

## Prediksi Indeks Bursa Efek Indonesia 2023 Pendekatan ARIMA, Machine Learning dengan R Programming

### *Indonesia Stock Exchange Index Prediction 2023 with the ARIMA Approach, Machine Learning with R Programming*

R. A. E. Virgana Targa Sapanji <sup>1\*</sup>, Sri Lestari <sup>2</sup>, Murnawan <sup>3</sup>, Rosalin Samihardjo <sup>4</sup>

Program Studi Sistem Informasi, Universitas Widyatama, Bandung, Indonesia

\*E-mail: [rae.virgana@widyatama.ac.id](mailto:rae.virgana@widyatama.ac.id)

#### Abstrak

Indeks Bursa Efek Indonesia (BEI) adalah indikator utama pasar saham Indonesia. Masalah dalam penelitian ini adalah *specific issue* untuk menyelesaikan masalah praktis prediksi pergerakan indeks BEI di masa depan memiliki nilai strategis bagi investor, trader, dan perusahaan di Indonesia, tujuan penelitian ini untuk dapat memprediksi indeks BEI sampai akhir tahun 2023, dikarenakan bursa efek memberikan peran besar bagi perekonomian Indonesia sebagai fungsi ekonomi dan keuangan. Pendekatan tradisional seperti ARIMA (Autoregressive Integrated Moving Average) telah digunakan dalam memprediksi pergerakan indeks BEI. Namun, dalam beberapa tahun terakhir, machine learning dan teknik data mining telah menjadi populer sebagai pendekatan alternatif yang lebih canggih. Penelitian ini menggunakan pendekatan gabungan antara ARIMA dan machine learning dengan R Programming. Data indeks BEI harian dari periode Januari 2012 hingga Desember 2022 akan diambil dari Yahoo Finance. Data ini kemudian akan dibersihkan dan diproses menggunakan R Programming. Pendekatan ARIMA akan digunakan sebagai baseline untuk membandingkan performa machine learning. Hasilnya fokus pada perkiraan harga penutupan saham untuk 365 hari ke depan atau rata-rata sampai akhir tahun 2023, nilai Time Series kemungkinan nilai minimal/maksimal prediksi saham BEI 2023, adalah nilai prediksi minimal 6786.212-6849.559, nilai prediksi maksimal 6850.093-7086.012. Kecenderungan menunjukkan pendekatan yang baik dalam memprediksi arah harga penutupan di masa depan.

**Kata kunci:** Bursa Efek Indonesia; ARIMA; Machine Learning, R Programming.

#### Abstract

The Indonesian Stock Exchange Index (BEI) is the main indicator of the Indonesian stock market. The problem in this research is a *specific issue* to solve the practical problem of predicting the future movement of the BEI index which has strategic value for investors, traders and companies in Indonesia. The aim of this research is to be able to predict the BEI index until the end of 2023, because the stock exchange plays a big role for the Indonesian economy as an economic and financial function. Traditional approaches such as ARIMA (Autoregressive Integrated Moving Average) have been used to predict the movement of the BEI index. However, in recent years, machine learning and data mining techniques have become popular as more sophisticated alternative approaches. This research uses a combined approach between ARIMA and machine learning with R Programming. Daily IDX index data from January 2012 to December 2022 will be taken from Yahoo Finance. This data will then be cleaned and processed using R Programming. The ARIMA approach will be used as a baseline to compare machine learning performance. The results focus on the estimated closing stock prices for the next 365 days or the average until the end of 2023. The Time Series value of the possible minimum/maximum predicted value for IDX shares in 2023 is a minimum predicted value of 6786,212 - 6849,559, a maximum predicted value of 6850,093 - 7086,012. Trends represent a good approach in predicting the future direction of closing prices.

**Keywords:** Indonesia Stock Exchange (IDX); ARIMA; Machine Learning, R Programming.

Naskah diterima 12 Ags. 2023; direvisi 26 Sept. 2023; dipublikasikan 04 Okt. 2023.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



## I. PENDAHULUAN

Indeks Bursa Efek Indonesia (BEI) adalah indikator utama pasar saham Indonesia. Masalah dalam penelitian ini adalah *specific issue* untuk menyelesaikan masalah praktis prediksi pergerakan indeks BEI di masa depan memiliki nilai strategis bagi investor, trader, dan perusahaan di Indonesia. Pendekatan tradisional seperti ARIMA (Autoregressive Integrated Moving Average) telah digunakan dalam memprediksi pergerakan indeks BEI. Namun, dalam beberapa tahun terakhir, machine learning dan teknik data mining telah menjadi populer sebagai pendekatan alternatif yang lebih canggih.

*Auto-Regressive Integrated Moving Average (ARIMA)* (Box dan Jenkins, 1976) didasarkan pada teori proses stokastik dan memiliki fitur kebutuhan data yang lebih sedikit, struktur sederhana dan pemodelan cepat. Selain itu, mereka dapat menangani rangkaian waktu yang tidak stasioner [1]. Tinjauan kemajuan pembelajaran mesin terbaru untuk peramalan, sudut kekuatan ARIMA : Membutuhkan sejumlah kecil data, sederhana, cepat, fleksibel, dan dapat beradaptasi dengan berbagai jenis deret waktu, dari sudut kelemahan ARIMA : Tidak dapat memodelkan pola non-linier dalam deret waktu dan tidak dapat diterapkan pada kasus multivariat [2].

Penambangan data cerdas, termasuk jaringan saraf, telah digunakan secara luas oleh para akademisi selama dekade terakhir. Dalam ekonomi yang berkembang pesat saat ini, prediksi dan analisis data pasar saham memainkan peran penting. Beberapa model non-linear seperti neural network, generalized autoregressive conditional heteroskedasticity (GARCH) dan autoregressive conditional heteroscedasticity (ARCH) serta model linier seperti Auto-Regressive Integrated Moving Average (ARIMA), Moving Average (MA) dan Auto Regressive (AR) dapat digunakan untuk peramalan saham. Arsitektur pembelajaran mendalam termasuk Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), Multilayer Perceptron (MLP) dan Support Vector Machine (SVM) digunakan dalam makalah ini untuk prediksi harga saham suatu organisasi dengan menggunakan harga saham yang tersedia sebelumnya. Dataset National Stock Exchange (NSE) India digunakan untuk melatih model dengan harga penutupan harian [3].

Memprediksi nilai saham yang diperdagangkan secara publik menggunakan machine learning berbasis SARIMA dan XGBoost menjadi topik penelitian ini. Saat memprediksi pasar saham, musim memainkan peran penting dalam model SARIMA, yang mirip dengan model ARIMA. Terbuka, harga akhir hari yang disesuaikan, puncak hari, rendah hari, dan volume total semuanya diperhitungkan dalam XGBoost, penerapan pohon keputusan yang ditingkatkan gradien yang dioptimalkan untuk kecepatan dan kinerja [4].

Peramalan indeks saham adalah isu hangat di arena keuangan. Karena pergerakan indeks saham bersifat non-linear dan tunduk pada banyak faktor internal dan eksternal, hal itu menimbulkan tantangan besar bagi para peneliti yang mencoba memprediksinya. Dalam makalah ini, memilih jaringan neural fungsi basis radial (RBFNN) untuk melatih data dan meramalkan indeks saham Bursa Efek Shanghai. Algoritme segerombolan ikan buatan (AFSA) untuk mengoptimalkan RBF. Untuk meningkatkan efisiensi peramalan, algoritma pengelompokan K-means dioptimalkan oleh AFSA dalam proses pembelajaran RBF [5].

Meningkatnya penggunaan data deret waktu telah memulai banyak upaya penelitian dan pengembangan di bidang penambangan data. Banyaknya penelitian tentang data mining deret waktu dalam dekade terakhir dapat menghambat masuknya peneliti yang tertarik, karena kerumitannya. Dalam makalah ini, diberikan revisi komprehensif pada penelitian data mining deret waktu yang ada. Mereka umumnya dikategorikan ke dalam representasi dan pengindeksan, ukuran kesamaan, segmentasi, visualisasi dan penambangan. Selain itu, isu-isu penelitian mutakhir juga disorot [6].

Peramalan keuangan adalah arah penelitian statistik matematika tradisional dan menggabungkan teknologi penambangan data yang muncul seperti pembelajaran mesin dan pembelajaran mendalam untuk membuat perkiraan. Makalah ini dimulai dari model deret waktu tradisional, yang dapat menangani heteroskedastisitas bersyarat dari data deret waktu dan menggabungkan keunggulan mesin vektor pendukung (SVM) yang dapat menangani bagian nonlinier dari data waktu finansial. Prakiraan pasar saham. Untuk menguji efek prediksi model dan membandingkan metode prediksi tradisional dengan metode prediksi pembelajaran mesin modern, makalah ini menetapkan model ARIMA-GARCH, SVM, dan kombinasi ARIMA-GARCH-SVM [7].

Peramalan data keuangan berfluktuasi cepat dan berfrekuensi tinggi selalu menjadi masalah yang menantang di bidang ekonomi dan pemodelan. Dalam studi ini, model hybrid baru dengan kekuatan derivatif urutan fraksional disajikan dengan fitur dinamis pembelajaran mendalam, memori jangka pendek jangka panjang (LSTM) jaringan, untuk memprediksi variasi stokastik pasar keuangan yang tiba-tiba. Harga pasar saham bersifat dinamis, sangat sensitif, nonlinier, dan kacau. Ada berbagai teknik untuk memperkirakan harga dalam domain varian waktu dan karena variabilitas dan perilaku tidak pasti dalam harga saham, metode tradisional, seperti penambangan data, pendekatan statistik, dan model jaringan saraf non-dalam tidak cocok untuk prediksi dan peramalan umum. harga saham. Sementara model autoregressive fractional integrated moving average (ARFIMA) menyediakan alat yang fleksibel untuk kelas model memori panjang. Kemajuan pemodelan non-linear mendalam berbasis pembelajaran mesin menegaskan bahwa model hybrid secara efisien mengekstraksi fitur mendalam dan memodelkan fungsi non-linear [8].

Data mining dan machine learning merupakan satu kemajuan konsep teknologi saat ini yang dapat memprediksi atau meramalkan. Metode Time Series merupakan satu metode peramalan yang dapat digunakan dengan mempertimbangkan nilai inflasi yang akan datang. Time Series adalah metode untuk melakukan peramalan pada interval waktu tertentu. Model Auto Regressive Integrated Moving Average (ARIMA) merupakan tool dalam peramalan dengan data time series sebagai landasan dalam peramalan [9].

Metode untuk menggabungkan informasi secara kreatif dari berita keuangan dan umpan Twitter ke dalam memprediksi harga portofolio saham, menggunakan kerangka model Multivariate Bayesian Structural Time Series (MBSTS). MBSTS adalah model pembelajaran mesin Bayesian yang dirancang untuk menangkap korelasi di antara beberapa deret waktu target, sambil menggunakan sejumlah prediktor kontemporer. Sebagai ilustrasi model saat ini, menggunakan data pada dua perusahaan perdagangan online terkemuka, yaitu Amazon dan eBay, dan menjalankan eksperimen empiris ekstensif untuk memeriksa mana, jika ada, prediktor penambahan teks akan menambah prediktibilitas harga saham. Evaluasi model bersaing seperti model autoregressive integrated moving average (ARIMA), dan model recurrent neural network (RNN) dengan long short term memory (LSTM), dalam hal kinerja mereka sehubungan dengan kesalahan perkiraan kumulatif satu langkah ke depan dengan dan tanpa prediktor sentimental [10].

Pembelajaran mesin adalah bagian dari kecerdasan buatan yang dapat membuat peramalan masa depan berdasarkan pengalaman sebelumnya. Dalam makalah ini, metode telah diusulkan untuk membangun model berbasis algoritma pembelajaran mesin seperti regresi linier, K-nearest neighbour (KNN), dan juga model statistik seperti Auto-ARIMA dan Facebook's Prophet (Fbprophet). Makalah ini menyajikan kinerja komparatif pembelajaran mesin dan algoritma pemodelan statistik untuk peramalan cryptocurrency [11].

Penggunaan pembelajaran mesin dan arsitektur pembelajaran mendalam untuk memprediksi tren pergerakan harga bitcoin (naik atau turun) untuk prediksi jangka pendek dan memprediksi harga bitcoin untuk prediksi jangka panjang. Mampu memprediksi masa depan pasar terbukti berguna di pasar saham dan karena itu makalah ini memutuskan untuk mereplikasi peluang tersebut untuk disajikan di pasar cryptocurrency juga. Perbandingan model pembelajaran mendalam yang canggih seperti Long Short-Term Memory (LSTM) dengan model pembelajaran mesin yang sukses secara tradisional seperti Random Forest dan ARIMA untuk mengetahui model mana yang menyediakan hasil terbaik [12].

Sepuluh tahun data turunan harian digunakan untuk tujuan pelatihan dan pengujian. Metode yang diadopsi untuk pekerjaan penelitian ini meliputi pembuatan model menggunakan ARIMA, kerangka kerja Hadoop yang terdiri dari pemetaan dan pengurangan untuk analisis data besar. Hasil: Hasil dari pekerjaan ini adalah pengamatan tren yang menunjukkan naik turunnya harga di derivatif, pembuatan grafik kesamaan deret waktu dan memplot frekuensi data temporal. Kesimpulan: Analitik data besar adalah topik yang belum dieksplorasi di pasar derivatif India [13].

Peramalan penjualan akan membantu dalam mengelola pendapatan dan inventaris yang sesuai. Makalah ini mengusulkan sebuah model yang dapat meramalkan segmen yang paling menguntungkan pada tingkat granular. Karena sebagian besar raksasa ritel memiliki banyak cabang di lokasi yang berbeda, konsolidasi penjualan sulit dilakukan dengan menggunakan data mining. Alih-alih menggunakan model pembelajaran mesin membantu mendapatkan hasil yang andal dan akurat [14].

Teknik pemodelan hybrid untuk prediksi harga saham berdasarkan berbagai model machine learning dan deep learning. Hasilnya jelas menunjukkan bahwa model univariat berbasis LSTM, yang menggunakan data historis satu minggu sebagai input untuk memperkirakan nilai penutupan deret waktu Reliance Industries Limited minggu depan, adalah yang paling akurat [15].

Pasar Saham Amerika (New York Stock Exchange dan NASDAQ Stock Exchange). Mempertimbangkan kompleksitas prediksi, Penelitian ini mengusulkan model Autoregressive Integrated Moving Average (ARIMA) untuk memperkirakan nilai harga saham di masa mendatang. ARIMA menunjukkan hasil prediksi yang lebih baik karena dapat menangani data deret waktu dengan sangat baik yang cocok untuk meramalkan indeks saham masa depan [16].

Teknik dan instrumen digunakan untuk meramalkan harga pasar saham seperti sistem saraf tiruan, logika fuzzy, pembelajaran mesin, Support Vector Machine, model ARIMA, pemrograman R. Algoritma yang berbeda digunakan untuk menjalankan semua metode ini dengan lebih tepat seperti Naïve Bayes, K-means, Algoritma genetika dan algoritma penambahan data, dan sebagainya. Tujuan utamanya adalah untuk meningkatkan akurasi peramalan pasar saham [17].

Pemahaman yang tepat terhadap data yang diselidiki tidak hanya mengetahui struktur data, tetapi juga perilaku dan karakteristik data. Analisis statistik menyediakan banyak metode untuk menganalisis dan memahami sifat data. Tujuan dari makalah ini adalah untuk menganalisis dataset deret waktu dan untuk menilai apakah model ARIMA dapat berhasil digunakan untuk memprediksi nilai data deret waktu di masa mendatang [18].

Prediksi harga saham di pasar saham memberikan peran utama dalam bisnis pasar saham. Penggunaan metode konvensional seperti studi fundamental dan teknis mungkin tidak menjamin konsistensi peramalan. Dalam banyak kasus, analisis regresi digunakan untuk meramalkan harga saham. Dalam makalah ini, mensurvei beberapa pendekatan regresi yang kompeten untuk prediksi harga saham di pasar saham [19].

Analitik Prediktif, yang melibatkan penggunaan berbagai model statistik, teknik penambangan data, kecerdasan buatan, pembelajaran mesin, dan lainnya untuk mengekstrak wawasan yang bermakna dari data. Usaha bisnis terus-menerus dibanjiri dengan sejumlah besar data yang dihasilkan dari waktu ke waktu. Jadi ada kebutuhan untuk mengembangkan pendekatan baru untuk meramalkan data deret waktu. Makalah penelitian ini menyajikan analitik prediktif data deret waktu dengan mengambil data dunia nyata, untuk membangun model prediksi yang efektif [20].

Tujuan dari penelitian ini adalah untuk memprediksi pergerakan indeks BEI menggunakan pendekatan ARIMA dan machine learning dengan R Programming. Penelitian ini menggunakan pendekatan gabungan antara ARIMA dan machine learning dengan R Programming. Data indeks BEI harian dari periode Januari 2012 hingga Desember 2022 akan diambil dari Yahoo Finance dan disimpan dalam format CSV. Data ini kemudian akan dibersihkan dan diproses menggunakan R Programming. Pendekatan ARIMA akan digunakan sebagai baseline untuk membandingkan performa machine learning. Beberapa algoritma machine learning seperti Random Forest, Support Vector Machines, dan Neural Networks akan diterapkan untuk memprediksi pergerakan indeks BEI. Performa dari setiap pendekatan akan diuji menggunakan metrik seperti Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE).

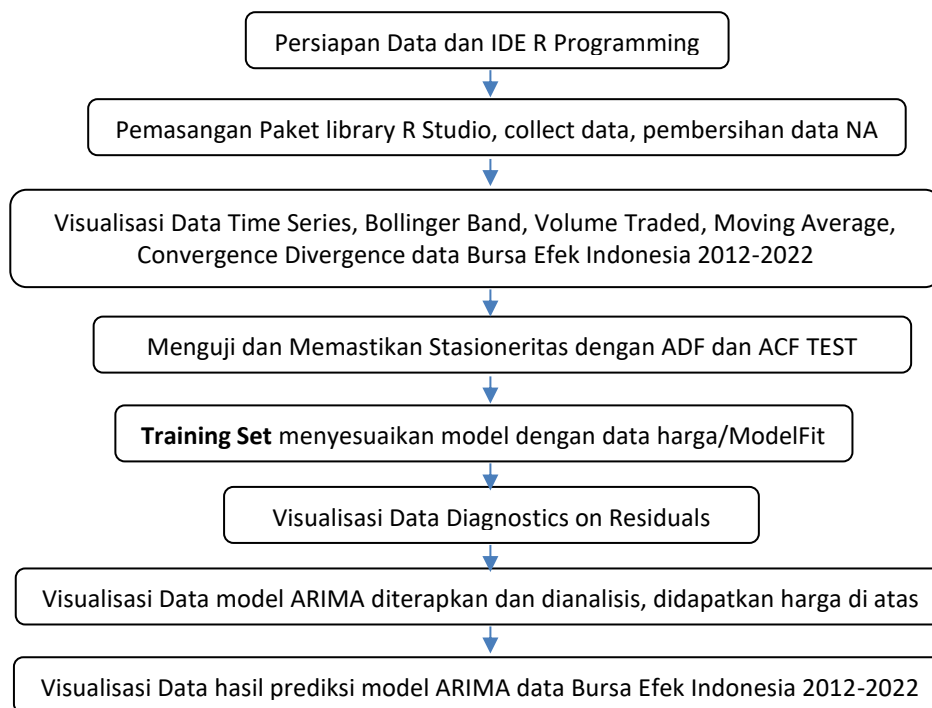
Penelitian ini menghasilkan model prediksi indeks BEI yang akurat dan dapat dipertanggungjawabkan menggunakan kombinasi pendekatan ARIMA dan machine learning. Selain itu, penelitian ini dapat memberikan wawasan baru tentang bagaimana teknik data mining dan machine learning dapat diterapkan dalam memprediksi pergerakan pasar saham di Indonesia.

## II. METODE PENELITIAN

Menurut C. R. Kothari [15], langkah-langkah detail proses untuk memberikan panduan prosedural yang berguna dalam proses penelitian adalah sebagai berikut:

1. *Formulating The Research Problem*, pada tahapan ini kami akan meneliti kembali pentingnya masalah penelitian Analisa Prediksi Indeks Bursa Efek Indonesia Menggunakan Pendekatan ARIMA, Data Mining dan Machine Learning dengan R Programming, diharapkan dapat menghasilkan model prediksi indeks BEI yang akurat dan dapat dipertanggungjawabkan menggunakan kombinasi pendekatan ARIMA dan machine learning. Selain itu, penelitian ini dapat memberikan wawasan baru tentang bagaimana teknik data mining dan machine learning dapat diterapkan dalam memprediksi pergerakan pasar saham di Indonesia. Secara singkat tahapannya adalah persiapan data diambil dari yahoo finance 2012-2022, persiapan IDE R Programming Studio dan *library* pendukung, *collect data* dan pembersihan data NA, menguji dan memastikan Stasioneritas dengan ADF dan ACF TEST, training Set menyesuaikan model dengan data harga/ModelFit, diagnostics on Residuals, model ARIMA diterapkan dan dianalisis untuk didapatkan harga di atas harga penutupan saham, terakhir didapatkan data hasil prediksi model ARIMA.
2. *Extensive Literature Survey*, peneliti akan melakukan survei literature awal telah dilakukan oleh peneliti, dengan menggunakan data eksternal seperti data perekonomian Indonesia, faktor politik, dan lingkungan global sebagai fitur tambahan dalam model prediksi. Selain itu, penelitian ini dapat dilanjutkan dengan mengembangkan strategi perdagangan berdasarkan hasil prediksi indeks BEI..
3. *Development Of Working Hypotheses*, peneliti akan menggunakan data eksternal seperti data perekonomian Indonesia, dan lingkungan global sebagai fitur tambahan dalam model prediksi. Selain itu, penelitian ini dapat dilanjutkan dengan mengembangkan strategi perdagangan berdasarkan hasil prediksi indeks BEI
4. *Preparing The Research Design*, rancangan penelitian sendiri merupakan metodologi ini sendiri, peneliti akan mulai dari memformulasikan masalah penelitian, kemudian melakukan pendekatan gabungan antara ARIMA, data mining dan machine learning dengan R Programming. Data indeks BEI harian dari periode Januari 2010 hingga Desember 2022 akan diambil dari Yahoo Finance dan disimpan dalam format CSV. Data ini kemudian akan dibersihkan dan diproses menggunakan R Programming. Pendekatan ARIMA akan digunakan sebagai baseline untuk membandingkan performa machine learning. Beberapa algoritma machine learning seperti Random Forest, Support Vector Machines, dan Neural Networks akan diterapkan untuk memprediksi pergerakan indeks BEI. Performa dari setiap

pendekatan akan diuji menggunakan metrik seperti Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE).



Gambar 1. Alur Proses Penelitian

5. *Determining Sample Design*, menentukan model prediksi indeks BEI yang akurat dan dapat dipertanggungjawabkan menggunakan kombinasi pendekatan ARIMA dan machine learning. Selain itu, penelitian ini dapat memberikan wawasan baru tentang bagaimana teknik data mining dan machine learning dapat diterapkan dalam memprediksi pergerakan pasar saham di Indonesia.
6. *Collecting The Data*, yaitu mengambil data eksternal seperti data perekonomian Indonesia, dan lingkungan global sebagai fitur tambahan dalam model prediksi.
7. *Execution of The Project*, yaitu melakukan proses data mining, machine learning dan R Programming terhadap data-data sekunder dan primer yang ada, melakukan prediksi dan membuat visualisasi masa depan Bursa Efek Indonesia.
8. *Analysis of Data*, yaitu hasil keluaran analisa gabungan antara ARIMA, data mining dan machine learning dengan R Programming. Data indeks BEI harian dari periode Januari 2010 hingga Desember 2022 akan diambil dari Yahoo Finance dan disimpan dalam format CSV. Data ini kemudian akan dibersihkan dan diproses menggunakan R Programming. Pendekatan ARIMA akan digunakan sebagai baseline untuk membandingkan performa machine learning. Beberapa algoritma machine learning seperti Random Forest, Support Vector Machines, dan Neural Networks akan diterapkan untuk memprediksi pergerakan indeks BEI.
9. *Hypothesis-Testing*, pada tahap ini dengan analisis data yang data akan disimpulkan prediksi pergerakan indeks BEI menggunakan pendekatan ARIMA dan machine learning dengan R Programming.
10. *Generalisations And Interpretation*, yaitu melakukan genearisasi dan interpretasi atas penelitian ini, dan akan memberikan evaluasi dan rekomendasi yang bisa diberikan atas penelitian ini.
11. *Of The Report*, yaitu mempersiapkan laporan akhir penelitian ini.

### III. HASIL DAN PEMBAHASAN

#### Persiapan Data dan IDE R Programming Studio

Langkah-langkah detail proses pembahasan penelitian dimulai dari pengumpulan populasi data Indeks Saham Bursa Efek Indonesia (BEI)/IDX, pada data Yahoo Finance dengan kode *IDX Composite* dengan kode  $\wedge$ JKSE, antara tahun 2012-2022 dengan jumlah data 2733 baris, kemudian persiapan tools IDE R Programming, yaitu R Studio.

≡ Your Workspace / ARIMA - Peramalan Data Deret Waktu Indeks Saham  $\wedge$ JKSE 2012-2022 by RAE

	JKSE.Open	JKSE.High	JKSE.Low	JKSE.Close	JKSE.Volume	JKSE.Adjusted
2012-01-02	3821.626	3825.097	3797.002	3809.140	24861600	3809.140
2012-01-03	3808.689	3858.188	3808.689	3857.882	49712400	3857.882
2012-01-04	3858.063	3921.168	3858.063	3907.421	39077900	3907.421
2012-01-05	3907.511	3924.050	3893.190	3906.264	34539000	3906.264
2012-01-06	3906.444	3906.676	3852.750	3869.415	21634800	3869.415
2012-01-09	3868.784	3890.197	3848.343	3889.072	35483700	3889.072
2012-01-10	3888.712	3941.311	3886.870	3938.842	50193700	3938.842
2012-01-11	3938.932	3954.856	3902.345	3909.640	31509900	3909.640
2012-01-12	3910.000	3918.586	3890.928	3909.497	42546300	3909.497
2012-01-13	3909.046	3943.261	3909.046	3935.326	42892200	3935.326
2012-01-16	3935.236	3935.236	3894.831	3909.693	31210900	3909.693
2012-01-17	3910.059	3955.736	3910.059	3954.755	48105000	3954.755
2012-01-18	3954.575	3978.455	3947.855	3978.128	40722700	3978.128
2012-01-19	3978.489	4027.893	3978.489	4001.073	43535600	4001.073
2012-01-20	4001.805	4024.948	3967.336	3986.515	36347900	3986.515
2012-01-24	3986.965	4038.785	3985.507	3994.583	20738300	3994.583
2012-01-25	3995.123	3997.110	3948.960	3963.605	26529100	3963.605
2012-01-26	3963.785	3983.506	3950.735	3983.434	32829700	3983.434
2012-01-27	3983.524	3992.899	3969.456	3986.410	27714300	3986.410
2012-01-30	3986.140	3986.140	3896.400	3915.160	28566300	3915.160
2012-01-31	3914.620	3941.693	3902.155	3941.693	34190200	3941.693
2012-02-01	3941.783	3965.323	3936.809	3964.976	36201200	3964.976

Showing 1 to 23 of 2,733 entries, 6 total columns

≡ Your Workspace / ARIMA - Peramalan Data Deret Waktu Indeks Saham  $\wedge$ JKSE 2012-2022 by RAE

	JKSE.Open	JKSE.High	JKSE.Low	JKSE.Close	JKSE.Volume	JKSE.Adjusted
2022-12-01	7081.195	7090.277	7018.269	7020.803	167994300	7020.803
2022-12-02	7020.803	7021.814	6967.950	7019.639	139459800	7019.639
2022-12-05	7019.639	7053.903	6987.328	6987.328	165306000	6987.328
2022-12-06	6987.368	6987.368	6862.565	6892.570	149697900	6892.570
2022-12-07	6892.570	6892.663	6799.300	6818.752	158724300	6818.752
2022-12-08	6818.791	6821.051	6683.629	6804.227	162316700	6804.227
2022-12-09	6804.227	6804.227	6695.382	6715.118	127123900	6715.118
2022-12-12	6715.118	6734.452	6641.810	6734.452	134471100	6734.452
2022-12-13	6734.491	6826.116	6714.872	6810.316	396136700	6810.316
2022-12-14	6810.356	6854.098	6774.132	6801.749	215430000	6801.749
2022-12-15	6801.789	6801.838	6740.955	6751.860	145093400	6751.860
2022-12-16	6751.860	6812.193	6693.765	6812.193	156126900	6812.193
2022-12-19	6812.154	6827.815	6754.088	6779.698	127896600	6779.698
2022-12-20	6779.698	6792.203	6715.045	6768.316	151817000	6768.316
2022-12-21	6768.316	6820.663	6763.697	6820.663	139079800	6820.663
2022-12-22	6820.702	6844.121	6800.617	6824.432	129012900	6824.432
2022-12-23	6824.393	6824.393	6784.587	6800.673	120430300	6800.673
2022-12-26	6800.713	6858.147	6796.167	6835.808	117543200	6835.808
2022-12-27	6835.847	6933.889	6835.662	6923.027	121669200	6923.027
2022-12-28	6923.066	6953.039	6828.140	6850.520	120571100	6850.520
2022-12-29	6850.520	6879.585	6786.982	6860.077	141016100	6860.077
2022-12-30	6860.116	6888.736	6838.589	6850.619	129302000	6850.619

Showing 2,711 to 2,733 of 2,733 entries, 6 total columns

Gambar 2. Contoh 2733 Populasi Data IDX Composite/ $\wedge$ JKSE antara tahun 2012-2022

#### Pemasangan Paket library R Studio, collect data, pembersihan data NA

Setelah *new project* di IDE R Programming Studio disiapkan, langkah selanjutnya adalah instalasi dan pemasangan paket *library* untuk pengolahan data sebagai berikut :

```
#libraries
install.packages("quantmod")
install.packages("ggplot2")
install.packages("forecast")
```

```
install.packages("tseries")
install.packages("rugarch")
install.packages("prophet")
install.packages("tsfknn")
```

```
library(quantmod)
library(ggplot2)
library(forecast)
library(tseries)
library(rugarch)
library(prophet)
library(tsfknn)
```

Mulai penarikan data *IDX Composite* dari Yahoo Finance data tahun 2012-2022 :  
`getSymbols("^JKSE",src="yahoo",from="2012-01-01",to = "2022-12-31").`

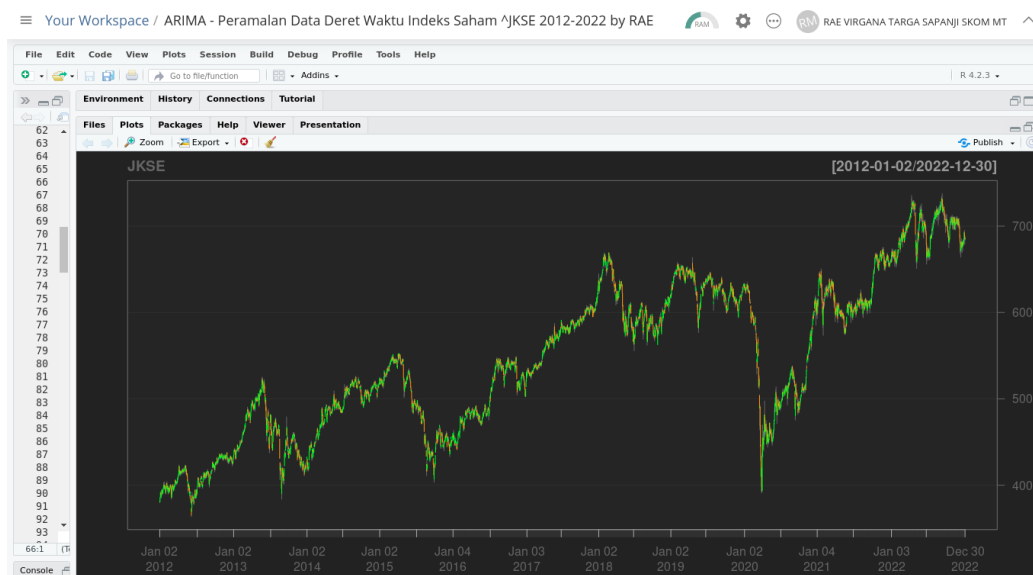
Data mulai terbentuk dengan nama variabel data frame JKSE.

Pemeriksaan data NA, data dengan kondisi kosong/NA tidak dapat proses, data kosong dimungkinkan karena ada libur nasional atau terjadi bursa di hentikan sementara, dan alasan lainnya, untuk melakukan proses pembersihan data kosong/NA ini dilakukan proses berikut :

```
# print ("Posisi Baris dan Kolom dari nilai NA")
which(is.na(JKSE), arr.ind=TRUE)
# hapus semua row yang mengandung NA
JKSE <- JKSE[-c(..indek row data NA.. ), ]
```

#### Visualisasi Data *Time Series*, *Bollinger Band*, *Volume Traded*, *Moving Average*, *Convergence Divergence* data Bursa Efek Indonesia 2012-2022

Visualisasikan data harga penutupan dengan grafik dengan `chartSeries(JKSE,TA = NULL)`



Gambar 3. Visualisasi Data Time Series data harga penutupan Bursa Efek Indonesia 2012-2022

Tampak pada gambar 2., Bursa Efek Indonesia data 10 tahun (2012-2022), menunjukkan posisi terus naik (*bullish stock price*), menunjukkan bursa efek Indonesia memiliki kinerja yang baik dan meningkat, walaupun pada tahun 2020 terjadi pandemi Covid19 terjadi penurunan harga yang kembali ke tahun 2012, setelah pandemi selesai hanya dalam 2 tahun terjadi peningkatan meroket ke harga sebelum pandemi.

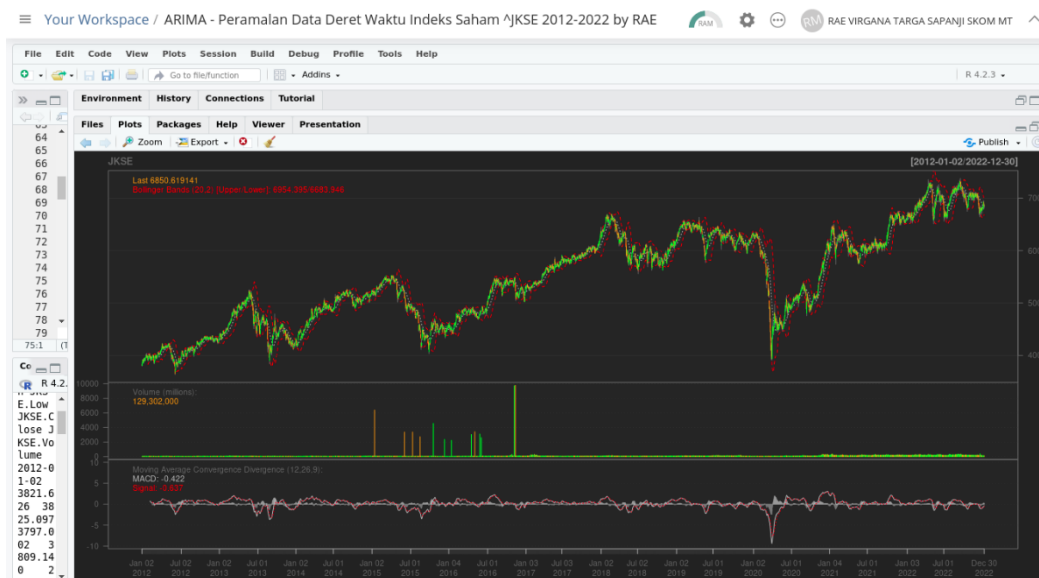


Menambahkan grafik :

```
# Bollinger Band chart, % Bollinger change,
```

```
# Volume Traded and Moving Average Convergence Divergence
```

```
chartSeries(JKSE,TA=c(addVo(),addBBands(),addMACD()))
```



Gambar 4. Visualisasi Data *Bollinger Band*, *Volume Traded*, *Moving Average*, *Convergence Divergence* data Bursa Efek Indonesia 2012-2022 data harga penutupan Bursa Efek Indonesia 2012-2022

Pada gambar 3., pada *candle stick* data Januari 2022 sampai dengan Juli 2022 terjadi *double shoulder candle stick*, dengan *bollinger band* garis merah putus-putus menunjukkan posisi turun/*bearish* sepertinya akan menunjukkan bursa saham BEI akan turun?

Di sisi bawah chart data *volume trade* menunjukkan transaksi normal, walaupun di akhir desember 2022 terjadi penurunan *volume trade*. Disisi paling bawah chart *Moving Average Convergence Divergence* (MACD), diujung akhir tahun 2022 memprediksi sepertinya akan terjadi kenaikan/*bullish* transaksi, terlihat garis merah putus-putus pada MACD menanjak.

### Menguji dan Memastikan Stasioneritas dengan ADF dan ACF TEST

Menguji dan Memastikan Stasioneritas dengan ADF TEST, `print(adf.test(JKSE$JKSE.Close))`

Augmented Dickey-Fuller Test

data: JKSE\$JKSE.Close

Dickey-Fuller = -2.7994, Lag order = 13, p-value = 0.2399

alternative hypothesis: stationary

Menguji stasioneritas menggunakan uji akar unit Augmented Dickey-Fuller, Nilai p yang dihasilkan dari uji ADF harus kurang dari 0,05 atau 5% agar deret waktu stasioner. Jika nilai p lebih besar dari 0,05 atau 5%, disimpulkan bahwa deret waktu memiliki akar unit yang berarti proses tersebut tidak stasioner, tampak nilai p-value tidak stasioner

Nanti akan dilakukan proses Differencing, untuk mengubah proses non-stasioner menjadi proses stasioner.

Setelah uji ADF, dilakukan proses fungsi ACF (fungsi autokorelasi) dan PACF (fungsi autokorelasi parsial) ke dataset, sebagai berikut :

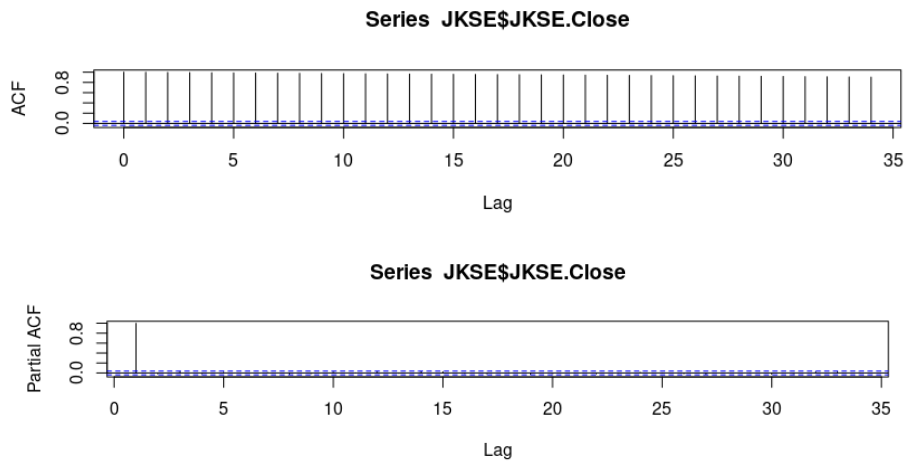
```
#Plot ACF and PACF
```

```
par(mfrow = c(1, 2))
```

```
acf(JKSE$JKSE.Close)
```



```
pacf(JKSE$JKSE.Close)
par(mfrow = c(1, 1))
```



Gambar 5. Proses fungsi ACF (fungsi autokorelasi) dan PACF (fungsi autokorelasi parsial) ke dataset

Autokorelasi mengacu pada seberapa berkorelasi deret waktu dengan nilai masa lalu. Dalam model AR, ACF akan berkurang secara eksponensial. ACF adalah plot yang digunakan untuk melihat korelasi antar titik, sampai dengan dan termasuk satuan lag. Dapat dilihat bahwa autokorelasi signifikan untuk sejumlah besar lag, tetapi mungkin autokorelasi pada lag posterior hanya disebabkan oleh propagasi autokorelasi pada lag pertama.

Plot ACF dan PACF untuk mengidentifikasi urutan (q) dan PACF akan berkurang secara eksponensial. Dapat terlihat lonjakan yang signifikan hanya pada lag pertama, berarti semua autokorelasi orde tinggi secara efektif dijelaskan oleh autokorelasi lag pertama.

#### ***Training Set menyesuaikan model dengan data harga/ModelFit***

```
Proses selanjutnya menyesuaikan model dengan data harga.
## Applying auto.arima() to the dataset
modelfit <- auto.arima(JKSE$JKSE.Close, lambda = "auto")
summary(modelfit)
```

Hasil proses sebagai berikut :

```
Series: JKSE$JKSE.Close
ARIMA(0,1,2)
Box Cox transformation: lambda= 1.269195
```

Coefficients:

```
ma1 ma2
0.0327 -0.0411
s.e. 0.0193 0.0195
```

```
sigma^2 = 283132: log likelihood = -20623.67
AIC=41253.34 AICc=41253.35 BIC=41271.02
```

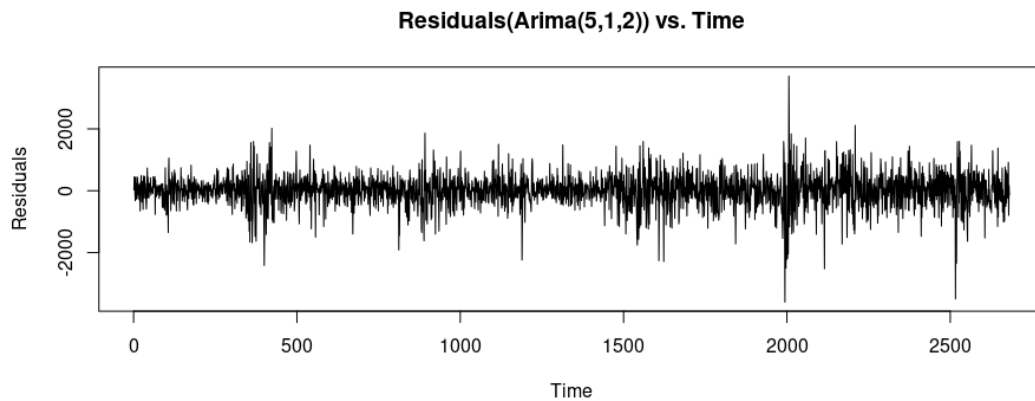
Training set error measures:

```
ME RMSE MAE MPE MAPE MASE ACF1
Training set 1.144278 52.63567 37.9438 0.01691938 0.7120236 0.9987452 0.003696831
```

#### ***Visualisasi Data Diagnostics on Residuals***

Ringkasan model memeriksa residu model dengan parameter ARIMA yang dipilih

```
# Diagnostics on Residuals
plot(resid(modelfit),ylab="Residuals",main="Residuals(Arima(5,1,2)) vs. Time")
```



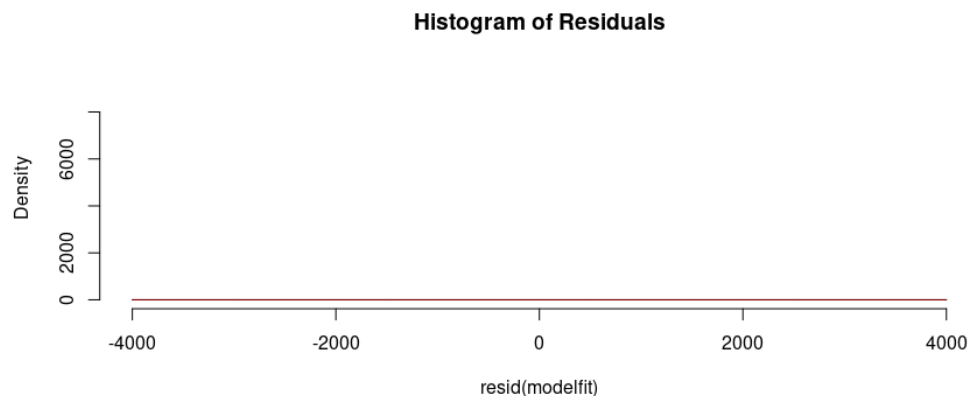
Gambar 6. Ringkasan model memeriksa residu model dengan parameter ARIMA yang dipilih

Residual dalam model deret waktu adalah apa yang tersisa setelah memasang model. Pada sebagian besar model deret waktu, residu sama dengan perbedaan antara pengamatan dan nilai yang dipasang:  $e_t = y_t - \hat{y}_t$ .

Melakukan beberapa tes korelasi pada dataset, memeriksa residual di atas kurva normal.

```
# Histogram of Residuals & Normality Assumption
hist(resid(modelfit),freq=F,ylim=c(0,9500),main="Histogram of Residuals")
e=resid(modelfit)
curve(dnorm(x, mean=mean(e), sd=sd(e)), add=TRUE, col="darkred")
```

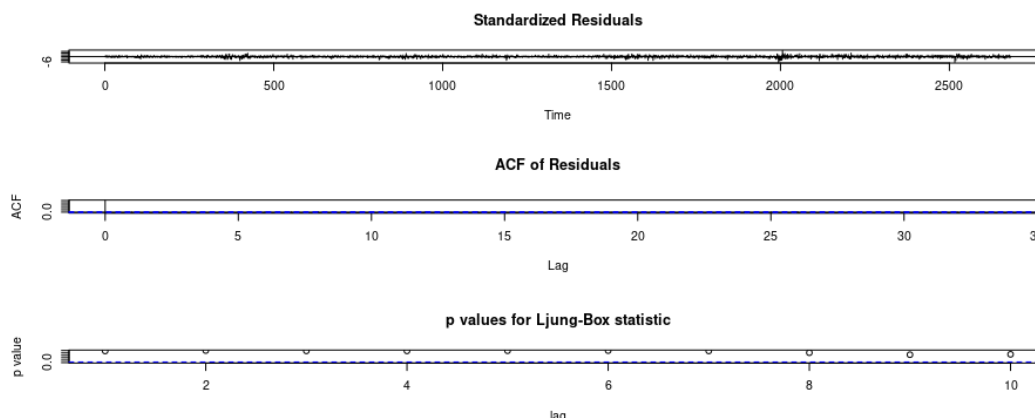
hasilnya sebagai berikut:



Gambar 7. Histogram of Residuals

Seperti yang dapat dilihat, plot residual memiliki penyesuaian kurva normal, memberi poin yang baik untuk melanjutkan penelitian ini.

```
Sekarang membuat plot residual terakhir residual standar,
# ACF residual, dan nilai-p untuk plot statistik Ljung-Box.
# Diagnostics tests for Arima
tsdiag(modelfit)
```



Gambar 8. Plot residual terakhir residual standar

Dengan 3 grafik fokus pada nilai-p Kotak-Ljung. Untuk uji Ljung-Box hipotesis nol adalah:

$H_0$ : The dataset points are independently distributed.

$H_0$ : Titik-titik dataset terdistribusi secara independen.

Dengan hipotesis nol ini, nilai p signifikan lebih besar dari 0,05, tidak menolak fakta bahwa titik dataset tidak berkorelasi.

Pada plot Ljung-Box nilai-p sebelumnya, dapat dilihat bahwa pada lag 1, memiliki nilai-p yang lebih kecil.

Mengingat inspeksi visual diatas, dilanjutkan menganalisis lag ini dengan tes independen.

```
# Box test for lag=2
```

```
Box.test(modelfit$residuals, lag= 2, type="Ljung-Box")
```

Hasilnya:

Box-Ljung test

```
data: modelfit$residuals
```

```
X-squared = 0.0022545, df = 2, p-value = 0.9989
```

Seperti yang dapat dilihat, nilai p masih belum menolak hipotesis nol, memungkinkan untuk membuat uji box test.

```
Box.test(modelfit$residuals, type="Ljung-Box")
```

Hasilnya :

Box-Ljung test

```
data: modelfit$residuals
```

```
X-squared = 0.0020147, df = 1, p-value = 0.9642
```

Dalam tes umum ini dapat dilihat bahwa hipotesis nol masih belum ditolak, memungkinkan untuk melanjutkan penelitian ini.

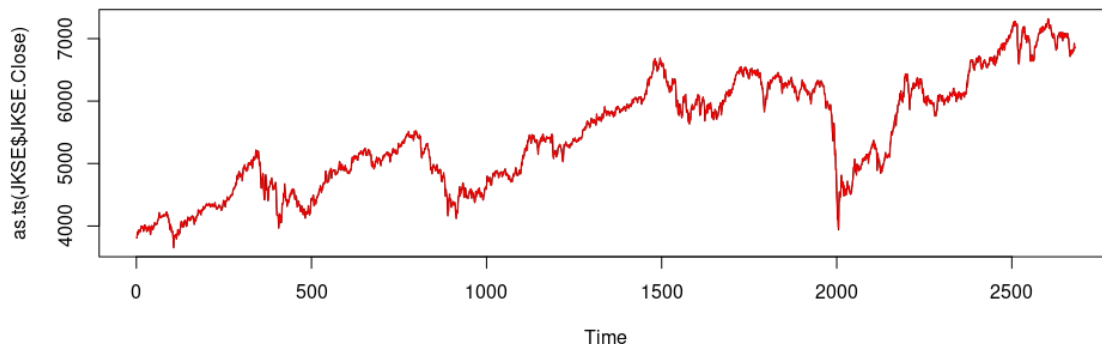
### ***Visualisasi Data model ARIMA diterapkan dan dianalisis, didapatkan harga di atas harga penutupan saham***

Setelah model ARIMA diterapkan dan dianalisis, didapatkan harga di atas harga penutupan saham dengan cara berikut:

```
plot(as.ts(JKSE$JKSE.Close))
```

```
lines(modelfit$fitted,col="red")
```

Hasilnya:



Gambar 9. Model ARIMA diterapkan dan dianalisis, didapatkan harga di atas harga penutupan saham

### Visualisasi Data hasil prediksi model ARIMA data Bursa Efek Indonesia 2012-2022

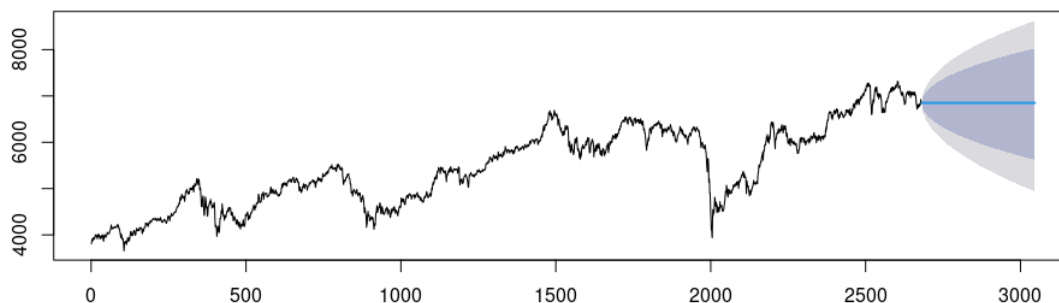
Hasil Arima, sekarang dengan model yang dipasang, dapat dilanjutkan memperkirakan nilai harga penutupan harian ke masa depan, fokus pada perkiraan harga penutupan saham untuk 365 hari ke depan atau rata-rata sampai akhir tahun 2023.

Sekarang merencanakan perkiraan untuk 365 hari ke depan.

```
plot(forecast(modelfit,h=365))
```

Hasilnya :

#### Forecasts from ARIMA(0,1,2)



Gambar 10. Visualisasi Data prediksi harga saham sampai akhir tahun 2023 dengan model ARIMA dengan data Bursa Efek Indonesia 2012-2022

Seperti yang bisa dilihat, grafik memiliki garis biru yang mewakili rata-rata prediksi sampai akhir 2023.

Nilai *Time Series*, kemungkinan nilai minimal/maksimal prediksi saham BEI 2023 :

```
price_forecast <- forecast(modelfit,h=30)
```

```
plot(price_forecast)
```

```
head(price_forecast$mean)
```

Hasilnya:

Time Series:

Start = 2682

End = 2687

Frequency = 1

```
[1] 6849.559 6850.093 6850.093 6850.093 6850.093 6850.093
```

Kesimpulannya nilai *Time Series*, kemungkinan nilai minimal/maksimal prediksi saham BEI 2023, adalah nilai prediksi minimal 6849.559, nilai prediksi maksimal 6850.093

Pada gambar 9 pada garis biru, dapat dilihat area yang lebih gelap dan lebih gelap, mewakili interval kepercayaan 80% dan 95% masing-masing dalam skenario nilai bawah dan atas prediksi saham BEI sampai akhir tahun 2023.

Untuk prediksi kemungkinan nilai saham tertinggi di tahun 2023,

```
head(price_forecast$upper)
```

Hasilnya:

Time Series:

Start = 2682

End = 2687

Frequency = 1

80% 95%

2682 6912.749 6946.137

2683 6940.878 6988.808

2684 6960.342 7018.514

2685 6976.835 7043.677

2686 6991.404 7065.898

2687 7004.595 7086.012

Untuk prediksi kemungkinan nilai saham terbawah di tahun 2023,

```
head(price_forecast$lower)
```

hasilnya :

Time Series:

Start = 2682

End = 2687

Frequency = 1

80% 95%

2682 6786.212 6752.613

2683 6758.982 6710.617

2684 6739.364 6680.550

2685 6722.716 6655.024

2686 6707.993 6632.441

2687 6694.647 6611.965

Menyelesaikan model ARIMA dilakukan *test* dan *training set*, yang membagi data harga penutupan dipilih set nilai 70 persen dari dataset alias menetapkan 30 persen sisa set data.

Pengujian menetapkan 30 persen sisa set data :

```
N = length(JKSE$JKSE.Close)
```

```
n = 0.8*N
```

```
train = JKSE$JKSE.Close[1:n, ]
```

```
test = JKSE$JKSE.Close[(n+1):N,]
```

```
trainarimafit <- auto.arima(train$JKSE.Close ,lambda= "auto")
```

```
summary(trainarimafit)
```

```
predlen= length(test)
```

```
trainarima_fit <- forecast(trainarimafit, h= predlen)
```

Hasilnya:

Series: train\$JKSE.Close

ARIMA(0,1,2)

Box Cox transformation: lambda= 1.958564

Coefficients:

ma1 ma2

0.0591 -0.0581

s.e. 0.0216 0.0221

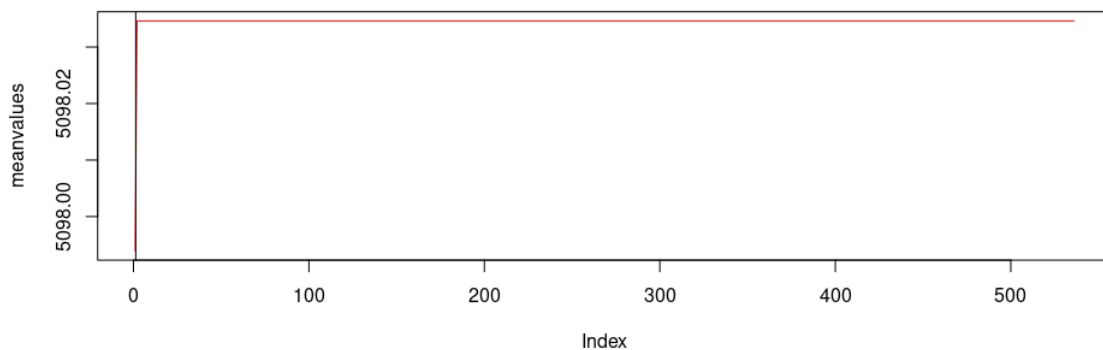
$\sigma^2 = 3.495e+10$ : log likelihood = -29052.85  
 AIC=58111.7 AICc=58111.71 BIC=58128.71

Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1  
 Training set 0.6013339 51.66737 36.73998 0.008161229 0.7248434 0.9979393 0.006640711

Setelah prediksi diterapkan pada rangkaian data set latihan, di plot kecenderungan rata-rata prediksi di atas pergerakan harga penutupan set uji.

```
#Plotting mean predicted values vs real data
meanvalues<- as.vector(trainarima_fit$mean)
precios <- as.vector(test$JKSE.Close)
plot(meanvalues, type = "l",col="red")
lines(precios, type = "l")
#dev.off()
```



Gambar 11. Visualisasi Data nilai real saham dengan nilai prediksi harga saham sampai akhir tahun 2023 dengan model ARIMA dengan data Bursa Efek Indonesia 2012-2022

Pada garis merah, dapat dilihat kecenderungan prediksi perkiraan rata-rata di atas harga penutupan saham yang sebenarnya. Kecenderungan menunjukkan pendekatan yang baik dalam memprediksi arah harga penutupan di masa depan.

#### IV. KESIMPULAN

Masalah dalam penelitian ini adalah *specific issue* untuk menyelesaikan masalah praktis prediksi pergerakan indeks BEI di masa depan memiliki nilai strategis bagi investor, trader, dan perusahaan di Indonesia. Pendekatan tradisional seperti ARIMA (Autoregressive Integrated Moving Average) telah digunakan dalam memprediksi pergerakan indeks BEI. Penelitian ini menggunakan pendekatan gabungan antara ARIMA dan machine learning dengan R Programming. Data indeks BEI harian dari periode Januari 2012 hingga Desember 2022 diambil dari Yahoo Finance. Data kemudian dibersihkan dan diproses menggunakan R Programming. Hasilnya fokus pada perkiraan harga penutupan saham untuk 365 hari ke depan atau rata-rata sampai akhir tahun 2023, nilai Time Series kemungkinan nilai minimal/maksimal prediksi saham BEI 2023 adalah nilai prediksi minimal 6786.212-6849.559, nilai prediksi maksimal 6850.093-7086.012. Kecenderungan menunjukkan pendekatan yang baik dalam memprediksi arah harga penutupan di masa depan. Model ARIMA yang disajikan memiliki hasil yang sangat baik, tetapi tidak memuaskan seperti yang diharapkan, bias hasil terutama dijelaskan oleh pengamatan volatil dari dataset dan rangkaian pasar saham Bursa Efek Jakarta (BEI). Situasi ini menimbulkan kekhawatiran dalam hal memprediksi nilai-nilai baru.

UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada Yayasan Widyatama, Rektorat Universitas Widyatama, Lembaga Penelitian, Pengabdian kepada Masyarakat & Modal Intelektual Universitas Widyatama (LP2M UTama) atas dukungan penelitian ini.

DAFTAR PUSTAKA

- [1] Q. Chen, T. Guan, L. Yun, R. Li, and F. Recknagel, "Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials," *Harmful Algae*, vol. 43, pp. 58–65, Mar. 2015.
- [2] R. C. Cruz, P. R. Costa, S. Vinga, L. Krippahl, and M. B. Lopes, "A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination," *Journal of Marine Science and Engineering*, vol. 9, no. 3. MDPI AG, 01-Mar-2021.
- [3] C. Anand, "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks," *J. Ubiquitous Comput. Commun. Technol.*, vol. 3, no. 2, pp. 122–134, Jul. 2021.
- [4] D. S. Kumar, B. C. Thiruvarangan, C. Sai Nikhil Reddy, A. Vishnu, A. S. Devi, and D. Kavitha, "Analysis and Prediction of Stock Price Using Hybridization of SARIMA and XGBoost," *2022 Int. Conf. Commun. Comput. Internet Things, IC3IoT 2022 - Proc.*, 2022.
- [5] W. Shen, X. Guo, C. Wu, and D. Wu, "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," *Knowledge-Based Syst.*, vol. 24, no. 3, pp. 378–385, Apr. 2011.
- [6] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nat. Rev. Genet.*, vol. 16, no. 6, pp. 321–332, May 2015.
- [7] X. Cao and Z. Zhao, "Research on stock index forecasting based on ARIMA-GARCH and SVM mixed model," *Highlights Sci. Eng. Technol.*, vol. 4, pp. 40–46, Jul. 2022.
- [8] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," *IEEE Access*, vol. 8, pp. 71326–71338, 2020.
- [9] G. L. Ramadhan, D. A. R., and H. Sussanto, "Peramalan Inflasi Indonesia Dengan Seasonal Auto Regressive Integrated Moving Average," *SISTEMASI*, vol. 10, no. 3, p. undefined-undefined, Sep. 2021.
- [10] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.
- [11] H. Parikh, N. Panchal, and A. Sharma, "Cryptocurrency Price Prediction Using Machine Learning," *Lect. Notes Networks Syst.*, vol. 428, pp. 275–285, 2023.
- [12] E. Millikan, P. Subramanian, and M. H. Joseph, "Bitcoin Vision: Using Machine Learning and Data Mining to Predict the Short-Term and Long-Term Price of Bitcoin," *Webology*, vol. 18, no. Special Issue, pp. 751–760, 2021.
- [13] C. M. Abraham, M. S. Elayidom, and T. Santhanakrishnan, "Big Data Analysis for Trend Recognition Using Machine Learning Techniques," *Int. J. Sensors, Wirel. Commun. Control*, vol. 10, no. 4, pp. 540–550, Mar. 2020.
- [14] A. R. Banjanagari and B. Vijaykumar, "Retail giant sales forecasting using machine learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 11, pp. 2408–2411, Sep. 2019.
- [15] M. Hirey, J. Unagar, K. Prabhu, and R. Desai, "Analysis of Stock Price Prediction using Machine Learning Algorithms," *2022 Int. Conf. Adv. Technol. ICONAT 2022*, 2022.
- [16] L. C. Jiang and P. Subramanian, "Forecasting of stock price using autoregressive integrated moving average model," *J. Comput. Theor. Nanosci.*, vol. 16, no. 8, pp. 3519–3524, 2019.
- [17] S. Devi and V. Devaser, "Stock market price prediction using SAP predictive service," *Commun. Comput. Inf. Sci.*, vol. 955, pp. 135–148, 2019.
- [18] D. Borkin, M. Nemeth, and A. Nemethova, "Using Autoregressive Integrated Moving Average (ARIMA) for Prediction of Time Series Data," *Adv. Intell. Syst. Comput.*, vol. 1046, pp. 470–476, 2019.
- [19] B. K. Gupta, M. K. Mallick, and S. Hota, "Survey on stock price forecasting using regression analysis," *Smart Innov. Syst. Technol.*, vol. 153, pp. 147–156, 2021.
- [20] S. M. Idrees, M. A. Alam, P. Agarwal, and L. Ansari, "Effective Predictive Analytics and Modeling Based on Historical Data," *Commun. Comput. Inf. Sci.*, vol. 1046, pp. 552–564, 2019.