

Pemodelan dan Peramalan Data Saham dengan Analisis *Time Series* menggunakan Python

Bayu Dwi Prasetya^a, Fajar Sodik Pamungkas^a, Iqbal Kharisudin^{a*}

Universitas Negeri Semarang

Email : iqbalkharisudin@mail.unnes.ac.id

Abstrak

Dalam makalah ini dilakukan pemodelan time series menggunakan software Python. Data yang digunakan adalah data saham PT Bank Negara Indonesia. Secara umum diketahui bahwa pergerakan saham di Indonesia sangat fluktuatif oleh karena itu perlu dilakukan prediksi untuk mengetahui kondisi dan mempersiapkan strategi untuk menghadapi penurunan atau pelonjakan harga saham. Dalam makalah ini diselidiki pula factor seasonal atau musiman, sehingga pada identifikasi model digunakan analisis time series Seasonal Autoregressive Integrated Moving Average (SARIMA) berbantuan software Jupyter Notebook berbahasa Python. Dalam analisis ini Python membantu untuk melakukan perhitungan dan visualisasi data agar lebih mudah dan efisien. Dalam menganalisis data saham dengan metode SARIMA menggunakan Python, langkah pertama yang dilakukan adalah mempersiapkan modul yang akan digunakan untuk analisis SARIMA, kemudian dilakukan preprocessing data dan indexing. Langkah selanjutnya adalah visualisasi data untuk mengetahui pola data secara umum. Mengikuti langkah metode Box Jenkins, dilakukan identifikasi dan diagnosa model. Pada studi kasus ini didapat model SARIMA (1,1,1) (0,1,1,12). Berdasarkan studi kasus dari model tersebut, dilakukan fitting model dan didapat Mean Square Error (MSE) sebesar 460418,82. Langkah terakhir adalah melakukan peramalan untuk beberapa periode mendatang.

Kata kunci:

Metode Box Jenkins, pemodelan data saham, *time series*, Python

© 2020 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

1.1. Latar Belakang

Saham adalah salah satu instrument keuangan jangka panjang yang diperjual belikan di pasar modal dan merupakan salah satu alternatif investasi yang populer bagi investor di Indonesia. Saham merupakan surat berharga yang dikeluarkan oleh perusahaan dan sebagai tanda kepemilikan suatu perusahaan. Salah satu saham yang terdapat di pasar modal Indonesia adalah saham PT. Bank Negara Indonesia (Persero) Tbk. PT ini telah tercatat di dalam pasar modal sejak 25 November 1996 dengan bidang usaha Perbankan dan berada pada sektor finance dan subsektor Bank ("PT Bursa Efek Indonesia," n.d.). PT BNI mempunyai 10 anak perusahaan jenis jasa keuangan, sekuritas dan asuransi dengan laba yang mencapai Rp11,4 Triliun pada tahun 2016.

Posisi penawaran dan permintaan atas saham di pasar modal Indonesia membuat saham memiliki harga jual. Semakin tinggi permintaan maka harga saham akan meningkat dan semakin tinggi penawaran maka harga saham akan menurun. Pergerakan harga saham di Indonesia bersifat fluktuatif, oleh karena itu perlu dilakukan pemodelan dan prediksi untuk mengetahui kondisi dan mempersiapkan strategi untuk menghadapi penurunan atau pelonjakan harga saham. Setiap detik, pergerakan saham di Indonesia tercatat di bursa efek Indonesia. Salah satu analisis yang dapat digunakan untuk membuat prediksi dan pemodelan adalah analisis Time Series dengan memanfaatkan data pergerakan saham yang tercatat di bursa efek Indonesia.

Data yang digunakan adalah data pergerakan saham PT BNI dari Desember 2013 sampai Oktober 2019 dengan total data sebanyak 3947 data. Untuk menganalisis data sebanyak ini menggunakan bantuan program Jupyter Notebook dengan bahasa pemrograman python. Python telah banyak digunakan oleh

To cite this article:

Prasetya, B. D., Pamungkas, F. S., & Kharisudin, I. (2019,20). Pemodelan dan Peramalan Data Saham dengan Analisis *Time Series* menggunakan Python. *PRISMA, Prosiding Seminar Nasional Matematika 3*, 714-718

para Data Scientist untuk mengolah data. Python juga bahasa pemrograman yang populer dan relatif mudah penggunaannya. Dalam proses analisa menggunakan konsep Machine Learning dimana mesin mempelajari data historik dari pergerakan saham sebelumnya untuk memprediksi harga saham dimasa depan.

1.2. Rumusan Masalah

Pergerakan saham di Indonesia bersifat fluktuatif, oleh karena itu perlu dilakukan pemodelan dan prediksi untuk mengetahui kondisi dan mempersiapkan strategi untuk menghadapi penurunan atau pelonjakan harga saham. Untuk melakukan pemodelan dan prediksi perlu dilakukan dengan analisis yang tepat untuk memperkecil nilai *error* pada hasil prediksi.

1.3. Tujuan

Melakukan pemodelan dan prediksi harga saham PT Bank Negara Indonesia (Persero) Tbk 3 tahun mendatang

2. Metode (untuk artikel hasil kajian, bagian ini tidak ada)

2.1. Analisis Time Series

Analisis Time Series adalah analisis yang dilakukan berdasarkan data atau pengamatan yang berorientasi pada waktu atau kronologis pada variable yang diamati (Montgomery, Jennings, & Kulahci, 2016). Analisis ini sangat berguna pada data yang perubahannya dipengaruhi oleh waktu atau pengamatan sebelumnya. Dalam pengembangannya Analisis *Time Series* banyak digunakan di beberapa bidang seperti bidang ekonomi, bidang keuangan, bidang transportasi dan masih banyak lagi. Untuk membuat model yang cocok untuk meramalkan data terdapat beberapa tahapan proses dalam analisis *Time Series* yaitu: Stasioneritas data, *Parameter Estimation*, *Model Specification*, *Model Checking*, *Unit Root Test*, dan *Forecasting*.

2.1.1. Stasioneritas Data

Data stasioner harus memenuhi 2 kondisi. Kondisi pertama berarti bahwa semua anggota proses stokastik memiliki rata-rata konstan yang sama. Oleh karena itu, deret waktu yang dihasilkan oleh proses stokastik stasioner harus berfluktuasi di sekitar rata-rata konstan dan tidak memiliki tren. Kondisi kedua memastikan bahwa semua anggota proses stokastik memiliki varian yang sama. (Lutkepohl & Kratzig, 2004).

2.1.2. Parameter Estimation

Untuk data yang berdistribusi normal, metode estimasi parameter yang populer yaitu menggunakan metode *MLE (Maximum Likelihood Estimator)*. Dalam statistik, estimasi kemungkinan maksimum (MLE) adalah metode estimasi parameter distribusi probabilitas dengan memaksimalkan fungsi probabilitasnya, sehingga di bawah model statistik yang diasumsikan, data yang diamati adalah data yang paling memungkinkan. Titik dalam ruang parameter yang memaksimalkan fungsi kemungkinan disebut *Maksimum likelihood estimate* Rossi, Richard J. (2018).

2.1.3. Model Specification

Menentukan jenis-jenis model memerlukan penentuan dan kemungkinan syarat dan asumsi distribusi. Ini dapat dilakukan dengan mencocokkan suatu model, yang mencakup semua ketentuan yang mungkin menarik, dan kemudian melakukan pengujian untuk kecocokan model dan pengurangan model dengan cara yang biasa. Pendekatan ini dibatasi oleh fakta, bahwa parameter dalam model ARMA yang terlalu ditentukan mungkin tidak unik. Oleh karena itu, estimator tidak memiliki sifat asimptotik seperti biasa. Oleh karena itu, prosedur pemilihan model sering diterapkan untuk menentukan *order*.

2.1.4. Model Checking

Setelah model ditentukan, serangkaian alat diagnostik tersedia untuk memeriksa keakuratannya. Banyak dari alat diagnostik didasarkan pada residual model. Akhirnya, kekokohan dan stabilitas model dapat diperiksa dengan memperkirakannya secara rekursif untuk sub-sampel yang berbeda.

2.1.5. Unit Root Test

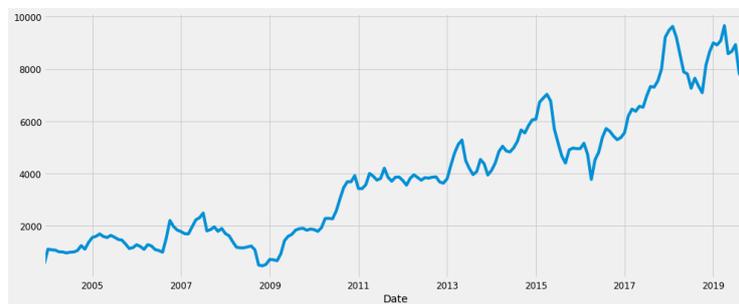
Karena urutan integrasi deret waktu sangat penting untuk analisis, beberapa uji statistik telah dikembangkan untuk menyelidikinya. Serangkaian tes memeriksa hipotesis nol bahwa ada unit root terhadap alternatif stasioneritas DGP yang mungkin memiliki istilah rata-rata nol, tren linear adeterministik, dan mungkin variabel dummy musiman. Salah satu terst yang dapat digunakan adalah *Augmented Dicky Fuller Test*.

2.1.6. Forecasting

Jika model yang cocok untuk Data dari seri waktu tertentu telah ditemukan, dapat digunakan untuk memperkirakan pengembangan variabel di masa depan yang sedang dipertimbangkan.

3. Hasil dan Pembahasan

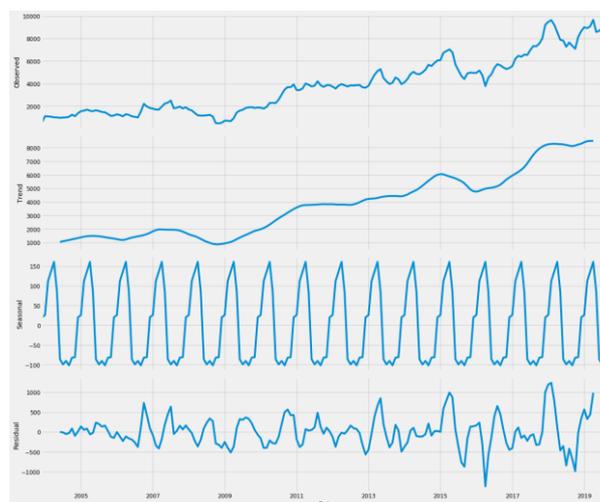
3.1. Visualisasi Data



Gambar 1. Visualisasi data harga saham PT BNI tahun 2015-2018

Beberapa pola yang dapat dibedakan muncul ketika data divisualisasikan. Rangkaian waktu memiliki pola musiman. Selalu ada tren naik dalam satu tahun saja dengan beberapa bulan rendah di pertengahan tahun. Data juga dapat divisualisasikan menggunakan metode yang disebut dekomposisi deret waktu yang memungkinkan untuk menguraikan deret waktu menjadi tiga komponen berbeda: tren, musiman, dan *noise*.

3.2. Dekomposisi



Gambar 2. Hasil dekomposisi data

Plot di atas jelas menunjukkan bahwa pergerakan saham tidak stabil, bersamaan dengan musim yang jelas dan memiliki tren naik. Karena memiliki musiman maka metode peramalan yang dapat digunakan adalah metode peramalan SARIMA.

3.3. Identifikasi model terbaik

Identifikasi model ARIMA menggunakan *Machine Learning*. Dengan menggunakan Iterasi dari SARIMA ((0,0,0),(0,0,0,12)) sampai SARIMA ((2,2,2),(2,2,2,12)). Dan untuk model terbaik dipilih model dengan *MSE (Mean Square Error)* terendah dan nilai *AIC* terbaik.

```

p = d = q = range(0, 2)
pdq = list(itertools.product(p, d, q))
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

params = []
params_s = []
aics = []

for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            model = sm.tsa.statespace.SARIMAX(array,
                                              order = param,
                                              seasonal_order = param_seasonal,
                                              enforce_stationarity=False,
                                              enforce_invertibility=False)

            results = model.fit()

            params.append(param)
            params_s.append(param_seasonal)
            aics.append(results.aic)

            print('ARIMA({})x{}12 - AIC:{}'.format(param, param_seasonal, round(results.aic, 2)))
        except:
            continue

print('Best Parameter :')
min_ind = aics.index(min(aics))
bestparam = (params[min_ind], params_s[min_ind])
print('ARIMA', bestparam, '- AIC:', min(aics))
    
```

Gambar 3. Kode python untuk mencari model terbaik

Dari iterasi tersebut didapatkan model terbaik adalah SARIMA ((1,1,1),(0,1,1,12)).

3.4. Fitting & Diagnostic Model

	coef	std err	z	P> z	[0.025	0.975]

ar.L1	0.9039	0.024	38.163	0.000	0.857	0.950
ma.L1	-1.0000	0.155	-6.457	0.000	-1.304	-0.696
ma.S.L12	-1.1640	0.090	-12.945	0.000	-1.340	-0.988
sigma2	1.339e+05	1.16e-06	1.15e+11	0.000	1.34e+05	1.34e+05

Ljung-Box (Q):			51.22	Jarque-Bera (JB):		45.82
Prob(Q):			0.11	Prob(JB):		0.00
Heteroskedasticity (H):			6.04	Skew:		-0.00
Prob(H) (two-sided):			0.00	Kurtosis:		5.59

Gambar 4. Hasil diagnosa model

Pada semua parameter terlihat menunjukkan nilai *P-Value* < 0,05 jadi dapat disimpulkan bahwa semua parameter model signifikan terhadap data. Uji Ljung-Box bertujuan apakah data bersifat independent. Dan dari uji Ljung-Box didapatkan *P-Value* sebesar 0,11 yang artinya bahwa data bersifat independent.

3.5. Validasi prediksi



Gambar 5. Hasil perbandingan nilai prediksi dengan nilai asli

Validasi prediksi diperlukan untuk mengetahui tingkat akurasi model dalam memprediksi data yang telah ada. Dalam proses ini didapatkan *RMSE (Root Mean Square Error)* sebesar 666,1.

3.6. Peramalan

Dari model SARIMA((1,1,1),(0,1,1,12)) didapatkan hasil peramalan harga saham PT BNI sebagai berikut



Gambar 6. Grafik hasil peramalan

2019-11-01	7162.736875	2021-05-01	8715.888464
2019-12-01	7610.625636	2021-06-01	8547.318668
2020-01-01	7659.487425	2021-07-01	8651.454203
2020-02-01	7938.899857	2021-08-01	8619.559894
2020-03-01	8077.714872	2021-09-01	8480.014059
2020-04-01	7965.918631	2021-10-01	8562.344458
2020-05-01	7954.753552	2021-11-01	8739.568249
2020-06-01	7764.496285	2021-12-01	9135.966023
2020-07-01	7923.853906	2022-01-01	9152.674162
2020-08-01	7941.636414	2022-02-01	9372.385243
2020-09-01	7826.795493	2022-03-01	9495.899228
2020-10-01	7943.207197	2022-04-01	9365.970218
2020-11-01	8132.670915	2022-05-01	9279.140977
2020-12-01	8538.317405	2022-06-01	9103.561667
2021-01-01	8564.041206	2022-07-01	9204.492917
2021-02-01	8790.579690	2022-08-01	9169.680468
2021-03-01	8921.981210	2022-09-01	9026.637429
2021-04-01	8799.011236	2022-10-01	9106.290871

Gambar 7. Hasil peramalan

4. Simpulan

Pada hasil peramalan didapatkan kesimpulan bahwa harga saham PT BNI dalam 3 tahun kedepan memiliki tren naik. Dalam peramalan ini terdapat kemungkinan bahwa harga saham tidak hanya dipengaruhi oleh waktu tetapi juga dapat dipengaruhi oleh faktor lainnya. Diharapkan di penelitian selanjutnya dapat menggunakan analisis Time Series Multivariat.

Daftar Pustaka (Style PRISMA-BagianNoNumber)

Lutkepohl, H., & Kratzig, M. (2004). *Applied Times Series*. Retrieved from file:///Files/1C/1C868265-79C2-4BD4-B423-9CCAAB1E7E69.pdf

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2016). *Introduction Time Series Analysis and Forecasting*. 671.

PT Bursa Efek Indonesia. (n.d.). Retrieved February 14, 2020, from <https://www.idx.co.id/perusahaan-tercatat/profil-perusahaan-tercatat/detail-profile-perusahaan-tercatat/?kodeEmiten=BBNI>

Rossi, Richard J. *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons, 2018.