

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Penelitian terdahulu**

Penelitian yang dilakukan oleh Ajeng Afifah Muhartini, Oman Sahroni, Septi Dwi Rahmawati, Tanti Febrianti, dan Isnaini Mahuda yang berjudul “Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana”. Penelitian ini bertujuan untuk melakukan prediksi jumlah mahasiswa baru pada Program Studi Manajemen di Fakultas Ekonomi dan Bisnis Universitas Bina Bangsa menggunakan metode regresi sederhana linear. Analisis data prodi Manajemen selama 5 tahun dari 2015 hingga 2019 memberikan akurasi setiap tahunnya. Hasil prediksi menunjukkan tingkat error sebesar 3,444% atau tingkat akurasi sebesar 96,556%. Penelitian ini memberikan informasi yang berguna sebagai acuan untuk memperkirakan jumlah mahasiswa prodi Manajemen di Universitas Bina Bangsa pada tahun-tahun berikutnya, meskipun perlu diingat bahwa prediksi ini belum pasti dan jumlah penerimaan mahasiswa dapat dipengaruhi oleh faktor lain[8].

Penelitian yang dilakukan oleh Ghebyla Najla Ayuni dan Devi Fitriana yang berjudul “Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ”. Penelitian ini bertujuan untuk melakukan prediksi penjualan properti berdasarkan data penjualan properti di PT XYZ. Metode regresi linear digunakan dengan variabel kuantitas properti terjual sebagai variabel akibat dan periode penjualan sebagai variabel penyebab. Pengujian menggunakan *MSE*, *RMSE*, dan *MAPE* dilakukan untuk menguji keakuratan prediksi. Berdasarkan analisis data penjualan selama lima tahun, hasil prediksi menunjukkan jumlah penjualan properti pada bulan berikutnya untuk berbagai tipe properti. Prediksi ini dikategorikan sangat baik berdasarkan hasil pengujian keakuratan yang memenuhi standar dengan nilai *MSE*, *RMSE*, dan *MAPE* yang sesuai[7].

Penelitian yang dilakukan oleh Doni Winarso yang berjudul “Perbandingan Metode Regresi Linier Dan Weighted Moving Average Dalam Meramalkan Jumlah Mahasiswa Pada Periode Tertentu”. Penelitian ini membandingkan metode regresi linear dengan metode weighted moving average (WMA) untuk meramalkan jumlah calon mahasiswa di Universitas Muhammadiyah Riau pada periode tertentu. Pengukuran akurasi menggunakan *MAE*, *MSE*, dan *MAPE*. Hasil penelitian menunjukkan bahwa metode regresi linear memiliki persentase kesalahan sebesar 23%, sedangkan metode WMA memiliki kesalahan sebesar 27,14%. Berdasarkan persentase kesalahan tersebut, metode regresi linear lebih baik dalam meramalkan jumlah calon mahasiswa [9].

Penelitian yang dilakukan oleh Dio Yudha Perdana dan Magdalena A. Ineke Pakereng yang berjudul “Prediksi Tingkat Pengangguran Berdasarkan Data *Time series* Menggunakan Regresi Linear (Studi Kasus : Kota Salatiga)”. Penelitian ini bertujuan untuk memprediksi tingkat pengangguran berdasarkan jumlah angkatan kerja, terutama dalam konteks dampak pandemi COVID-19 yang menyebabkan PHK massal. Metode regresi linear digunakan untuk melakukan peramalan penduduk pengangguran selama 5 tahun ke depan. Hasil prediksi menunjukkan peningkatan jumlah penduduk pada tahun 2022 menjadi 144189 jiwa dan penurunan jumlah pengangguran menjadi 4513 jiwa dibandingkan tahun sebelumnya. Meskipun prediksi ini tidak 100% akurat, disarankan untuk menggunakan metode lain sebagai pembanding dan untuk mengatasi tingkat pengangguran yang tinggi dengan memperluas lapangan pekerjaan, meningkatkan indeks pembangunan manusia, dan memberikan pelatihan kepada tunawisma. Untuk pengembangan penelitian di masa depan, disarankan untuk melibatkan data pengangguran dari kota lain dan menggunakan metode seperti *Fuzzy Time series*, *ARIMA*, atau *Moving Average*[10].

Penelitian yang dilakukan oleh Diva Christalivea dan Magdalena A. Ineke Pakereng yang berjudul “Prediksi Volume Penyaluran Air Minum Kota Salatiga Tahun 2021 Berdasarkan Time Data Series Menggunakan Regresi

Linear”. Penelitian ini bertujuan untuk memprediksi jumlah volume air yang akan didistribusikan berdasarkan jumlah pelanggan. Data yang valid dan jangka waktu yang panjang digunakan untuk membuat prediksi yang akurat. Metode regresi linear digunakan dalam penelitian ini untuk menentukan hubungan antara variabel efek dan variabel kausal. Hasil prediksi menunjukkan perubahan jumlah volume distribusi dan pelanggan pada tahun 2021, dengan total volume diprediksi menjadi 4.927.284 dan jumlah pelanggan sebanyak 28.126 [11].

Penelitian yang dilakukan oleh Laila Qadrini, Asrirawan, Nur Mahmudah, Muhammad Fahmuiddin S dan Ihsan Fathoni Amri yang berjudul “Peramalan Inflow dan Outflow Uang Kartal Bank Indonesia dengan Pendekatan *ARIMA*, *Time series Regression (TSR)*, *ARIMAX*, dan *NN* di Lampung”. Penelitian ini menggunakan data deret waktu untuk meramalkan Inflow dan Outflow uang kartal yang merupakan kebijakan dalam peredaran uang kartal oleh Bank Indonesia. Beberapa metode peramalan seperti *ARIMA*, *Time series Regression (TSR)*, *ARIMAX*, dan *NN* digunakan dalam penelitian ini, dan metode terbaik dipilih berdasarkan akurasi *RMSE*. Hasil penelitian menunjukkan bahwa metode *ARIMAX* memiliki *RMSE* terkecil, sehingga merupakan metode yang paling sesuai untuk meramalkan *Inflow* dan *Outflow* uang kartal [12].

Penelitian yang dilakukan oleh Lisa Widyarsi dan Hardius Usman yang berjudul “Penggunaan Data *Google Trends* untuk Peramalan Tingkat Pengangguran Terbuka di Tingkat Nasional dan Regional di Provinsi Jawa Barat”. Penelitian ini bertujuan untuk memanfaatkan data *Google Trends* dalam meramalkan Tingkat Pengangguran Terbuka di Indonesia dan Jawa Barat. Metode *ARIMAX* digunakan dalam penelitian ini dengan periode penelitian dari Februari 2005 hingga Februari 2020. Hasil penelitian menunjukkan bahwa indeks *Google Trends* dapat digunakan untuk meramalkan Tingkat Pengangguran Terbuka baik secara nasional maupun regional di Jawa Barat dengan tingkat akurasi yang baik [13].

Penelitian yang dilakukan oleh Solikhah Novita Intan, Etik Zukhronah dan Supriyadi Wibowo yang berjudul “Peramalan Banyaknya Pengunjung Pantai Glagah Menggunakan Metode Autoregressive Integrated Moving Average Exogenous (*ARIMAX*) dengan Efek Variasi Kalender”. Penelitian ini berfokus pada Pantai Glagah di Kabupaten Kulon Progo, Yogyakarta, yang merupakan destinasi wisata yang banyak dikunjungi, terutama saat bulan Ramadan. Data kunjungan wisatawan Pantai Glagah menunjukkan peningkatan signifikan selama bulan Idul Fitri. Oleh karena itu, diperlukan metode yang dapat digunakan untuk menganalisis data deret waktu yang mengandung efek variasi kalender, yaitu metode *ARIMAX*. Tujuan dari penelitian ini adalah menemukan model *ARIMAX* terbaik dan memprediksi jumlah wisatawan Pantai Glagah. Hasil penelitian menunjukkan bahwa model *ARIMAX* terbaik adalah *ARIMAX*([24],0,0). Ramalan jumlah pengunjung dari Januari hingga September 2016 adalah 37.211, 21.306, 26.247, 24.148, 28.402, 29.309, 81.724, 26.029, dan 23.688 pengunjung [14].

Penelitian yang dilakukan oleh Cita Meliana, Rochdi Wasono, M. Al Haris, Zamni Haquel Alfiyani dan Eka Yuni Kartika Sari yang berjudul “Peramalan Indeks Harga Saham Gabungan (Ihsg) Menggunakan *ARIMAX* Dengan Variabel Eksogen Covid-19”. Penelitian ini membahas dampak pandemi Covid-19 terhadap perekonomian Indonesia, terutama terlihat pada Indeks Harga Saham Gabungan (IHSG) Jakarta. Data IHSG sebelum pandemi Covid-19 pada tanggal 30 Desember 2019 adalah 6299,54. Setelah pandemi Covid-19 meluas di Indonesia mulai 2 Maret 2020, angka IHSG menurun menjadi 5361,25 dan semakin tidak stabil hingga 24 Maret 2020, mencapai 3937,63. Pandemi Covid-19 menyebabkan IHSG menjadi tidak stabil. Data yang digunakan adalah data harian IHSG kecuali hari libur, sebanyak 388 data, dari 2 Januari 2019 hingga 30 Juli 2020. Dalam penelitian ini, metode *ARIMAX* digunakan untuk meramalkan IHSG dengan variabel eksogen covid-19. Pemodelan IHSG menggunakan variabel eksogen covid-19 dilakukan dengan regresi berganda dan menghasilkan model *ARIMAX* dengan tren deterministik dan sisa model *ARIMA* (1,1,2) [15].

Penelitian yang dilakukan oleh Azeilla Putri Bulu Laga, Sri Wahyuningsih dan Memi Nor Hayati yang berjudul “Peramalan Penjualan Pakaian dengan Autoregressive Integrated Moving Average with Exogeneous Input (*ARIMAX*) (Studi Kasus: Penjualan Pakaian di Toko M~Al Samarinda Tahun 2012 s.d 2016)”. Penelitian ini mengkaji tingkat penjualan pakaian yang akan meningkat pada hari libur, perayaan agama, dan persiapan tahun baru. Dalam transaksi jual beli, terkadang barang yang diinginkan oleh konsumen habis terjual, sehingga penting bagi pemilik toko untuk menyediakan pakaian yang akan dijual. Oleh karena itu, metode *ARIMAX* digunakan. *ARIMAX* adalah model *ARIMA* yang memiliki variabel eksogen. Variabel eksogen adalah variabel yang dianggap mempengaruhi variabel lain, seperti hari libur dan perayaan hari besar, dan Hari Raya Idul Fitri merupakan indikator penting dari penjualan pakaian. Tujuan dari penelitian ini adalah untuk menemukan model *ARIMAX* terbaik dalam penjualan pakaian di Toko M~Al dan untuk mengetahui hasil ramalan data penjualan pakaian di Toko M~Al untuk 12 periode mendatang. Hasil peramalan menggunakan metode *ARIMAX* adalah *SARIMA* (1,0,2)(0,0,2)<sub>6</sub> dengan nilai *MAPE* terkecil yaitu 9,01089. Hasil peramalan tertinggi terjadi pada bulan Juni sebesar 945 dan perkiraan terendah pada bulan Desember sebesar 187 [6].

Penelitian yang dilakukan oleh Ilham Rizki Julianto, Indwiarti dan Aniq Atiqi Rohmawati yang berjudul “Prediksi Jumlah Kunjungan Wisatawan Di Jawa Barat Dengan Model *ARIMAX* Dan *SARIMAX* Menggunakan Data Google Trends”. Penelitian ini menggunakan metode peramalan *time series* *ARIMAX* dan *SARIMAX* untuk memprediksi jumlah kunjungan wisatawan mancanegara di Provinsi Jawa Barat. Data kunjungan wisatawan mancanegara dari BPS digunakan sebagai data utama penelitian, sedangkan data dari Google Trends digunakan sebagai variabel eksternal dalam bentuk kueri pencarian. Hasil peramalan menunjukkan bahwa kedua model *ARIMAX* dan *SARIMAX* memberikan hasil yang baik, dengan penurunan nilai *MAPE* pada kisaran 1-2% untuk model *ARIMA* dan *ARIMAX*, serta 1-3% untuk model *SARIMA* dan *SARIMAX*. Model *SARIMAX* memiliki *MAPE* terendah

sebesar 16,61%. Penelitian ini dapat membantu pengelola destinasi wisata dalam meningkatkan layanan dan infrastruktur wisata berdasarkan kebutuhan yang diperoleh dari peramalan tersebut [5].

Ringkasan penelitian yang relevan ditunjukkan pada Tabel 2.1.

Tabel 2. 1 Penelitian terdahulu

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana oleh Ajeng Afifah Muhartini, Oman Sahroni, Septi Dwi Rahmawati, Tanti Febrianti, dan Isnaini Mahuda	Penelitian ini membandingkan jumlah penerimaan mahasiswa baru pada Prodi Manajemen Fakultas Ekonomi dan Bisnis di Universitas Bina Bangsa selama periode 2015-2019. Data tersebut digunakan untuk melakukan analisis regresi sederhana linear guna memperkirakan jumlah mahasiswa	Meskipun penelitian ini memberikan prediksi peningkatan jumlah penerimaan mahasiswa baru pada Prodi Manajemen di Universitas Bina Bangsa, perlu diingat bahwa hasil prediksi tersebut tidak dapat dijamin 100% akurat. Faktor lain, seperti perubahan kebijakan penerimaan mahasiswa atau perubahan kondisi eksternal, dapat mempengaruhi jumlah	Penelitian ini hanya menggunakan metode regresi sederhana linear untuk meramalkan jumlah mahasiswa baru. Metode ini mungkin terlalu sederhana untuk mempertimbangkan faktor-faktor kompleks yang dapat memengaruhi penerimaan mahasiswa baru, seperti tren demografi, perubahan preferensi siswa, atau kondisi pasar tenaga kerja. Penggunaan metode yang lebih kompleks mungkin	Penelitian ini menggunakan data penerimaan mahasiswa baru selama lima tahun di Prodi Manajemen Fakultas Ekonomi dan Bisnis Universitas Bina Bangsa untuk memprediksi jumlah mahasiswa baru di masa yang akan datang. Dengan tingkat akurasi sebesar 96,556%, hasil prediksi ini dapat digunakan sebagai panduan untuk memperkirakan jumlah mahasiswa baru di tahun	Penelitian ini menggunakan metode regresi sederhana linear untuk memprediksi jumlah penerimaan mahasiswa baru pada Prodi Manajemen dan Bisnis di Universitas Bina Bangsa. Hasil prediksi menunjukkan tingkat akurasi sebesar 96,556%. Namun, prediksi ini perlu dipertimbangkan

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
	baru di masa yang akan datang.	penerimaan mahasiswa baru.	diperlukan untuk memperoleh prediksi yang lebih akurat.	berikutnya. Namun, perlu diingat bahwa faktor-faktor lain dapat mempengaruhi hasil prediksi dan menyebabkan variasi dari angka yang diprediksi.	dengan faktor lain yang dapat mempengaruhi jumlah penerimaan mahasiswa baru.
Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ oleh Ghebyla Najla Ayuni dan Devi Fitriannah	Penelitian ini menggunakan regresi linear sebagai metode prediksi penjualan properti berdasarkan data penjualan properti PT XYZ. Hasil prediksi penjualan properti pada bulan berikutnya untuk berbagai tipe properti disajikan.	Meskipun metode regresi linear memberikan hasil prediksi penjualan properti yang memenuhi standar dan tergolong baik, penelitian ini menyatakan bahwa hasil prediksi yang baik hanya dapat diperoleh dengan data yang memiliki rentang signifikan dan tidak mengalami fluktuasi drastis. Garis	Penelitian ini hanya menggunakan regresi linear dan variabel periode dalam prediksinya, sehingga mungkin ada faktor lain yang dapat mempengaruhi penjualan properti yang tidak dipertimbangkan. Penggunaan variabel lain atau metode yang lebih kompleks mungkin	Penelitian ini menggunakan regresi linear untuk memprediksi penjualan properti berdasarkan data penjualan properti PT XYZ. Hasil prediksi menunjukkan penjualan yang diharapkan untuk berbagai tipe properti pada bulan berikutnya. Metode ini tergolong baik dan dapat digunakan	Penelitian ini menggunakan metode regresi linear untuk memprediksi penjualan properti berdasarkan data penjualan PT XYZ. Hasil prediksi menunjukkan penjualan yang diharapkan untuk berbagai tipe properti. Namun,



<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
		linear hanya menunjukkan peningkatan atau penurunan, tanpa menggambarkan fluktuasi naik-turun.	diperlukan untuk meningkatkan akurasi prediksi.	sebagai pertimbangan dalam pengambilan keputusan pembangunan di sektor properti.	hasil prediksi yang baik hanya dapat diperoleh dengan data yang memiliki rentang signifikan dan tanpa fluktuasi drastis. Penelitian ini dapat digunakan sebagai panduan dalam pengambilan keputusan pembangunan sektor properti, tetapi perlu dipertimbangkan faktor-faktor lain yang mungkin mempengaruhi penjualan properti.
Perbandingan Metode Regresi Linier Dan Weighted Moving	Penelitian ini membandingkan metode regresi linier	Dalam perbandingan antara metode regresi linier dan metode WMA,	Meskipun penelitian ini membandingkan dua metode peramalan, yaitu	Penelitian ini membandingkan metode regresi linier dan metode	Penelitian ini membandingkan metode regresi linier

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
Average Dalam Meramalkan Jumlah Mahasiswa Pada Periode Tertentu oleh Doni Winarso	dengan metode weighted moving average (WMA) dalam meramalkan jumlah calon mahasiswa pada Universitas Muhammadiyah Riau. Hasil penelitian menunjukkan bahwa metode regresi linier memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode WMA.	penelitian ini menemukan bahwa metode regresi linier memiliki persentase error sebesar 23%, sedangkan metode WMA memiliki persentase error sebesar 27.14%. Hal ini menunjukkan bahwa metode regresi linier lebih unggul dalam meramalkan jumlah calon mahasiswa.	regresi linier dan WMA, tidak disertakan penjelasan mengenai kelebihan dan kekurangan masing-masing metode. Selain itu, penelitian ini hanya mempertimbangkan tiga ukuran error peramalan (MAD, MSE, dan MAPE) sebagai indikator akurasi, tanpa mempertimbangkan faktor-faktor lain yang mungkin mempengaruhi hasil peramalan.	weighted moving average (WMA) dalam meramalkan jumlah calon mahasiswa pada Universitas Muhammadiyah Riau. Hasil penelitian menunjukkan bahwa metode regresi linier memiliki tingkat akurasi yang lebih baik dengan persentase error sebesar 23% dibandingkan dengan metode WMA yang memiliki persentase error sebesar 27.14%. Penelitian ini menyarankan untuk membandingkan metode peramalan lainnya guna mencari metode terbaik	dan metode weighted moving average (WMA) dalam meramalkan jumlah calon mahasiswa di Universitas Muhammadiyah Riau. Hasil penelitian menunjukkan bahwa metode regresi linier memiliki tingkat akurasi yang lebih baik, dengan persentase error sebesar 23%. Penelitian ini menyarankan untuk melibatkan metode peramalan lainnya dalam penelitian selanjutnya guna

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
				untuk meramalkan jumlah calon mahasiswa di masa depan.	memperoleh metode terbaik untuk meramalkan jumlah calon mahasiswa.
<p>Prediksi Tingkat Pengangguran Berdasarkan Data <i>Time series</i> Menggunakan Regresi Linear (Studi Kasus : Kota Salatiga) oleh Dio Yudha Perdana dan Magdalena A. Ineke Pakereng</p>	<p>Penelitian ini membandingkan metode regresi linear dengan metode lain seperti <i>Fuzzy Time series</i>, <i>ARIMA</i>, atau <i>Moving Average</i> untuk meramalkan tingkat pengangguran berdasarkan jumlah angkatan kerja. Namun, penelitian ini tidak memberikan hasil perbandingan antara metode regresi</p>	<p>Dalam penelitian ini, penggunaan metode regresi linear digunakan untuk memprediksi tingkat pengangguran berdasarkan jumlah angkatan kerja. Namun, penelitian ini tidak membandingkan secara langsung metode regresi linear dengan metode lainnya, sehingga tidak ada perbandingan antara keunggulan dan</p>	<p>Penelitian ini tidak menjelaskan secara rinci bagaimana validitas data yang digunakan untuk peramalan. Selain itu, penelitian ini hanya menggunakan metode regresi linear tanpa mempertimbangkan metode lain yang mungkin lebih sesuai atau memiliki tingkat akurasi yang lebih baik untuk meramalkan tingkat pengangguran.</p>	<p>Penelitian ini bertujuan untuk memprediksi tingkat pengangguran berdasarkan jumlah angkatan kerja menggunakan metode regresi linear. Hasil penelitian menunjukkan bahwa pada tahun 2022, jumlah penduduk mengalami peningkatan sedangkan jumlah pengangguran mengalami penurunan dibandingkan tahun</p>	<p>Penelitian ini menggunakan metode regresi linear untuk memprediksi tingkat pengangguran berdasarkan jumlah angkatan kerja. Hasil penelitian menunjukkan peningkatan jumlah penduduk dan penurunan jumlah pengangguran pada tahun 2022. Namun, penelitian ini tidak</p>

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
	linear berganda dan metode lainnya.	kelemahan masing-masing metode.	Sebuah penelitian yang lebih komprehensif dan mempertimbangkan berbagai metode peramalan dapat memberikan hasil yang lebih baik.	sebelumnya. Penelitian ini menyarankan penggunaan metode lain seperti <i>Fuzzy Time series</i> , <i>ARIMA</i> , atau <i>Moving Average</i> serta melibatkan data pengangguran dari kota lain untuk pengembangan penelitian selanjutnya.	membandingkan secara langsung metode regresi linear dengan metode lainnya dan tidak menjelaskan validitas data yang digunakan. Penelitian selanjutnya disarankan untuk mempertimbangkan metode peramalan lainnya dan melibatkan data pengangguran dari kota lain untuk hasil yang lebih komprehensif.
Prediksi Volume Penyaluran Air Minum Kota Salatiga Tahun	Penelitian ini menggunakan metode regresi linear	Dalam penelitian ini, metode regresi linear digunakan untuk	Penelitian ini hanya menggunakan metode regresi linear untuk	Penelitian ini menggunakan metode regresi linear untuk	Penelitian ini menggunakan metode regresi linear

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
2021 Berdasarkan Time Data Series Menggunakan Regresi Linear oleh Diva Christalivea dan Magdalena A. Ineke Pakereng	untuk memprediksi jumlah volume air yang akan didistribusikan berdasarkan jumlah pelanggan. Namun, penelitian ini tidak membandingkan metode regresi linear dengan metode lainnya untuk melihat keunggulan dan kelemahan masing-masing metode dalam melakukan prediksi volume air.	memprediksi volume air yang akan didistribusikan berdasarkan jumlah pelanggan. Namun, penelitian ini tidak memberikan perbandingan antara metode regresi linear dan metode lainnya, sehingga kita tidak dapat mengetahui apakah metode ini merupakan metode terbaik untuk melakukan prediksi volume air.	memprediksi volume air berdasarkan jumlah pelanggan. Metode regresi linear memiliki asumsi yang cukup ketat, seperti adanya hubungan linier antara variabel kausal dan variabel efek. Penelitian ini juga tidak menyebutkan faktor-faktor lain yang mungkin memengaruhi volume air yang akan didistribusikan, seperti musim, cuaca, atau kebijakan pengelolaan air. Penggunaan metode lain atau mempertimbangkan faktor-faktor tambahan	memprediksi jumlah volume air yang akan didistribusikan berdasarkan jumlah pelanggan. Namun, penelitian ini tidak membandingkan metode regresi linear dengan metode lainnya dan tidak mempertimbangkan faktor-faktor lain yang mungkin memengaruhi volume air. Sebagai hasilnya, perlu dipertimbangkan penggunaan metode lain dan memperhatikan faktor-faktor tambahan dalam penelitian selanjutnya untuk	untuk memprediksi jumlah volume air yang akan didistribusikan berdasarkan jumlah pelanggan. Namun, penelitian ini tidak melakukan perbandingan dengan metode lainnya dan tidak mempertimbangkan faktor-faktor tambahan yang dapat memengaruhi volume air. Oleh karena itu, penelitian selanjutnya disarankan untuk mempertimbangkan penggunaan metode lain dan memasukkan

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
			dapat meningkatkan akurasi prediksi.	meningkatkan akurasi prediksi.	faktor-faktor tambahan untuk meningkatkan akurasi prediksi.
Peramalan Inflow dan Outflow Uang Kartal Bank Indonesia dengan Pendekatan <i>ARIMA</i> , <i>Time series</i> Regression (TSR), <i>ARIMAX</i> , dan NN di Lampung oleh Laila Qadrini , Asrirawan , Nur Mahmudah, Muhammad Fahmuddin S dan Ihsan Fathoni Amri	Penelitian ini membandingkan beberapa metode peramalan, yaitu <i>ARIMA</i> , <i>Time series</i> Regression (TSR), <i>ARIMAX</i> , dan NN, untuk meramalkan Inflow dan Outflow uang kartal. Penelitian ini menggunakan nilai Root Mean Squared Error (RMSE) sebagai metrik akurasi untuk memilih metode	Dalam penelitian ini, terdapat perbandingan antara beberapa metode peramalan, tetapi tidak dijelaskan secara rinci bagaimana masing-masing metode bekerja dan mengapa <i>ARIMAX</i> dipilih sebagai metode terbaik. Selain itu, tidak ada perbandingan dengan metode lain yang tidak termasuk dalam penelitian ini, sehingga tidak dapat diketahui apakah <i>ARIMAX</i> benar-	Penelitian ini hanya menggunakan RMSE sebagai metrik akurasi untuk memilih metode terbaik. Meskipun RMSE adalah metrik yang umum digunakan, penting untuk mempertimbangkan metrik lainnya seperti Mean Absolute Percentage Error (MAPE) atau membandingkan grafik peramalan dengan data aktual untuk melihat	Penelitian ini membandingkan beberapa metode peramalan, yaitu <i>ARIMA</i> , TSR, <i>ARIMAX</i> , dan NN, untuk meramalkan Inflow dan Outflow uang kartal. Dalam penelitian ini, metode <i>ARIMAX</i> dipilih sebagai metode terbaik berdasarkan RMSE yang terkecil. Namun, penelitian ini tidak memberikan penjelasan rinci tentang bagaimana masing-masing metode bekerja	Penelitian ini membandingkan beberapa metode peramalan, seperti <i>ARIMA</i> , TSR, <i>ARIMAX</i> , dan NN, untuk meramalkan Inflow dan Outflow uang kartal. Metode <i>ARIMAX</i> dipilih sebagai metode terbaik berdasarkan RMSE terkecil. Namun, penelitian ini tidak memberikan penjelasan rinci tentang metode yang

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
	terbaik. Hasilnya menunjukkan bahwa metode <i>ARIMAX</i> memiliki RMSE terkecil dibandingkan dengan metode lainnya.	benar metode terbaik yang tersedia.	performa metode secara holistik.	dan tidak membandingkan dengan metode lain yang tidak termasuk dalam penelitian ini.	digunakan dan tidak membandingkan dengan metode lain yang tidak termasuk dalam penelitian ini. Saran untuk penelitian selanjutnya adalah mempertimbangkan penggunaan metrik akurasi yang lain dan menjelaskan dengan lebih detail tentang masing-masing metode yang digunakan.
Penggunaan Data <i>Google Trends</i> untuk Peramalan Tingkat Pengangguran Terbuka di Tingkat Nasional	Penelitian ini membandingkan penggunaan data <i>Google Trends</i> dengan data official	Penelitian ini mengontraskan antara official statistics yang memiliki lag time dan frekuensi survei terbatas	Meskipun penelitian ini menunjukkan bahwa data <i>Google Trends</i> dapat memberikan akurasi yang baik dalam	Penelitian ini memanfaatkan data <i>Google Trends</i> sebagai komplementer official statistics dalam	Penelitian ini memanfaatkan data <i>Google Trends</i> sebagai komplementer

Penelitian	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
<p>dan Regional di Provinsi Jawa Barat oleh Lisa Widyarsi dan Hardius Usman</p>	<p>statistics dalam meramalkan Tingkat Pengangguran Terbuka (TPT) di Indonesia dan Jawa Barat. Hasil penelitian menunjukkan bahwa data <i>Google Trends</i> dapat digunakan sebagai komplementer <i>official statistics</i> dan memberikan akurasi yang baik dalam meramalkan TPT.</p>	<p>dengan data <i>Google Trends</i> yang memiliki frekuensi lebih banyak dan disajikan secara real time. Dengan menggunakan data <i>Google Trends</i>, penelitian ini mencoba mengatasi keterbatasan <i>official statistics</i> dalam meramalkan TPT.</p>	<p>meramalkan TPT, namun perlu diingat bahwa penggunaan data historis dari <i>Google Trends</i> untuk peramalan harus hati-hati terutama dalam kondisi ekstrim seperti pandemi COVID-19. Selain itu, penelitian ini tidak memberikan penjelasan rinci tentang metode <i>ARIMAX</i> yang digunakan, sehingga sulit untuk mengevaluasi kehandalan metode tersebut.</p>	<p>meramalkan Tingkat Pengangguran Terbuka di Indonesia dan Jawa Barat. Hasil penelitian menunjukkan bahwa data <i>Google Trends</i> dapat digunakan untuk meramalkan TPT dengan akurasi yang baik, baik pada tingkat nasional maupun regional di Provinsi Jawa Barat. Namun, penelitian ini juga mencatat bahwa penggunaan data historis dari <i>Google Trends</i> harus dilakukan dengan hati-hati, terutama dalam situasi ekstrim seperti pandemi COVID-19.</p>	<p><i>official statistics</i> dalam meramalkan Tingkat Pengangguran Terbuka di Indonesia dan Jawa Barat. Hasil penelitian menunjukkan bahwa data <i>Google Trends</i> dapat digunakan untuk meramalkan TPT dengan akurasi yang baik, baik pada tingkat nasional maupun regional di Provinsi Jawa Barat. Namun, penelitian ini juga mencatat bahwa penggunaan data historis dari <i>Google Trends</i> harus</p>



Penelitian	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
					dilakukan dengan hati-hati, terutama dalam situasi ekstrim seperti pandemi COVID-19.
Peramalan Banyaknya Pengunjung Pantai Glagah Menggunakan Metode Autoregressive Integrated Moving Average Exogenous ( <i>ARIMAX</i> ) dengan Efek Variasi Kalender oleh Solikhah Novita Intan , Etik Zukhronah dan Supriyadi Wibowo	Penelitian ini membandingkan beberapa model <i>ARIMAX</i> untuk meramalkan jumlah pengunjung Pantai Glagah. Hasilnya menunjukkan bahwa model <i>ARIMAX</i> ([24],0,0) memberikan hasil yang paling baik dalam meramalkan jumlah pengunjung.	Penelitian ini mengontraskan antara data pengunjung Pantai Glagah pada bulan Idul Fitri yang mengalami peningkatan signifikan dengan data pengunjung pada bulan-bulan lainnya. Hal ini menunjukkan adanya efek variasi kalender Idul Fitri. Oleh karena itu, penelitian ini menggunakan metode <i>ARIMAX</i> yang dapat mengakomodasi efek	Penelitian ini tidak memberikan penjelasan yang rinci mengenai proses pemilihan model <i>ARIMAX</i> yang terbaik. Selain itu, tidak dijelaskan juga mengenai evaluasi dan validasi model yang digunakan dalam penelitian ini. Informasi-informasi tersebut akan membantu dalam memahami kehandalan hasil penelitian.	Penelitian ini menggunakan metode <i>ARIMAX</i> untuk meramalkan jumlah pengunjung Pantai Glagah, khususnya pada bulan Idul Fitri yang mengalami peningkatan signifikan. Hasil penelitian menunjukkan bahwa model <i>ARIMAX</i> ([24],0,0) memberikan hasil yang paling baik dalam meramalkan jumlah pengunjung. Namun,	Penelitian ini menggunakan metode <i>ARIMAX</i> untuk meramalkan jumlah pengunjung Pantai Glagah, terutama pada bulan Idul Fitri yang mengalami peningkatan signifikan. Hasilnya menunjukkan bahwa model <i>ARIMAX</i> ([24],0,0) memberikan hasil yang paling baik

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
		variasi kalender dalam analisis data deret waktu.		penelitian ini perlu memberikan penjelasan lebih rinci mengenai proses pemilihan model dan evaluasi yang dilakukan.	dalam meramalkan jumlah pengunjung. Namun, penelitian ini perlu memberikan penjelasan yang lebih rinci mengenai proses pemilihan model dan evaluasi yang dilakukan.
Peramalan Indeks Harga Saham Gabungan (Ihsg) Menggunakan <i>ARIMAX</i> Dengan Variabel Eksogen Covid-19 oleh Cita Meliana , Rochdi Wasono , M. Al Haris , Zamni Haquel Alfiyani dan Eka Yuni Kartika Sari	Penelitian ini membandingkan pergerakan Indeks Harga Saham Gabungan (IHSG) sebelum dan setelah pandemi Covid-19 di Indonesia. Data IHSG sebelum pandemi menunjukkan nilai yang lebih tinggi	Penelitian ini mengontraskan perbedaan nilai IHSG sebelum dan setelah pandemi Covid-19. Sebelum pandemi, IHSG memiliki nilai yang relatif stabil, sedangkan setelah pandemi, IHSG mengalami penurunan yang signifikan dan tidak stabil. Hal ini	Penelitian ini tidak memberikan penjelasan yang cukup detail mengenai metode yang digunakan dalam pemodelan <i>ARIMAX</i> dan evaluasi kinerja model. Informasi tersebut akan membantu pembaca dalam memahami kehandalan hasil	Penelitian ini menganalisis pergerakan Indeks Harga Saham Gabungan (IHSG) sebelum dan setelah pandemi Covid-19 di Indonesia. Hasilnya menunjukkan adanya penurunan signifikan dan ketidakstabilan IHSG setelah pandemi. Penelitian ini	Penelitian ini menganalisis dampak pandemi Covid-19 terhadap pergerakan Indeks Harga Saham Gabungan (IHSG) di Indonesia. Hasilnya menunjukkan adanya penurunan signifikan dan ketidakstabilan IHSG setelah pandemi. Penelitian

<b>Penelitian</b>	<b><i>Comparing</i></b>	<b><i>Contrasting</i></b>	<b><i>Criticize</i></b>	<b><i>Synthesize</i></b>	<b><i>Summarize</i></b>
	dibandingkan dengan periode setelah pandemi, menunjukkan adanya dampak negatif dari pandemi terhadap IHSG.	menunjukkan dampak negatif dari pandemi terhadap IHSG.	peramalan yang dihasilkan.	menggunakan metode <i>ARIMAX</i> dengan model deterministik tren covid-19 dan model residual <i>ARIMA</i> (1,1,2) untuk meramalkan IHSG. Namun, penelitian ini perlu memberikan informasi yang lebih rinci mengenai metode dan evaluasi model yang digunakan.	ini menggunakan metode <i>ARIMAX</i> dengan model deterministik tren covid-19 dan model residual <i>ARIMA</i> (1,1,2) untuk meramalkan IHSG. Namun, penelitian ini perlu memberikan penjelasan yang lebih rinci mengenai metode dan evaluasi model yang digunakan.
Peramalan Penjualan Pakaian dengan Autoregressive Integrated Moving Average with Exogeneous Input	Penelitian ini membandingkan tingkat penjualan pakaian pada hari libur, perayaan agama, dan persiapan	Penelitian ini mengontraskan perbedaan antara penjualan pakaian pada hari-hari libur dan perayaan agama dengan	Penelitian ini menggunakan metode <i>ARIMAX</i> untuk meramalkan penjualan pakaian, namun tidak menjelaskan secara rinci	Penelitian ini menganalisis tingkat penjualan pakaian pada hari-hari libur, perayaan agama, dan persiapan tahun baru. Penelitian ini	Penelitian ini menganalisis tingkat penjualan pakaian pada hari libur, perayaan agama, dan persiapan tahun baru.

Penelitian	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
<p>(ARIMAX) (Studi Kasus: Penjualan Pakaian di Toko M~Al Samarinda Tahun 2012 s.d 2016) oleh Azeilla Putri Bulu Laga , Sri Wahyuningsih dan Memi Nor Hayati</p>	<p>tahun baru. Penelitian ini menunjukkan bahwa penjualan pakaian meningkat pada saat-saat tersebut.</p>	<p>hari-hari biasa. Hasilnya menunjukkan adanya peningkatan signifikan pada penjualan pakaian saat ada hari libur atau perayaan agama.</p>	<p>bagaimana variabel eksogen yang digunakan dalam model ARIMAX. Informasi lebih lanjut tentang variabel eksogen yang dipilih dan alasan mengapa variabel tersebut dipilih akan membantu dalam memahami kehandalan hasil peramalan.</p>	<p>menggunakan metode ARIMAX, yang merupakan model ARIMA dengan variabel eksogen. Hasil peramalan menunjukkan bahwa model ARIMAX terbaik untuk penjualan pakaian di Toko M~Al adalah SARIMA (1,0,2)(0,0,2)<sub>6</sub> dengan nilai MAPE terkecil. Penelitian ini juga menyajikan hasil peramalan penjualan pakaian selama 12 periode, dengan kenaikan tertinggi terjadi pada bulan Juni dan penurunan terendah terjadi pada bulan Desember.</p>	<p>Metode ARIMAX digunakan untuk meramalkan penjualan pakaian dengan memasukkan variabel eksogen seperti hari libur dan perayaan agama. Hasil peramalan menunjukkan bahwa model ARIMAX terbaik adalah SARIMA (1,0,2)(0,0,2)<sub>6</sub> dengan MAPE terendah. Penelitian ini juga menyajikan hasil peramalan penjualan pakaian selama 12 periode, dengan peningkatan</p>

Penelitian	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
					tertinggi terjadi pada bulan Juni dan penurunan terendah terjadi pada bulan Desember.
<p>Prediksi Jumlah Kunjungan Wisatawan Di Jawa Barat Dengan Model <i>ARIMAX</i> Dan <i>SARIMAX</i> Menggunakan Data <i>Google Trends</i> oleh Ilham Rizki Julianto, Indwiarti dan Aniq Atiqi Rohmawati</p>	<p>Penelitian ini membandingkan penggunaan metode peramalan <i>ARIMAX</i> dan <i>SARIMAX</i> untuk meramalkan kunjungan wisatawan mancanegara di Provinsi Jawa Barat. Hasilnya menunjukkan bahwa kedua model memberikan penurunan nilai MAPE yang relatif</p>	<p>Penelitian ini mengontraskan perbedaan antara model <i>ARIMAX</i> dan <i>SARIMAX</i> dalam meramalkan kunjungan wisatawan mancanegara. Meskipun keduanya memberikan hasil yang cukup baik dengan penurunan MAPE, model <i>SARIMAX</i> memiliki kinerja yang sedikit lebih baik dengan MAPE terendah.</p>	<p>Penelitian ini tidak memberikan informasi yang cukup tentang data dari <i>Google Trends</i> yang digunakan sebagai variabel eksternal. Penjelasan lebih lanjut tentang jenis kueri pencarian yang digunakan dan bagaimana data tersebut berkontribusi dalam peramalan akan membantu memahami</p>	<p>Penelitian ini mengevaluasi penggunaan metode peramalan <i>ARIMAX</i> dan <i>SARIMAX</i> dalam meramalkan kunjungan wisatawan mancanegara di Provinsi Jawa Barat. Hasil peramalan menunjukkan bahwa kedua model memberikan penurunan MAPE yang relatif kecil, namun model <i>SARIMAX</i> memiliki kinerja yang</p>	<p>Penelitian ini mengevaluasi penggunaan metode peramalan <i>ARIMAX</i> dan <i>SARIMAX</i> untuk meramalkan kunjungan wisatawan mancanegara di Provinsi Jawa Barat. Hasilnya menunjukkan bahwa model <i>SARIMAX</i> (2, 1, 2) (2, 1, 2)<sup>12</sup> dengan kueri pencarian "west java"</p>

Penelitian	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
	kecil, tetapi model <i>SARIMAX</i> memberikan MAPE terendah.		keandalan hasil peramalan.	sedikit lebih baik dengan MAPE terendah. Namun, penelitian ini juga mencatat bahwa penambahan kueri pencarian dari Google <i>Trends</i> memberikan hasil yang berbeda antara model <i>ARIMAX</i> dan <i>SARIMAX</i> .	memiliki MAPE terendah sebesar 16,61%. Meskipun penurunan akurasi pada MAPE tidak signifikan antara model <i>ARIMAX</i> dan <i>SARIMAX</i> , penelitian ini menyarankan bahwa pemilihan model dapat menghasilkan hasil yang berbeda.

## 2.2. Dasar Teori

### 2.2.1 Industri *Real estate*

Industri *Real estate* adalah sektor bisnis yang berkaitan dengan pembelian, penjualan, dan pengelolaan properti seperti rumah, apartemen, tanah, dan bangunan komersial. Dalam industri ini, pemasaran online dan strategi *SEO* menjadi sangat penting untuk meningkatkan visibilitas dan menarik calon pembeli atau penyewa properti. Untuk mencapai tujuan tersebut, analisa kata kunci yang tepat untuk menentukan target *keyword SEO* dibutuhkan, dengan menggunakan metode regresi linear dan *ARIMA* dalam memprediksi *trend keyword SEO* dapat memberikan rekomendasi yang berharga bagi tim pemasaran di industri *real estate* [10].

#### 2.2.1.1 Rekomendasi *Keyword SEO*

Dengan menggunakan metode regresi linear dan *ARIMA*, tim pemasaran di industri *real estate* dapat menganalisis data historis dan *trend keyword SEO* yang relevan dengan properti yang mereka jual [16]. Dengan pemodelan ini, tim dapat mengidentifikasi *keyword* yang memiliki korelasi positif dengan permintaan properti dan memprediksi tren *keyword* yang akan populer di masa depan. Rekomendasi *keyword* yang didasarkan pada analisis regresi linear dan *ARIMA* ini akan membantu tim pemasaran dalam mengoptimalkan kampanye *SEO* mereka, memperluas jangkauan target audiens, dan meningkatkan lalu lintas organik ke situs web mereka.

Dalam industri *real estate*, pemasaran online dan strategi *SEO* sangat penting untuk meningkatkan visibilitas dan menarik calon pembeli atau penyewa properti. Dengan menggunakan metode regresi linear dan *ARIMA*, tim pemasaran dapat menganalisis data *keyword SEO*, mengidentifikasi tren permintaan, dan memberikan rekomendasi *keyword* yang akan tren di masa depan. Dengan landasan teori yang kuat tentang industri *real estate*, tim dapat mengoptimalkan kampanye *SEO* mereka dan mencapai hasil yang lebih baik dalam memasarkan properti secara online

## 2.2.2 *Search Engine Optimization (SEO)*

*Search Engine Optimization (SEO)* adalah kumpulan praktik dan strategi yang digunakan untuk meningkatkan visibilitas dan peringkat suatu situs web pada mesin pencari seperti Google, Bing, dan lainnya. Tujuan utama *SEO* adalah untuk mendapatkan lalu lintas organik yang tinggi, yaitu pengunjung yang datang ke situs web melalui hasil pencarian alami, tanpa membayar iklan.

*SEO* melibatkan serangkaian tindakan yang dilakukan untuk mengoptimalkan situs web agar lebih mudah ditemukan oleh mesin pencari dan meningkatkan peringkatnya dalam hasil pencarian. Hal ini dilakukan dengan memahami bagaimana mesin pencari bekerja dan mengoptimalkan faktor-faktor yang memengaruhi peringkat.

### 2.2.4.1 *Algoritma Mesin Pencari*

Mesin pencari menggunakan algoritma kompleks untuk menentukan peringkat situs web dalam hasil pencarian. Beberapa algoritma terkenal yang digunakan oleh mesin pencari adalah

- a. Google menggunakan berbagai algoritma, seperti Google PageRank, Google Panda, Google Penguin, dan Google Hummingbird, untuk menentukan peringkat situs web.
- b. Bing juga memiliki algoritma yang unik untuk menentukan peringkat situs web dalam hasil pencarian mereka.
- c. RankBrain adalah elemen algoritma Google yang menggunakan kecerdasan buatan (AI) untuk memahami maksud di balik pencarian pengguna dan memberikan hasil yang relevan.

## 2.2.3 *EDA (Exploratory Data Analysis)*

Analisis Data Eksplorasi atau (EDA) adalah memahami kumpulan data dengan merangkum karakteristik utamanya, pola dan hubungan didalam dataset, sering kali memplotnya secara visual. Langkah ini sangat penting terutama ketika kita sampai pada pemodelan data untuk menerapkan *Machine Learning*. Plotting dalam EDA terdiri dari *Histogram*, *Box plot*,



*Scatter plot* dan masih banyak lagi. Seringkali dibutuhkan banyak waktu untuk mengeksplorasi data. Melalui proses EDA, kita dapat meminta untuk mendefinisikan pernyataan masalah atau definisi pada kumpulan data kita yang sangat penting [17].

Konsep EDA sendiri dibagi menjadi 4 bagian diantaranya adalah:

- a. Memahami struktur data: Mengetahui jenis data yang ada dalam dataset seperti numerik, kategorikal, tanggal, dll. Ini membantu Anda dalam memilih teknik analisis yang tepat untuk setiap variabel.
- b. Statistik deskriptif: Menghitung statistik deskriptif seperti mean, median, modus, kuartil, deviasi standar, dan lain-lain untuk memahami distribusi dan variabilitas data.
- c. Visualisasi data: Menggunakan grafik dan plot untuk menerangkan pola, hubungan, dan tren dalam dataset.
- d. Pembersihan data: Mengidentifikasi dan menangani *missing values*, *outliers*, dan data yang tidak konsisten atau tidak valid.

#### 2.2.4 Regresi Linear

Regresi linear merupakan metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen (*search volume keyword*) dengan satu atau lebih variabel independen (waktu). Dalam konteks prediksi *trend keyword SEO*, regresi linear dapat membantu dalam mengidentifikasi pola dan tren dari data *time series* yang dikumpulkan dari sumber seperti mesin pencari atau alat analisis *SEO*. Dengan menggunakan regresi linear, dapat mengestimasi nilai atau perilaku masa depan *keyword SEO* berdasarkan tren historis yang ada.

Dalam regresi linear, variabel dependen (*search volume keyword*) ditentukan oleh variabel independen berupa waktu. Tujuan utama regresi linear adalah untuk menemukan persamaan garis terbaik yang mewakili hubungan antara waktu dan tren *keyword SEO*. Persamaan ini kemudian dapat digunakan untuk memperkirakan nilai *search volume keyword* di masa depan dan juga arah trennya.

Metode regresi linear banyak digunakan dalam peramalan data *time series* karena memiliki interpretasi yang intuitif dan mudah dipahami. Koefisien regresi dapat diinterpretasikan sebagai perubahan rata-rata dalam nilai *keyword SEO* untuk setiap satuan waktu yang berlalu. Dengan banyaknya kelebihan yang dimiliki regresi linear seperti kemampuan untuk menangkap tren dan pola, peramalan yang relatif akurat, dapat membantu analisis kausalitas [18].

Rumus regresi linear untuk data *time series* menggambarkan hubungan linier antara variabel dependen (y) dengan satu atau lebih variabel independen (x) dalam konteks data yang terurut secara waktu. Berikut adalah rumus regresi linear:

$$y = a + bX \quad (2.1)$$

Dimana:

y = Variabel terikat (*dependent*) atau variabel *search volume*.

x = Variabel bebas (*independent*) atau variabel kolom minggu.

a = Variabel Konstanta.

b = Variabel Koefisien

dimana nilai a dan b sebagai berikut:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n \sum x^2 - (\sum x)^2} \quad (2.2)$$

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \quad (2.3)$$

Rumus regresi linear ini digunakan untuk memodelkan hubungan antara variabel dependen dan independen dalam konteks data yang terorganisir secara waktu. Dengan menggunakan rumus ini, kita dapat melakukan analisis, prediksi, dan evaluasi tren atau perubahan yang terjadi dalam variabel dependen berdasarkan variabel independen dalam rangka menggambarkan pola atau hubungan linier di dalam data *time series* tersebut.

### 2.2.5 Autoregressive Integrated Moving Average (ARIMA)

Metode *ARIMA* adalah salah satu metode yang digunakan dalam analisis dan peramalan *data time series*. memperkenalkan variabel eksternal sebagai faktor tambahan dalam analisis dan prediksi *data time series*. Dalam konteks prediksi *trend keyword SEO*, variabel eksternal dapat berupa data eksternal yang mempengaruhi perubahan *keyword SEO*, seperti kampanye iklan, peristiwa khusus, atau faktor ekonomi [19]. Dalam konteks peramalan *data time series*, metode *ARIMA* sangat berguna untuk mengidentifikasi pola, tren, dan sifat dependensi dalam data. Dengan menganalisis data historis dan menggunakan kombinasi komponen *AR*, *I*, dan *MA*. Metode *ARIMA* dapat memberikan prediksi yang akurat untuk nilai-nilai masa depan dalam deret waktu.

Metode *ARIMA* paling banyak digunakan dalam peramalan *data time series* karena memiliki kemampuan untuk mengidentifikasi pola dan tren dalam data, menangani non-stasioneritas, menyesuaikan model sesuai dengan karakteristik data, memanfaatkan data historis, serta sederhana dan mudah diimplementasikan. Metode ini memberikan hasil peramalan yang akurat dan dapat diandalkan dalam berbagai konteks, membuatnya menjadi pilihan utama dalam meramalkan data *time series* [18].

Metode *ARIMA* pertama kali dikembangkan pada tahun 1970-an oleh George Box dan Gwilym Jenkins. Mereka memperkenalkan model *ARIMA* sebagai salah satu pendekatan yang efektif dalam analisis dan peramalan *data time series*. Sejak diperkenalkannya, metode *ARIMA* telah menjadi salah satu metode yang paling banyak digunakan dalam bidang peramalan dan analisis *data time series* [20].

Rumus untuk model *ARIMA* yang dikenalkan oleh George Box dan Gwilym Jenkins dibagi menjadi 3 bagian yaitu: *Autoregressive (AR)*, *Moving Average (MA)*, dan model gabungan dari model *AR* dan *MA* yaitu *ARMA (Autoregressive Moving Average)*

#### 1. Model *Autoregressive (AR)*

Model  $AR(p)$  adalah model dimana  $Z_t$  merupakan fungsi dari data diwaktu yang lalu yaitu  $t-1, t-2, t-3, \dots, t-p$ . Rumus umum dari  $AR$  adalah:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t$$

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = a_t$$

Maka dapat ditulis dalam bentuk berikut:

$$(1 - \phi_1 B - \dots - \phi_p B^p) = a_t \quad (2.4)$$

Dimana  $BZ_t = Z_{t-1}$

Atau

$$\phi_1 (B)Z_t = a_t$$

Dengan  $\phi_1 (B)Z_t = (1 - \phi_1 B - \dots - \phi_p B^p)$

Dimana:

$Z$  = peubah tak bebas

$Z_{t-p}$  =himpunan peubah bebas

$a$  = *Error*

## 2. Model *Moving Average (MA)*

Model  $MA(q)$  adalah salah satu cara untuk menghaluskan data deret waktu yang bertujuan untuk menghilangkan atau meminimalisir dampak dari factor *siklisk*, musiman atau random hingga mendapatkan sebuah *trend*. Dari faktor tersebut bisa membuat data menjadi naik-turun (*frukturasi*) sehingga diperlukan *Moving Average* untuk mengurangi frukturasi tersebut.

Model  $MA$  memprediksi  $Z_t$  sebagai fungsi dari kesalahan prediksi masa lalu (*past forecast error*) dalam memprediksi  $Z_t$ . Bentuk umum model  $MA(q)$  sebagai berikut:

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

$$Z_t = (1 - \theta_1 B^p - \dots - \theta_q B^q) a_t$$

Dimana  $Ba_t = a_{t-1}$

Atau

$$Z_t = \theta_1 (B) a_t \quad (2.5)$$

Dengan  $\theta_1 (B) = (1 - \theta_1 B - \dots - \theta_q B^q)$

### 3. Model Autoregressive Moving Average (ARMA)

Model ARMA merupakan gabungan dari model AR ( $p$ ) dan MA ( $q$ ) dengan rumus umum sebagai berikut:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Sehingga dapat ditulis sebagai berikut:

$$(1 - \phi_1 B - \dots - \phi_p B^p)Z_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (2.6)$$

Atau

$$\phi_1(B)Z_t = \theta_1(B)a_t \quad (2.7)$$

### 4. Model Autoregressive Integrated Moving Average (ARIMA)

Dalam model ARIMA ( $p, d, q$ ) dimana apabila  $d=0$  dan  $q=0$  maka model dinotasikan sebagai AR ( $p$ ). Sedangkan sebaliknya jika  $p=0$  dan  $d=0$  maka model dinotasikan sebagai MA ( $q$ ). Model ARIMA dapat digunakan pada data yang stationer maupun tidak stationer. Jika data deret waktu tidak stationer maka perlu dilakukan proses *diffencing* sebanyak  $d$  kal, sehingga data menjadi stationer. Berikut rumus umum untuk model ARIMA sebagai berikut:

$$\phi_1(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (2.8)$$

Dengan

$$\phi_1(B) = (1 - \phi_1 B - \dots - \phi_p B^p), AR(p)$$

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q), MA(q)$$

Dimana:

$$(1 - B)^d = \text{differencing orde } d$$

$$a_t = \text{nilai residual pada } t$$

Dalam model ARIMA terdapat uji stationer melalui perhitungan ACF (*Autocorrelation Function*), PCAF (*Partial Autocorrelation Function*) dan ADF (*Augmented Dickey Fuller*). Berikut adalah penjelasannya

#### 1. ACF dan PCAF

*Autocorrelation Function* (ACF) merupakan suatu hubungan linier antara pengamatan  $Z_t$  dengan pengamatan  $Z_{t+k}$  pada data yang dipisahkan oleh waktu ke-  $k$ . Berikut merupakan fungsi autokorelasi untuk sampel data yang diambil.

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \quad (2.9)$$

Untuk  $k = 1, 2, 3, \dots, n$  dimana  $\bar{Z} = \sum_{t=1}^n \frac{Z_t}{n}$

Dengan:

$Z_t$  = nilai aktual pada waktu ke- $t$

$\hat{\rho}_k$  = nilai estimasi fungsi autokorelasi lag ke- $k$

*Partial Autocorrelation Function* (PACF) digunakan untuk mengukur tingkat keeratan hubungan antara pasangan data  $Z_t$  dan  $Z_{t+k}$  setelah pengaruh variabel  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$  dihilangkan. Perhitungan nilai PACF sampel lag ke- $k$  dimulai dengan menghitung  $\hat{\phi}_{1,1} = \hat{\rho}_1$ , sedangkan untuk menghitung  $\hat{\phi}_{k,k}$  dilakukan dengan menggunakan rumus sebagai berikut:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (2.10)$$

dan  $\hat{\phi}_{k+1,j} = \hat{\phi}_{k,j} - \hat{\phi}_{k+1,k+1} - \hat{\phi}_{k,k+1-j}; j = 1, 2, 3, \dots, k$

ACF dan PACF dapat digunakan untuk mengidentifikasi model dugaan yaitu dengan mengidentifikasi nilai  $p$  dan  $q$  seperti pada Tabel 2.2 berikut:

Tabel 2. 2 Struktur ACF dan PCAF

Model	ACF	PCAF
<b>AR (p)</b>	<i>Dies Down</i> (menurun mengikuti bentuk eksponensial atau gelombang sinus)	Cut off (terpotong) setelah lag ke- $p$

<b>MA (q)</b>	Cut off (terpotong) setelah lag ke- $q$ .	Dies Down (menurun mengikuti bentuk eksponensial atau gelombang sinus)
<b>ARMA (p, q)</b>	<i>Dies Down</i> (menurun mengikuti bentuk eksponensial atau gelombang sinus)	Dies Down (menurun mengikuti bentuk eksponensial atau gelombang sinus)
<b>AR (p) atau MA (q)</b>	Cut off (terpotong) setelah lag ke- $q$	Cut off (terpotong) setelah lag ke- $p$

## 2. ADF

Uji ADF digunakan untuk menguji kestasioneran data. Uji ADF dilakukan menggunakan suatu model  $y_t$  yang diasumsikan mengikuti proses AR orde ke- $p$ . Model  $y_t$  ditulis sebagai berikut:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.11)$$

Keterangan:

- $y_t$  : amatan pada periode ke- $t$ .
- $y_{t-p}$  : amatan pada periode ke- $(t-p)$ .
- $\phi_p$  : koefisien model AR pada waktu ke- $p$ .
- $t$  : 1, 2, 3, ...,  $T$
- $\varepsilon_t$  : galat pada waktu ke- $t$ .

### 2.2.6 Data Time series

*Data time series* (deret waktu) adalah sekumpulan data yang dikumpulkan secara berurutan dalam interval waktu yang terdefinisi. Dalam konteks analisis statistik, data *time series* digunakan untuk memodelkan dan memprediksi perilaku suatu variabel dalam periode waktu yang berkesinambungan. Data ini mencerminkan perubahan dalam variabel tersebut dari waktu ke waktu, memungkinkan kita untuk mengidentifikasi pola, tren, dan faktor-faktor yang mempengaruhi [21].

Pada penelitian ini ada 2 variabel utama dari *data time series* yang harus diperhatikan agar model yang digunakan untuk melakukan prediksi *trend keyword SEO* dapat bekerja dengan baik. Berikut adalah penjelasan 2 variabel utama dalam *data time series*:

### **2.2.3.1 Variabel waktu (*time variable*) atau Variabel kolom Minggu**

Variabel waktu adalah sumbu horizontal dalam data *time series* yang mewakili periode waktu dari observasi. Ini bisa berupa tahun, bulan, minggu, hari, jam, atau interval waktu lainnya, tergantung pada sifat data dan tujuan analisis. Variabel waktu memungkinkan kita untuk mengurutkan observasi dalam urutan temporal yang benar, sehingga memahami pola dan perubahan dari waktu ke waktu [1].

### **2.2.3.2 Variabel dependen (*dependent variable*) atau Variabel Search Keyword**

Variabel dependen (juga dikenal sebagai variabel respons) adalah variabel yang kita ingin analisis atau prediksi dalam *data time series*. Ini adalah variabel yang berubah seiring waktu dan menjadi fokus utama analisis. Dalam konteks *data time series* untuk prediksi *trend keyword SEO*, variabel dependen dapat berupa tingkat permintaan bulanan, jumlah klik, atau peringkat *keyword* pada mesin pencari [1].

## **2.2.7 Uji Akurasi**

Ketepatan peramalan dalam menganalisis data sangat diperlukan untuk mengukur kesesuaian metode prediksi yang digunakan. Kriteria yang digunakan untuk menguji ketepatan prediksi dalam penelitian ini, menggunakan dua metode uji akurasi, yaitu *mean absolute error* (MAE) *root mean square error* (RMSE) dan *mean absolute percentage error* (MAPE), untuk mengevaluasi performa model dalam memprediksi tren *keyword SEO* menggunakan metode regresi linear dan *ARIMA*.

*RMSE* merupakan metode yang digunakan untuk mengukur tingkat kesalahan rata-rata dari prediksi model. Rumus *RMSE* adalah sebagai berikut:



$$RMSE = \sqrt{\frac{\sum_{t=1}^n (At - Ft)^2}{n}} \quad (2.12)$$

Dimana:

$At$  = Nilai data aktual

$Ft$  = Hasil Nilai Peramalan

$n$  = Banyaknya data

interpretasi  $RMSE$  yaitu menghasilkan angka positif dan semakin kecil nilainya, semakin baik performa prediksi model.  $RMSE$  memberikan informasi tentang tingkat deviasi antara nilai prediksi dan nilai actual.[25].

$MAPE$  digunakan untuk mengukur persentase kesalahan rata-rata relatif dari prediksi model. Rumus  $MAPE$  adalah sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Ai - Fi}{Ai} \right| \times 100\% \quad (2.13)$$

Dimana:

$n$  = Ukuran Sampel

$Ai$  = Nilai data Aktual

$Fi$  = Nilai data Peramalan

Interpretasi nilai  $MAPE$  sebesar 10% dapat diartikan selisih rata-rata nilai data peramalan dengan nilai sebenarnya adalah 10%. Semakin kecil nilai  $MAPE$  maka semakin akurat sebuah model dalam melakukan peramalan. Untuk hasil  $MAPE$  lebih dari 10% dapat dikegerikan menjadi beberapa interpretasi. Berikut adalah Tabel Interpretasi  $MAPE$  berdasarkan nilai intervalnya. Dapat dilihat pada Tabel 2.2 berikut untuk nilai Interpretasi  $MAPE$ .

Tabel 2. 3 Interpretasi MAPE

Nilai $MAPE$	Interpretasi
$\leq 10$	Hasil peramalan sangat akurat
10 - 20	Hasil peramalan baik
20 – 50	Hasil peramalan layak (Cukup baik)

> 50	Hasil peramalan tidak baik
------	----------------------------

*MAPE* menghasilkan angka dalam bentuk persentase dan semakin kecil nilainya, semakin baik performa prediksi model. *MAPE* memberikan informasi tentang tingkat kesalahan relatif antara nilai prediksi dan nilai aktual.[25].

MAE adalah rata-rata selisih mutlak nilai sebenarnya (aktual) dengan nilai prediksi (peramalan). MAE digunakan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi atau peramalan. Rumus MAE adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i| \quad (2.14)$$

Dimana:

$n$  = nilai ukuran sampel.

$A_i$  = nilai data aktual ke- $i$ .

$F_i$  = nilai data peramalan ke- $i$ .

Karena pada rumus MAE di atas terdapat tanda mutlak ( $| \ |$ ), maka nilai MAE akan selalu bernilai positif. Interpretasi MAE digunakan dalam membandingkan dua model dalam melakukan peramalan, semakin kecil nilai MAE, semakin baik model tersebut dalam melakukan peramalan. Oleh karena itu, dalam melakukan perbandingan dua atau beberapa model statistik, maka model terbaik adalah model yang memiliki nilai MAE terkecil[22].