memahami tren, musiman, dan pola lainnya. Biasanya juga dilakukan dengan visualisasi data dalam plot histogram. Dalam tahap ini juga kita menentukan kolom mana yang akan digunakan sebagai tokoh utama dalam peramalan.
2. Stasionerisasi: Memastikan data deret waktu stasioner dengan menerapkan diferensiasi jika diperlukan. Stasioneritas dapat diperiksa dengan melakukan tes Augmented Dickey-Fuller (ADF):

> p-value > 0,05: Gagal menolak hipotesis nol (H0), data mempunyai akar unit dan tidak stasioner.

> p-value <= 0,05 : Menolak hipotesis nol (H0), data tidak mempunyai unit root dan stasioner.

1. Identifikasi Model: Menentukan parameter p, d, q untuk model ARIMA yang optimal berdasarkan analisis ACF (Autocorrelation Function) dan PACF (Partial Autocorrelation Function) dari data deret waktu. Dengan ketentuan ACF dan PACF mengasumsikan stasioneritas deret waktu yang mendasarinya. auto correlation: memperlihatkan hubungan antar data. ACF dan PACF digunakan untuk menentukan orde AR dan MA. orde/ordo: hari keberapa yang signifikan, hari keberapa yang berpengaruh dengan hari ini dengan melihat langsung dari plotnya.

* ACF : MA yang dilihat error yang signifikan (leg keberapa) tau error dilihat dari AR nya.
* PACF : AR hari yang signifikan

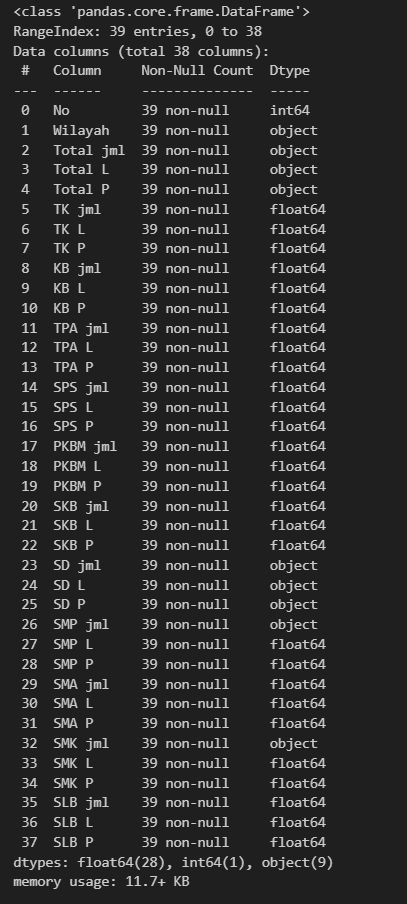
1. Estimasi Model: Mengestimasi parameter model ARIMA berdasarkan data deret waktu yang ada. Dengan order (p,d,q). (Hasil summary ARIMA akan menunjukan residu normal dengan melihat prob JB dan Ljung-Box nya harus lebih dari 0,05).
2. Peramalan: Menggunakan model ARIMA untuk membuat peramalan pada data masa depan. Semakin kecil hasil RMSE hasilnya makin bagus dan forecast dikatakan berhasil ketika hasil alur forecast sama dengan data asli, ikut naik atau ikut turun (line merah: forecast, line biru: data asli)
3. **Hasil langkah- langkah**

* **Import data**

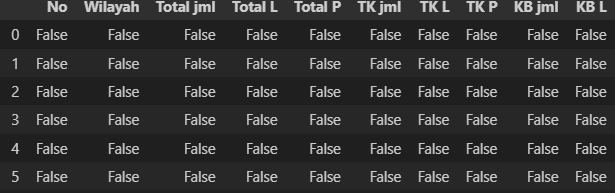
****

* **Memahami data**

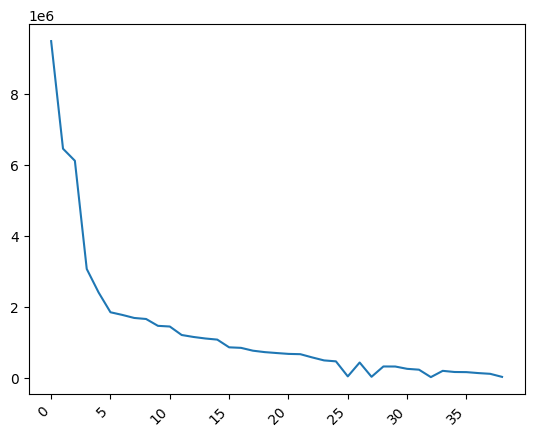
Dari data info didapat type data dan data yang berupa objek akan diubah ke float, kemudian kolom yang digunakan adalah kolom Total jumlah semua.

****

* **Missing Value**

****

* **Plot data**

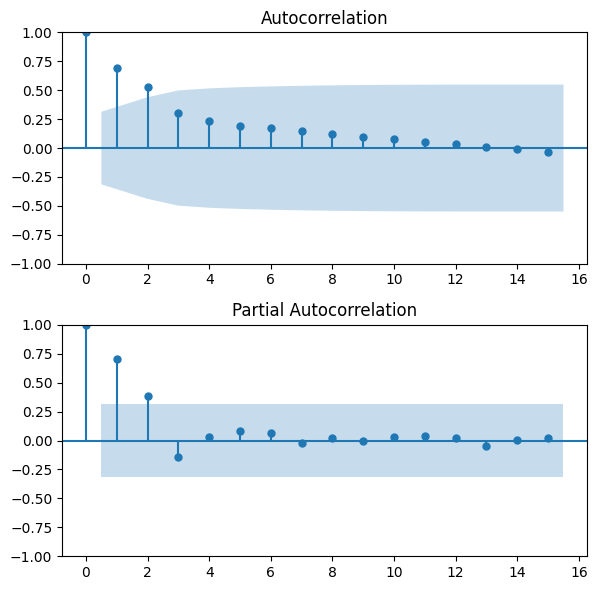
****

* **Stationarity**

Dalam stationarity didapat nilai pvalue= 2.138885340893453e-11 yang berarti menolak H0 / data sudah berupa stationer dan tidak dilakukan differencing lagi.

* **ACF dan PACF**

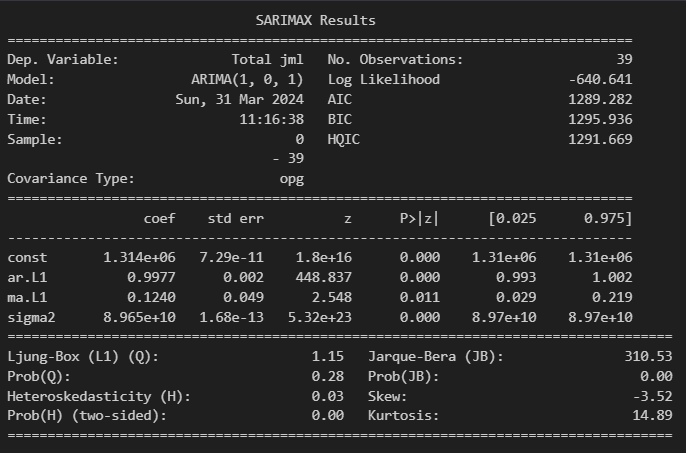
ACF dan PACF dilakukan dengan lag 15 setengah dari data terhadap data series (data asli). Lalu diambil lag yang paling tinggi selain 0 yaitu 1.

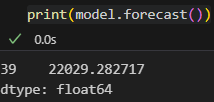
****

* **Model ARIMA**

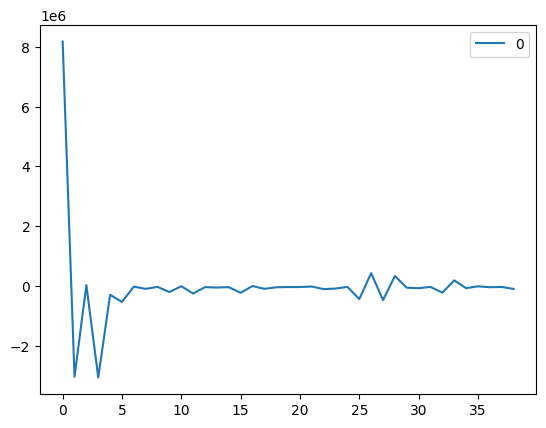
Pemodelan ARIMA (p,d,q) dilakukan dengan (0,0,0)

Hasil summary ARIMA didapat bahwa nilai L\_-jung Box dan JB lebih besar dari 0,05 yang menunjukkan residu normal.

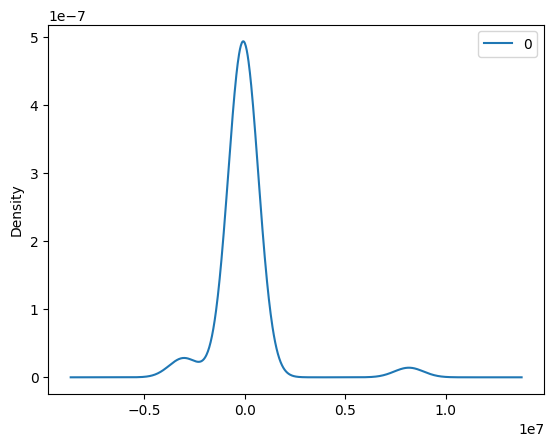


****

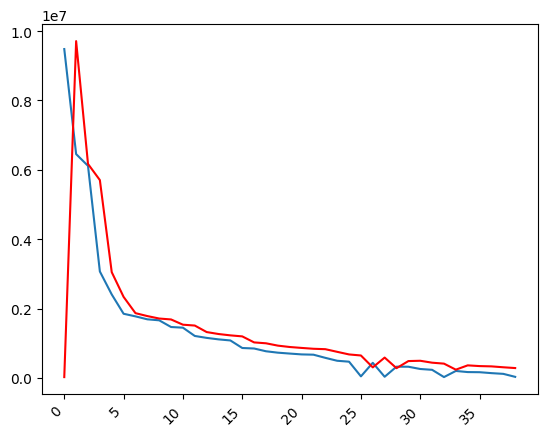
* **Residu plot**

****

* **Density Residu**

****

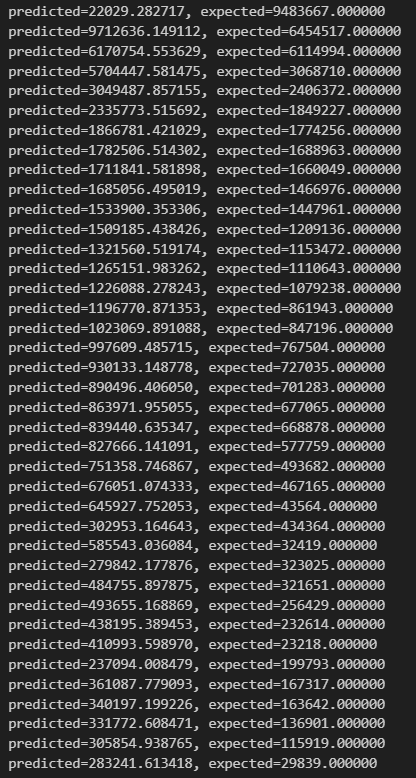
* **Forecast**

****

Pada hasil terlihat line forecasting searah dengan line data asli (forecasting berhasil**).**

1. **Kesimpulan**

Dengan menggunakan model ARIMA dan data yang diberikan, telah dilakukan proses prediksi jumlah peserta didik di Institut Kelembagaan Pendidikan (IKN) untuk beberapa tahun ke depan. Yang mana memang terdapat beberapa prediksi yang memiliki gap dengan ekspektasinya namun lebih banyak yang nilai prediksi dan ekspektasinya hampir sama. Berikut data hasil prediksi:



Nilai RMSE *(Root Mean Squared Error*) didapat sebesar 1676829.236. Nilai RMSE rendah menunjukkan bahwa model prediksi yang digunakan bekerja baik.