

Prediksi Penjualan *Marketplace* Menggunakan Regresi Linear Berganda Berdasarkan Produk Terjual dan Potongan Biaya Pengiriman (Studi Kasus : *Aufar Grosir Jeans*)

Melanie Putri^{a1}, Hannie^{a2}, Siska^{a3}

^aProgram Studi Sistem Informasi
Universitas Singaperbangsa Karawang
Puseurjaya, Telukjambe Timur, Karawang, Indonesia
¹melanieputri4501@gmail.com
²Hannie@staff.unsika.ac.id
²Siska@staff.unsika.ac.id

Abstrak

Aufar Grosir Jeans merupakan satu dari banyaknya *marketplace* yang saat ini tengah memasarkan produknya di *Shopee*, mereka menjual berbagai macam pakaian berbahan jeans. Saat ini *Aufar Grosir Jeans* terus berusaha untuk selalu meningkatkan penjualan mereka dengan memanfaatkan media promosi yang ada. Akan tetapi, dengan ketatnya persaingan pasar saat ini, *Aufar Grosir Jeans* tengah mengalami ketidakstabilan penjualan. Oleh karena itu, diperlukan cara agar dapat melihat apakah promosi yang telah mereka lakukan berpengaruh secara signifikan terhadap penjualan. Pada penelitian ini akan dilakukan prediksi penjualan dengan menggunakan metode Regresi Linear Berganda dengan variabel X_1 adalah produk terjual, X_2 adalah estimasi potongan biaya pengiriman dan Y Total pembayaran (penjualan). Tools yang akan digunakan adalah SPSS 26 dan Python. Kemudian untuk membuat alur penelitian efektif akan digunakan Metodologi SEMMA dalam penelitian ini. Berdasarkan Perhitungan Algoritma Regresi Linear Berganda dengan SPSS didapatkan nilai sig uji f adalah 0,000 dan $F_{hitung} = 2832,640 > F_{tabel} = 3,013$ yang menunjukkan X_1 , dan X_2 berpengaruh secara simultan terhadap Y . Serta nilai sig uji T adalah 0,000 dan $t_{hitung} = -54.271 > t_{tabel} = -1964$ yang menunjukkan bahwa variabel X_1 , dan X_2 bersama-sama mempengaruhi Y . Kemudian didapatkan juga R^2 sebesar 91,9% dan MAPE sebesar 8,99%, yang menandakan model dapat diandalkan untuk mendukung pengambilan keputusan dalam meningkatkan penjualan di *marketplace* *Aufar Grosir Jeans*.

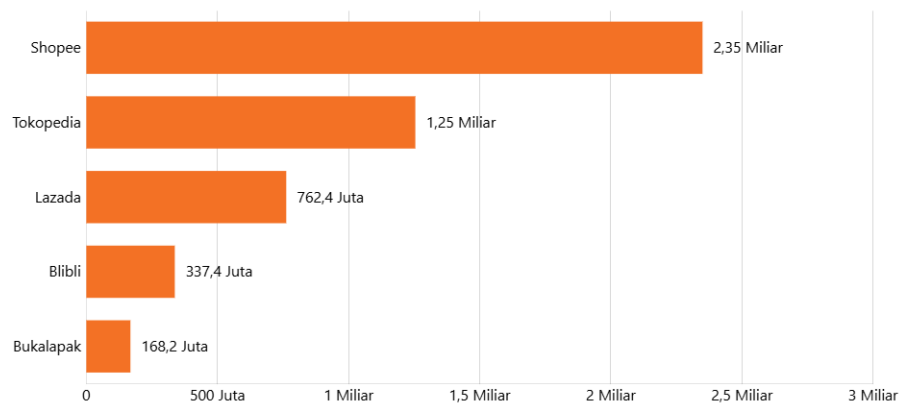
Keywords: Regresi Linear Berganda, SEMMA, Prediksi, Penjualan, *Marketplace*

1. Pendahuluan

Di era digital saat ini, kemajuan teknologi telah mengubah cara kita dalam melakukan sesuatu, salah satunya adalah mengubah cara kita dalam melakukan proses jual beli. Lalu munculah kebiasaan berbelanja online dengan berbagai kemudahannya. Terutama setelah hadir berbagai platform *Marketplace* seperti *Shopee*, *Tokopedia*, dan *Lazada* yang mempermudah akses berbagai produk dan layanan. Kini *Marketplace* telah menjadi sarana utama bagi bisnis untuk memperluas jangkauan pasar, meningkatkan penjualan dan membangun loyalitas dengan pelanggan secara lebih efektif.

Marketplace merupakan platform jual beli digital yang menghubungkan penjual dan pembeli secara online. Penjual dapat menawarkan produk atau pun jasa mereka, sedangkan pembeli dapat memilih, membandingkan, dan membeli barang dari berbagai penjual di dalam satu platform yang sama. Dengan berbagai kemudahannya, *Marketplace* telah membantu penjual dan pembeli dalam melakukan jual beli dengan lebih efektif dan efisien. Penjual tidak perlu memiliki toko fisik, dan pembeli dapat berbelanja darimana saja dan kapan saja. *Marketplace* biasanya juga menyediakan layanan pembayaran, pengiriman, dan sistem penilaian untuk memudahkan transaksi.

Prediksi Penjualan *Marketplace* Menggunakan Regresi Linear Berganda Berdasarkan Produk Terjual dan Potongan Biaya Pengiriman (Studi Kasus : AUFAR Grosir Jeans)



Gambar 1. Grafik pengunjung *e-commerce* per tahun 2023

(Sumber : (databoks.katadata.co.id))

Terlihat dari gambar [1] bahwa platform-platform *Marketplace* di Indonesia telah mencapai jutaan hingga milyaran pengunjung. Shopee menjadi platform dengan pengunjung paling banyak per tahun 2023 dengan 2,35 miliar pengunjung. Dengan banyaknya pengunjung dan akses yang lebih mudah menjangkau pembeli, tentu saja membuat para penjual berlomba-lomba untuk dapat menjual produknya di Shopee. Salah satu penjual atau toko yang masih aktif menjual produknya di Shopee adalah AUFAR Grosir Jeans.

AUFAR Grosir Jeans telah bergabung menjadi seller di Shopee sejak 16 bulan lalu dan menjual produk berupa atasan dan bawahan berbahan jeans. Dengan produk terupload 142 produk, sebanyak 7,7 ribu pengikut, dan 13,9 ribu ulasan pembeli, AUFAR Grosir Jeans terus berusaha mengembangkan tokonya. Akan tetapi, dengan ketatnya persaingan penjual di *Marketplace* khususnya Shopee, AUFAR Grosir Jeans mengalami ketidakstabilan penjualan. Pada penelitian ini, penulis akan menerapkan algoritma regresi linear berganda terhadap data penjualan AUFAR Grosir Jeans selama 5 bulan terakhir sejak penelitian ini dibuat. Metodologi yang akan diterapkan adalah *SEMMA* (*Sample, Explore, Modify, Model, Assess*) dan *tools* yang digunakan adalah SPSS dan Python.

Menurut Triyanto (2019), regresi linier berganda merupakan teknik analisis prediktif yang dipakai guna menggambarkan keterkaitan antara variabel terikat (variabel respons) bersama sejumlah variabel bebas (variabel penentu) yang berpengaruh terhadapnya [2]

Penelitian sebelumnya telah dilakukan oleh Badzlin 2024, ia meneliti mengenai pengimplementasian data mining untuk memprediksi penjualan produk semen di PT Moyo Mortar Indonesia dengan menggunakan regresi linear, dapat disimpulkan bahwa hasil uji melalui penerapan regresi linear, didapatkan nilai RMSE sebesar 493.125,701, Absolute Error sebesar 303.270,965, dan Relative Error mencapai 1,89%. Nilai-nilai ini menunjukkan performa model yang optimal dan prediksi yang cukup tepat [3]

Selanjutnya pada pengkajian yang dilakukan Ferdinan 2024, yang memprediksi banyaknya penjualan tahun 2024 dengan menggunakan metode regresi linear, didapatkan hasil prediksi penjualan di tahun 2024 pada toko dinan_s akan mengalami kenaikan dari tahun sebelumnya yaitu sebanyak 11614 pesanan. Hasil ini diharapkan dapat menjadi acuan bagi toko dinan_s untuk merencanakan strategi penjualan dan inventaris yang lebih efektif di tahun 2024 [4]

Kajian lainnya yang menggunakan teknik regresi linear berganda telah dikerjakan juga oleh Al-Fadhilah Nur Wahyudin 2020 yang mengestimasi penjualan mobil Astra Isuzu, dapat disimpulkan bahwa Algoritma regresi linear berganda mampu memperkirakan penjualan di tahun mendatang. Berdasarkan formula regresi, dapat dihasilkan penjualan PT. Astra Isuzu International Tbk-Isuzu diprediksi hanya mencapai 12.223 unit, yang menunjukkan penurunan sebesar 1.350 unit dibandingkan dengan tahun sebelumnya. Prediksi ini mengindikasikan adanya tren penurunan yang mungkin memengaruhi kinerja perusahaan. Temuan ini dapat menjadi dasar bagi PT. Astra Isuzu untuk mengevaluasi strategi penjualannya ke depan [5]

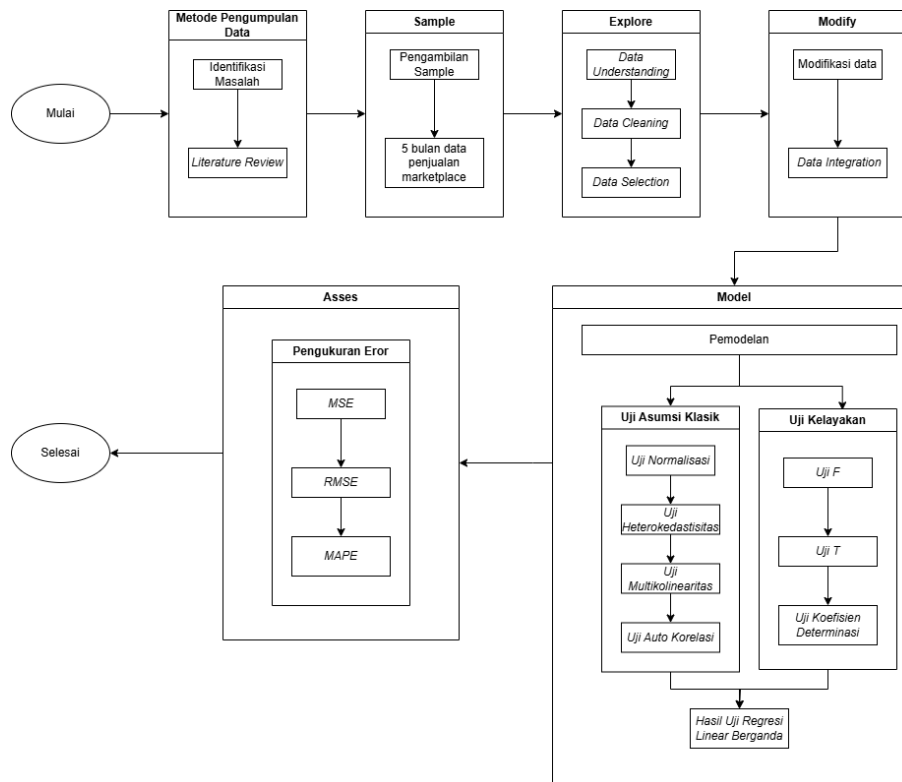
Dari beberapa contoh penelitian terdahulu yang sudah dilakukan, menunjukkan metode regresi linear berganda cukup akurat dalam memprediksi penjualan di masa depan. Oleh karenanya, penulis berencana untuk melakukan penelitian dengan judul “**Prediksi Penjualan *Marketplace* Menggunakan Regresi Linear Berganda Berdasarkan Produk Terjual dan Potongan Biaya Pengiriman (Studi Kasus : AUFAR Grosir Jeans)**”.

2. Metode Penelitian

Pada penelitian ini akan digunakan metodologi SEMMA (*Sample, Explore, Modify, Model, Asses*), yang merupakan pendekatan sistematis dalam data mining. SEMMA dirancang untuk mengoptimalkan proses analisis data dan mengungkap pola-pola tersembunyi dalam data. Metodologi ini dipilih karena fleksibilitasnya dalam menangani dataset besar dan kompleks serta kemampuannya untuk memberikan hasil analisis yang terstruktur dan akurat.

2.1 Rancangan Penelitian

Untuk memprediksi penjualan *Marketplace* menggunakan regresi linear berganda berdasarkan produk terjual dan potongan biaya pengiriman, setiap tahap harus disesuaikan dengan metodologi yang akan digunakan yaitu metodologi SEMMA. Rancangan penelitian yang akan dilakukan akan dijelaskan dalam ilustrasi berikut :



Gambar 2. Rancangan penelitian

Pada gambar 2. merupakan rancangan penelitian yang akan dilakukan. Berikut adalah penjelasan tentang tahapan penelitian tersebut :

a. Metode Pengumpulan Data

Pada tahapan ini, akan dipaparkan cara pengumpulan data yang digunakan dalam penelitian, mulai dari identifikasi masalah, dan *literature review*

1 Identifikasi Masalah

Identifikasi masalah digunakan guna menemukan isu utama yang menjadi dasar penelitian. Dalam konteks penelitian ini, fokusnya adalah memahami hubungan antara produk terjual dan potongan biaya pengiriman terhadap penjualan *marketplace*. Permasalahan yang diidentifikasi meliputi ketidakpastian dalam prediksi penjualan, yaitu banyak faktor yang memengaruhi penjualan di *marketplace*, sehingga sulit bagi pemilik usaha untuk membuat prediksi akurat. Selain itu, pengaruh diskon pengiriman terhadap penjualan digunakan sebagai strategi pemasaran, tetapi dampaknya terhadap penjualan belum sepenuhnya dipahami. Oleh karenanya penelitian ini berfokus dalam menganalisis dan memprediksi penjualan di *marketplace* Shopee, khususnya pada toko Aufar Grosir Jeans, dengan menggunakan pendekatan data mining berbasis metode SEMMA.

2 *Literatur Review*

Kajian literatur ini bertujuan untuk mendapatkan dasar teori yang mendukung penelitian.

Penelitian ini mengacu pada teori penjualan yang menyatakan bahwa penjualan dipengaruhi oleh berbagai faktor, seperti strategi pemasaran, potongan biaya pengiriman, dan jumlah produk yang terjual. Potongan biaya pengiriman dianggap memengaruhi keputusan pembelian karena memberikan insentif tambahan bagi konsumen untuk menyelesaikan transaksi.

Dari sisi metodologi, penelitian ini menggunakan pendekatan SEMMA (Sample, Explore, Modify, Model, Assess) untuk menganalisis data secara sistematis. Metode ini membantu menemukan pola dan hubungan antara variabel independen dan dependen. Algoritma regresi linear berganda digunakan untuk memprediksi total penjualan berdasarkan jumlah produk terjual dan potongan biaya pengiriman. Pendekatan ini dipilih karena mampu menghasilkan analisis yang akurat dan mendalam.

b. *Sample*

Penelitian ini memilih data yang relevan yang sejalan dengan tujuan penelitian. Data yang digunakan adalah data penjualan dari toko Afar Grosir Jeans di *marketplace* Shopee selama lima bulan (Mei–September 2024). Sampel yang dipilih terdiri dari 8 kolom utama, yaitu nomor pesanan, nama produk, harga awal, harga setelah diskon, total harga produk, jumlah produk yang dipesan, estimasi potongan biaya pengiriman, dan total pembayaran.

c. *Explore*

1 *Data Understanding*

Tahap *Data Understanding* bertujuan untuk memperoleh pemahaman awal terhadap data yang akan digunakan dalam penelitian. Proses ini juga mencakup analisis deskriptif untuk mengidentifikasi pola dan tren awal dari variabel-variabel yang ada.

2 *Data Cleaning*

Pada tahapan ini, dilakukan pembersihan data untuk memastikan kualitas data yang akan dianalisis. Proses ini melibatkan penghapusan data yang tidak relevan atau tidak lengkap, pemeriksaan dan penghapusan data duplikat, pengisian nilai yang hilang (*missing values*), serta memastikan konsistensi data yang digunakan.

3 *Data Selection*

Tahapan ini bertujuan untuk memilih data yang relevan dari dataset yang telah dibersihkan dan tersedia. Data yang dipilih hanya mencakup variabel yang sesuai dengan tujuan penelitian, yaitu produk terjual, potongan biaya pengiriman, dan total penjualan. Data yang tidak relevan atau tidak diperlukan akan dihapus untuk memastikan analisis lebih fokus dan efisien.

d. *Modify*

1 Modifikasi Data

Pada bagian modifikasi, data akan disesuaikan agar siap digunakan dalam analisis. Proses ini meliputi penanganan *missing values* yaitu mengisi data yang hilang dengan nilai rata-rata, median, atau metode lain yang sesuai, transformasi data yaitu mengubah format data jika diperlukan, seperti mengonversi nilai kategori menjadi angka atau normalisasi data untuk skala yang seragam, serta penghapusan outlier yaitu mengidentifikasi dan menghapus data yang tidak wajar agar tidak memengaruhi hasil analisis.

2 *Data Integration*

Tahap ini menggabungkan data penjualan *marketplace* dari bulan Mei hingga September 2024. Proses ini memastikan seluruh data dari periode tersebut digabungkan ke dalam satu dataset terpadu. Data dari masing-masing bulan diintegrasikan berdasarkan struktur yang sama, seperti kolom produk terjual, potongan biaya pengiriman, dan total penjualan.

e. *Model*

Tahapan ini merupakan proses utama dalam analisis data, di mana model regresi linear berganda digunakan untuk memprediksi pengaruh variabel independen terhadap variabel dependen. Sebelum membangun model, dilakukan serangkaian uji untuk memastikan bahwa data memenuhi asumsi dasar regresi linear dan model yang digunakan layak untuk analisis lebih lanjut. Untuk memastikan model regresi linear berganda valid dan reliabel, dilakukan serangkaian pengujian berikut:

1 Uji Asumsi Klasik

Uji ini dilakukan guna memastikan model regresi memenuhi kriteria statistik yang valid. Uji

yang dilakukan meliputi:

- Uji Normalitas : Memastikan bahwa data residu dari model regresi terdistribusi normal. Hal ini penting untuk validitas uji statistik dalam model. Uji normalitas akan dilakukan dengan menggunakan uji *Kolmogorov Smirnov*
- Uji Heterokedastisitas : Memeriksa apakah kesalahan dalam model regresi (selisih antara nilai yang diramalkan dan nilai yang sesungguhnya) memiliki variasi yang sama di semua tingkat variabel yang kita gunakan.
- Uji Multikolinearitas : Melihat apakah ada hubungan yang sangat kuat antar variabel independen dalam model. Multikolinearitas dapat menyebabkan hasil estimasi menjadi bias. Uji ini biasanya dilakukan dengan melihat nilai *Variance Inflation Factor (VIF)* dan *Tolerance*.
- Uji Autokorelasi : Memeriksa apakah ada korelasi antara residual dalam model regresi. Autokorelasi dapat membuat model kurang akurat, terutama jika data memiliki urutan waktu. Uji Durbin-Watson akan digunakan untuk mendeteksi masalah ini.

2 Uji Kelayakan

Uji kelayakan dilakukan untuk menilai apakah model regresi yang dihasilkan dapat digunakan untuk memprediksi atau menjelaskan hubungan antarvariabel. Uji ini meliputi:

- Uji F (Simultan) : Melihat apakah semua variabel independen bersama-sama mempengaruhi variabel dependen secara signifikan. Jika nilai signifikan lebih kecil dari α (0,05) maka model dianggap signifikan secara keseluruhan.
- Uji T (Parsial) : Memeriksa adanya pengaruh masing-masing variabel independen terhadap variabel dependen secara individu. Dianggap mempengaruhi apabila nilai signifikan lebih kecil dari α (0,05).
- Uji Aurokorelasi : Mengukur besarnya variasi variabel dependen terhadap variabel independent yang dapat dijelaskan dalam model. Semakin besar nilai R^2 , semakin baik model dalam menjelaskan hubungan antarvariabel.

3. Hasil Uji Regresi Linear Berganda

Uji Regresi linear berganda dalam penelitian ini sangat penting karena memungkinkan untuk menganalisis pengaruh dua atau lebih variabel independen, yakni produk terjual dan potongan biaya pengiriman, terhadap variabel dependen yaitu penjualan *marketplace*. Dengan metode ini, dapat diketahui sejauh mana variabel-variabel tersebut secara simultan maupun parsial memengaruhi penjualan, sehingga hasil penelitian dapat digunakan untuk memberikan prediksi yang akurat dan dukungan dalam pengambilan keputusan.

f. Asses

Tahap Asses bertujuan untuk mengevaluasi performa model regresi linear berganda yang telah dibuat. Evaluasi dilakukan dengan menghitung tingkat kesalahan prediksi menggunakan beberapa metrik, yaitu *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)* dan *Mean Absolute Percentage Error (MAPE)*,. Metrik-metrik ini memberikan gambaran sejauh mana model mampu memprediksi nilai variabel dependen dengan akurasi yang baik.

1. Pengukuran Error

- *Mean Squared Error (MSE)* : *Mean Squared Error (MSE)* digunakan untuk menghitung rata-rata dari kuadrat selisih antara nilai sebenarnya dengan nilai prediksi. Karena menggunakan perhitungan kuadrat, MSE memberikan penekanan lebih besar pada kesalahan yang besar. Nilai MSE yang lebih kecil menunjukkan bahwa model memiliki kinerja prediksi yang lebih baik.
- *Root Mean Squared error (RMSE)* : merupakan akar kuadrat dari nilai MSE, yang menghasilkan metrik kesalahan dalam skala yang sebanding dengan data asli. RMSE digunakan untuk memahami kesalahan prediksi dalam satuan yang sama dengan variabel dependen, sehingga lebih intuitif dibandingkan MSE.
- *Mean Absolute Percentage Error (MAPE)* : MAPE mengukur rata-rata persentase kesalahan antara nilai sebenarnya dan nilai ramalan. Semakin kecil nilai MAPE, semakin baik model dalam memprediksi. MAPE sering digunakan karena memberikan interpretasi dalam bentuk persentase, sehingga mudah dipahami.

2.2 Prediksi

Prediksi atau forecasting adalah kegiatan meramalkan atau memperkirakan situasi di masa depan didasarkan pada informasi atau data dari masa lalu, dengan tujuan mengurangi potensi kesalahan yang mungkin terjadi [6]

2.3 Penjualan

Secara umum, penjualan merupakan aktivitas transaksi yang melibatkan dua pihak atau lebih dengan memanfaatkan media pembayaran yang legal, dengan tujuan utama memperoleh keuntungan dari produk atau jasa yang dibeli. Menurut Soemarso dalam bukunya "Akuntansi Suatu Pengantar" (2009:164), penjualan adalah tindakan dunia usaha untuk menjual produknya, yang dapat dilakukan secara tunai atau secara kredit. Penjualan juga dapat diartikan sebagai kegiatan di mana penjual berusaha memuaskan kebutuhan dan keinginan pembeli, sehingga terjalin hubungan yang saling menguntungkan dan berkelanjutan bagi kedua belah pihak [7]

2.4 Marketplace

Marketplace dapat diartikan sebagai pasar dalam bahasa Inggris. Namun istilah *Marketplace* mempunyai arti yang lebih spesifik yaitu interaksi jual beli antara penjual dan pembeli di dalam tempat yang telah ditetapkan. Dalam konteks ini, yang ditunjukkan bukan pasar fisik, melainkan pasar online. Pasar online merupakan platform untuk memfasilitasi transaksi tanpa perlu menyewa lokasi fisik atau mendirikan booth penjualan, sehingga membuat transaksi menjadi lebih nyaman. *Marketplace* menawarkan fitur-fitur yang dapat meningkatkan kredibilitas toko online. Fitur-fitur tersebut antara lain integrasi pembayaran, escrow account, verifikasi toko, reputasi, dan mekanisme verifikasi yang dapat meningkatkan kepercayaan konsumen. Pada dasarnya, *Marketplace* merupakan ruang di mana para pengusaha dapat menawarkan produk dan jasanya kepada konsumen. Tidak hanya platform seperti shopee, tokopedia, bukalapak, dan lazada, aplikasi seperti Instagram dan TikTok juga berperan besar dalam perdagangan digital. Bisnis yang dikembangkan melalui platform seluler dapat menawarkan berbagai manfaat, antara lain pembentukan saluran pemasaran langsung, kemampuan memperkuat merek, meningkatkan keterlibatan pelanggan, dan membedakan diri dalam kompetensi [8]

2.5 SPSS

SPSS adalah perangkat lunak komputer yang berguna untuk analisis statistik. Pada awalnya, SPSS dirancang untuk ilmu sosial, tetapi seiring waktu, penggunaannya meluas ke berbagai bidang lainnya, sehingga namanya diubah menjadi Statistical Product and Service Solution (Nisfiannoor, Muhammad, 2009)

2.6 Python

Python merupakan bahasa pemrograman yang menggunakan interpreter guna menjalankan kode secara langsung. Dengan interpreter ini, Python dapat dijalankan di berbagai platform, seperti Windows, Linux, dan lainnya. Bahasa python mengambil berbagai paradigma pemrograman, termasuk pemrograman prosedural seperti pada C, pemrograman berorientasi objek seperti Java, serta pemrograman fungsional seperti Lisp. Kombinasi paradigma tersebut membuat Python menjadi fleksibel dan memudahkan pengembang dalam membuat berbagai proyek. Python banyak digunakan oleh programmer dan peneliti karena fleksibilitasnya untuk berbagai keperluan. Python juga mendukung integrasi dengan sistem basis data dan memungkinkan pengolahan serta modifikasi file. Oleh karena itu, Python sering dimanfaatkan untuk prototipe atau pengembangan perangkat lunak dengan cepat dan andal. Selain itu, kemampuan Python dalam menangani data skala besar dan menyelesaikan perhitungan matematika yang kompleks menjadikannya pilihan populer di kalangan peneliti [9]

2.7 Regresi Linear Berganda

Analisis regresi linier berganda, atau yang sering disebut sebagai multiple linear regression, diterapkan guna mengevaluasi seberapa jauh dua atau lebih variabel independen (X_1 dan X_2) mempengaruhi satu variabel dependen (Y) [10]. Ali Anwar berpendapat bahwa regresi linear berganda dikenal juga sebagai regresi ganda dua prediktor, di mana terdapat tiga variabel yang berperan, yaitu dua variabel bebas dan satu variabel terikat [11]

Regresi linier berganda merupakan metode yang digunakan untuk melakukan peramalan dengan mencakup dua atau lebih variabel, yaitu variabel yang mempengaruhi dan variabel yang dipengaruhi. Variabel-variabel ini saling terkait, menciptakan hubungan sebab-akibat. Dalam regresi linier berganda, keterkaitan antara variabel bebas dan variabel terikat direpresentasikan dalam bentuk garis lurus. Melalui analisis regresi linier, kita dapat memperoleh persamaan yang berfungsi untuk memprediksi nilai variabel dependen di waktu yang akan datang dengan menginput nilai variabel independen ke dalam persamaan

tersebut [12]

Analisis regresi linier berganda dapat dikatakan sebagai peningkatan dari analisis regresi linier sederhana, di mana lebih dari satu variabel dependen (x) yang dilibatkan. Metode ini bertujuan mengukur pengaruh beberapa variabel independen terhadap variabel dependen (y), dengan mempertimbangkan nilai dari masing-masing variabel independen tersebut [13]. Berikut merupakan rumus persamaannya :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \tag{1}$$

Keterangan:

- Y : Variabel terikat (*Dependent Variable*)
- a : Konstanta (titik potong Y)
- b : Koefisien dari variabel X(n)
- X : Variabel bebas (*Independent Variable*)

2.8 Pengukuran Error

Pengukuran *error* adalah metode untuk menghitung seberapa jauh perbedaan antara nilai yang diramalkan dan nilai asli dalam sebuah model. Pengukuran ini penting untuk menilai akurasi model prediksi atau peramalan.

3. Hasil dan Pembahasan

3.2 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan adalah metode dokumentasi, dimana data yang digunakan dalam penelitian didapatkan secara langsung dari toko AUFAR Grosir Jeans sebagai objek penelitian. Data yang diminta adalah data 5 bulan penjualan, tepatnya data di bulan Mei 2024 sampai September 2024 sebanyak 9.219 data, yang tersedia dalam arsip transaksi atau laporan penjualan. Teknik Pengumpulan data yang digunakan yaitu hanya dipilih beberapa data atau kolom saja yang relevan dengan kebutuhan analisis. Teknik ini digunakan agar data yang digunakan sesuai dengan tujuan penelitian yaitu mengeksplorasi hubungan antara variabel-variabel yang memengaruhi penjualan.

3.2.1 Identifikasi Masalah

Dalam penelitian ini masalah yang dihadapi adalah penurunan efektivitas strategi penjualan di *marketplace* AUFAR Grosir Jeans. Berdasarkan pengamatan awal, belum ada analisis mendalam mengenai hubungan antara jumlah produk terjual, potongan biaya pengiriman, dan total pembayaran terhadap peningkatan penjualan. Penelitian ini bertujuan untuk memberikan wawasan kepada pemilik toko mengenai pengaruh variabel-variabel tersebut untuk meningkatkan strategi pemasaran dan penjualan.

3.2.2 Literature Review

Penelitian ini didukung oleh kajian literatur yang berkaitan dengan teori penjualan dan strategi diskon biaya pengiriman. Penjualan dapat dorong oleh berbagai faktor, satu diantaranya adalah promosi atau diskon yang memberikan insentif kepada konsumen. Kajian sebelumnya juga menunjukkan bahwa metode analisis seperti regresi linear berganda dapat berguna untuk memprediksi hubungan antara variabel bebas dan terikat. Literature review ini memberikan landasan teoritis dalam memahami hubungan antara produk terjual, potongan biaya pengiriman, dan penjualan pada *marketplace*.

3.3 Sample

Sample yang digunakan dalam penelitian ini adalah data penjualan *marketplace* AUFAR Grosir Jeans di Shoppe yang diambil dari bulan Mei sampai September 2024. Berikut merupakan contoh data penjualan yang akan dijadikan sebagai sample dalam penelitian ini :

No. Pesanan	Nama Produk	Harga Awal	Harga Setelah	Total Harga Produk	Jumlah Produk	Estimasi Potong	Total Pembayaran
2405016A1RN7C2	celana viral jeans	58905	58905	58905	1	12000	57519
2405016UBY7655	CELANA KULOT JE	65800	65800	65800	1	18500	63947
2405029CY1DSM1	CELANA KULOT JE	55800	55800	55800	1	14000	55938
2405028R443CFU	celana anak laki l	50905	50905	50905	1	10000	41905
2405029Q2744U1	CELANA KULOT JE	56800	56800	56800	1	14000	57800
240501737UMBAX	overall anak pere	59865	59865	59865	1	15000	63995
2405029FCG13NA	celana jeans carg	65800	65800	65800	1	15000	63820
2405015P9YQEA0	celana anak laki l	64905	64905	64905	1	15000	68501
2405016XWNEXRN	CELANA JEANS AB	55800	55800	55800	1	18250	56130
2405029NANCJC3	Celana kulot jean	60015	60015	60015	1	18000	63416
2405017RXGTJ77	5-17 tahun cargo	63800	63800	63800	1	17000	55079
24050285RGGBV0	CARGO KULOT JE	58905	58905	58905	3	14000	182610
24050285RGGBV0	CELANA JEANS AB	61800	61800	61800	3	14000	182610
24050285RGGBV0	CELANA JEANS AB	59905	59905	59905	3	14000	182610
2405029JE83TRE	CELANA JEANS AB	55800	55800	55800	2	20000	104988
2405029JE83TRE	CELANA JEANS AB	61800	61800	61800	2	20000	104988
2405029KTEDNBM	CELANA KULOT JE	69905	69905	69905	1	20000	74741

Gambar 3. Data Penjualan AUFAR Grosir *Jeans*

3.3 Explore

Tahapan Explore bertujuan untuk memahami karakteristik dataset yang digunakan, membersihkan data dari masalah kualitas, dan memilih data yang relevan untuk analisis lebih lanjut. Tahapan ini meliputi tiga proses utama, yaitu *Data Understanding*, *Data Cleaning*, dan *Data Selection*.

3.3.1 Data Understanding

Data yang didapat dari toko AUFAR Grosir *Jeans* merupakan data penjualan di tahun 2024 dengan format excel, dimana data tersebut memiliki banyak atribut, diantaranya :

1. No. Pesanan : Nomor unik yang digunakan untuk mengenali produk yang dipesan
2. Nama Pesanan : Nama atau deskripsi dari barang yang dipesan
3. Harga Awal : Harga asli sebelum adanya diskon
4. Harga Setelah Diskon : Harga asli dikurangi dengan diskon yang diberikan
5. Total harga produk : Harga atau jumlah yang harus dibayarkan dalam pesanan
6. Jumlah Produk Dipesan : Kuantitas atau jumlah unit produk yang dibeli dalam satu pesanan oleh pelanggan.
7. Estimasi Potongan Biaya Pengiriman : Perkiraan besaran potongan atau subsidi biaya pengiriman yang diberikan oleh *marketplace* atau penjual dalam transaksi tersebut.
8. Total Pembayaran : Jumlah akhir yang harus dibayar oleh pelanggan setelah menghitung total harga produk dan dikurangi potongan biaya pengiriman, jika ada.

3.3.2 Data Cleaning

Pada tahap data cleaning dilakukan pengecekan apakah ada data yang hilang atau missing value pada data atau kolom sample yang dipilih dengan menggunakan Bahasa Python. Jika terdapat missing value, maka pada bagian tersebut harus dihapus agar tidak mengganggu akurasi data dalam tahap pemodelan nanti. Berikut hasil pengecekan missing value pada data *sample*.

```

0
Jumlah Produk di Pesan    0
Estimasi Potongan Biaya Pengiriman    0
Total Pembayaran         0

dtype: int64
    
```

Gambar 4. Data Missing Value

Pada gambar 4. ditunjukkan bahwa tidak terdapat missing value pada data *sample* yang digunakan.

3.3.3 Data Selection

Tahap ini melibatkan pemilihan data yang relevan untuk keperluan pemodelan. Data yang digunakan dalam penelitian ini sebanyak 504 data gabungan dari data bulan Mei-September 2024. Data yang dipakai mencakup :

1. Variabel Independen yaitu jumlah produk yang dipesan atau produk terjual (X1) dan estimasi potongan biaya pengiriman (X2)
2. Variabel Dependen yaitu total pembayaran atau penjualan (Y)

Berikut data yang akan digunakan :

No	Produk Terjual (X1)	Potongan Ongkir (X2)	Penjualan
1	2	24500	121214
2	4	30000	236210
3	1	40000	60359
4	2	33500	119260

5	2	20500	139944
6	1	18750	56370
7	1	11000	57992
8	1	15000	69432
9	3	40000	206057
10	2	16000	125384
11	1	20000	68300
12	1	20000	79208
13	2	11500	108772
14	1	20500	59600
15	1	20000	54030
504	1	15000	49672

Tabel 1. Data Variabel

3.4 Modify

3.4.1 Modifikasi Data

Pada tahapan sebelumnya yaitu tahap *explore*, tidak terdapat *missing value* pada data sehingga tidak perlu dilakukan penanganan *missing value*. Pada kolom estimasi potongan biaya pengiriman terdapat angka nol, yang menandakan bahwa tidak terdapat potongan biaya pengiriman pada pesanan tersebut. Untuk mencegah data yang tidak normal karena perbedaan *outlier*, oleh karena itu baris yang memiliki angka nol tidak akan digunakan dalam penelitian.

3.4.2 Data Integration

Penggabungan data akan dilakukan dalam tahapan ini. Penggabungannya yaitu data dari rentang waktu 5 bulan berdasarkan sample yang didapat, sehingga data menjadi konsisten dan siap untuk dianalisis lebih lanjut.

3.5 Model

3.5.1 Uji Asumsi Klasik

1. Uji Normalitas

Uji normalitas digunakan untuk memastikan bahwa data dan data residu yang digunakan terdistribusi secara normal. Pada penelitian ini, uji normalitas akan menggunakan uji metode statistik *Kolmogorov-Smirnov*.

One-Sample Kolmogorov-Smirnov Test

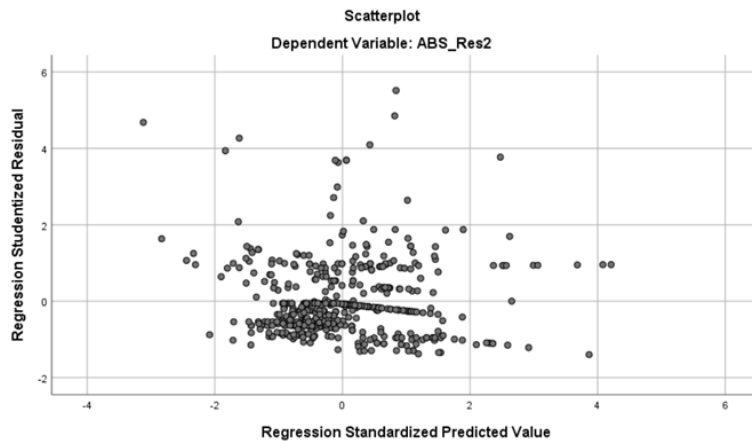
		Unstandardized Residual	
N		504	
Normal Parameters ^{a,b}	Mean	.0000000	
	Std. Deviation	11201.29218	
Most Extreme Differences	Absolute	.068	
	Positive	.068	
	Negative	-.049	
Test Statistic		.068	
Asymp. Sig. (2-tailed)		.000 ^c	
Monte Carlo Sig. (2-tailed)	Sig.	.018 ^d	
	99% Confidence Interval	Lower Bound	.014
		Upper Bound	.021

Gambar 5. Uji Normalitas

Dari gambar 5. pengujian yang dilakukan nilai sig adalah $0.18 > 0.05$ maka dapat disimpulkan data residual berdistribusi normal

2. Uji Heterokedastisitas

Ada 2 cara melihat adanya heterokedastisitas dalam suatu model, diantaranya yaitu dengan menggunakan metode grafik, dan metode non grafik atau uji glejser.



Gambar 6. Uji Heterokedastisitas

Kriteria terjadinya heterokedastisitas pada metode grafik seperti pada gambar 6, adalah terlihat adanya titik-titik menyebar dan tidak membentuk pola. Grafik tersebut tidak menunjukkan adanya heteroskedastisitas, karena titik-titiknya tersebar secara acak tanpa pola tertentu. Namun, interpretasi ini perlu didukung dengan uji statistik tambahan seperti uji Glejser untuk memastikan hasilnya.

Dalam uji glejser jika $sig > \alpha (0.05)$ disimpulkan tidak terjadi heterokedastisitas, sebaliknya jika $sig < \alpha (0.05)$ maka terjadi heterokedastisitas

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	22447.733	2291.224		9.797	.000
	X11	-2.233	1.571	-.063	-1.422	.156
	X22	.140	.101	.062	1.384	.167

a. Dependent Variable: ABS_Res2

Gambar 7. Uji Glejser

Dilihat dari hasil uji glejser yang dilakukan bahwa nilai signifikansi produk terjual yaitu 0.156 dan potongan biaya ongkir (pengiriman) adalah 0.167 menunjukkan nilai signifikansi lebih besar dari α , jadi kesimpulannya tidak terjadi heterokedastisitas.

3. Uji Multikolinearitas

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4806.341	1439.806		3.338	.001		
	produk terjual	55514.120	869.992	.929	63.810	.000	.759	1.318
	potongan ongkir	.263	.066	.058	4.010	.000	.759	1.318

a. Dependent Variable: penjualan

Gambar 8. Uji Multikolinearitas

Dalam uji multikolinearitas, apabila nilai VIF < 10 atau nilai *Tolerance* > 0,1, maka dinyatakan tidak terjadi multikolinearitas. Sebaliknya, jika nilai VIF > 10 atau nilai *Tolerance* < 0,1, maka dinyatakan terjadi multikolinearitas

Nilai *Tolerance* adalah 0.759 dan nilai VIF adalah 1,318 , maka kesimpulannya tidak terjadi multikolinearitas. Sehingga data produk terjual dan potongan biaya pengiriman baik digunakan dalam model regresi.

4. Uji Autokorelasi

Autokorelasi tidak terjadi apabila nilai Durbin Watson berada diantara du sampai dengan (4-du)

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.959 ^a	.919	.919	11223.628	2.118

a. Predictors: (Constant), potongan ongkir, produk terjual

b. Dependent Variable: penjualan

Gambar 9. Uji Autokorelasi

Nilai du dicari pada distribusi nilai tabel Durbin Watson berdasarkan k dan N dengan signifikansi 5%.

k = Banyaknya variabel *independent*

N = Banyaknya data

Dari persamaan tersebut, nilai k = 2, dan nilai N = 504. Lalu, tahap selanjutnya adalah mencari nilai k = 2, N = 504 pada tabel *Durbin Watson* dengan level *significance* sebesar $\alpha = 5\%$.

n\k	1	2	3	4	5	6	7	8	9	10										
250	1.785	1.801	1.777	1.809	1.769	1.817	1.760	1.825	1.752	1.834	1.744	1.842	1.736	1.851	1.727	1.859	1.719	1.868	1.710	1.876
300	1.804	1.817	1.797	1.824	1.791	1.831	1.784	1.838	1.777	1.845	1.770	1.852	1.763	1.859	1.756	1.866	1.749	1.873	1.742	1.880
350	1.819	1.830	1.813	1.836	1.807	1.842	1.802	1.848	1.796	1.854	1.790	1.860	1.784	1.866	1.778	1.872	1.772	1.878	1.766	1.884
400	1.831	1.841	1.826	1.846	1.821	1.851	1.816	1.856	1.811	1.861	1.806	1.866	1.800	1.872	1.795	1.877	1.790	1.882	1.785	1.887
450	1.841	1.850	1.836	1.854	1.832	1.859	1.827	1.863	1.823	1.868	1.818	1.872	1.814	1.877	1.809	1.882	1.805	1.886	1.800	1.891
500	1.849	1.857	1.845	1.861	1.841	1.865	1.837	1.869	1.833	1.873	1.829	1.877	1.825	1.882	1.821	1.886	1.817	1.890	1.812	1.894
550	1.856	1.864	1.853	1.867	1.849	1.871	1.845	1.875	1.842	1.878	1.838	1.882	1.834	1.886	1.831	1.890	1.827	1.893	1.823	1.897
600	1.863	1.869	1.859	1.873	1.856	1.876	1.853	1.879	1.849	1.883	1.846	1.886	1.842	1.890	1.839	1.893	1.836	1.896	1.832	1.900
650	1.868	1.874	1.865	1.877	1.862	1.880	1.859	1.884	1.856	1.887	1.853	1.890	1.849	1.893	1.846	1.896	1.843	1.899	1.840	1.902
700	1.873	1.879	1.870	1.882	1.867	1.884	1.864	1.887	1.861	1.890	1.859	1.893	1.856	1.896	1.853	1.899	1.850	1.902	1.847	1.905
750	1.877	1.883	1.875	1.885	1.872	1.888	1.869	1.891	1.867	1.893	1.864	1.896	1.861	1.899	1.859	1.902	1.856	1.904	1.853	1.907
800	1.881	1.886	1.879	1.889	1.876	1.891	1.874	1.894	1.871	1.896	1.869	1.899	1.866	1.901	1.864	1.904	1.861	1.907	1.859	1.909
850	1.885	1.890	1.883	1.892	1.880	1.894	1.878	1.897	1.875	1.899	1.873	1.902	1.871	1.904	1.868	1.906	1.866	1.909	1.864	1.911
900	1.888	1.893	1.886	1.895	1.884	1.897	1.882	1.899	1.879	1.902	1.877	1.904	1.875	1.906	1.873	1.908	1.870	1.911	1.868	1.913
950	1.891	1.895	1.889	1.898	1.887	1.900	1.885	1.902	1.883	1.904	1.881	1.906	1.879	1.908	1.876	1.910	1.874	1.913	1.872	1.915
1000	1.894	1.898	1.892	1.900	1.890	1.902	1.888	1.904	1.886	1.906	1.884	1.908	1.882	1.910	1.880	1.912	1.878	1.914	1.876	1.916

Tabel 2. *Durbin Watson* $\alpha = 0,05$
 (sumber : (real-statistics.com))

Dari Tabel 2 kita mendapati nilai du untuk N = 504 adalah (1,861), dan N 550 = (1,867), untuk mendapatkan nilai dengan N=504 karena dalam tabel tersebut tidak tersedia maka harus dilakukan interpolasi linear terlebih dahulu. Berikut merupakan rumus interpolasi linear :

$$du = d_{low} + ((n_{target} - n_{low}) / ((n_{high} - n_{low})) \times (d_{high} - d_{low})) \tag{2}$$

$$= d_{high} - d_{low} = 1.867 - 1.861 = 0.006$$

$$= n_{high} - n_{low} = 550 - 500 = 50$$

$$du = d_{low} + ((n_{target} - n_{low}) / ((n_{high} - n_{low})) \times (d_{high} - d_{low}))$$

$$du = 1.861 + ((504 - 500) / ((50)) \times (0.006))$$

$$du = 1.86148 = 1.8615 \text{ (dibulatkan ke 4 desimal)}$$

Jadi untuk nilai 4-du (4-1,8615) dan pada gambar mendapati nilai dari Durbin Watson adalah (2,118), dengan demikian didapatkan hasil seperti di bawah ini:

Diketahui :

$$du = 1,8615$$

$$4 - du = 2,1385$$

$$Durbin\ Watson = 2,118$$

Kesimpulannya adalah $1,8615 < 2,118 < 2,1385$ Tidak ada autokorelasi dalam model regresi ini, dikarenakan nilai Durbin Watson berada diantara nilai du dan $4 - du$.

3.5.2 Uji Kelayakan

1. Uji F (Simultan)

Ada 2 cara pengambilan keputusan dalam uji F simultan yaitu dengan melihat nilai signifikansi dari output anova dan perbandingan dari F_{hitung} dengan F_{tabel} .

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.204E+11	2	3.602E+11	2859.518	.000 ^b
	Residual	6.311E+10	501	125969820.4		
	Total	7.835E+11	503			

a. Dependent Variable: penjualan

b. Predictors: (Constant), potongan ongkir, produk terjual

Gambar 10. Uji F Simultan

Dari nilai signifikansi apabila $sig < 0.05$ maka H_a diterima artinya variabel independent berpengaruh secara simultan terhadap variabel dependen. Dari pengujian terlihat bahwa nilai signifikansi yaitu 0.000 artinya < 0.05 , maka dapat disimpulkan H_a diterima, variabel produk terjual dan variabel potongan biaya pengiriman mempengaruhi penjualan secara simultan.

Berdasarkan perbandingan F_{hitung} dan F_{tabel} , jika nilai $F_{hitung} > F_{tabel}$, maka hipotesis (H_a) diterima. Maka artinya variabel independent mempengaruhi variabel dependen. Rumus mencari F_{tabel} adalah :

$$F_{tabel} = (k ; n-k)$$

$$F_{tabel} = (2 ; 504-2)$$

$$F_{tabel} = (2 ; 502) = 3,013$$

Angka 3,013 didapatkan dari distribusi nilai F_{tabel} statistik dengan menggunakan excel. Dapat disimpulkan bahwa Variabel (X1) produk terjual dan (X2) potongan biaya pengiriman berhubungan secara simultan terhadap Variabel (Y) dikarenakan nilai $F_{hitung} = 2832,640 > F_{tabel} = 3,013$

2. Uji T (Parsial)

Sama dengan Uji F Simultan Uji T Parsial juga memiliki 2 cara dalam pengambilan keputusan yaitu dengan melihat nilai signifikansi dari output anova dan perbandingan dari t_{hitung} dengan t_{tabel} . Uji t-berpasangan (*Paired t-Test*) akan digunakan dalam penelitian ini karena tiap pasangan data berasal dari subjek yang sama atau terkait erat.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	produk terjual - potongan ongkir	-21276.540	8762.730	390.323	-22043.404	-20509.675	-54.510	503	.000

Gambar 11. Uji T Parsial

Berdasarkan nilai signifikansi apabila $sig < 0.05$ artinya H_a diterima, maka terdapat pengaruh antara variabel independent dan variabel dependen. Dari pengujian terlihat bahwa nilai signifikansi yaitu 0.000 artinya < 0.05 , maka dapat disimpulkan variabel produk terjual dan variabel potongan biaya pengiriman mempengaruhi penjualan.

Berdasarkan perbandingan t_{hitung} dengan t_{tabel} , jika nilai $t_{hitung} > t_{tabel}$, maka hipotesis (H_a) diterima. Maka artinya variabel independent mempengaruhi variabel dependen. Rumus mencari t_{tabel} adalah :

$$\begin{aligned}
 t_{tabel} &= a/2 ; n-k-1 \\
 t_{tabel} &= 0,05/2 ; 504-2-1 \\
 t_{tabel} &= 0,025 ; 501 = -1964
 \end{aligned}
 \tag{3}$$

Angka - 0674 didapatkan dari distribusi nilai *t* tabel statistik dengan menggunakan excel. Dapat disimpulkan bahwa Variabel (X1) produk terjual dan (X2) potongan biaya pengiriman berpengaruh terhadap Variabel (Y) dikarenakan nilai *thitung* = - 54.271 > *t* tabel = - 1964

3. Uji Koefisien Determinasi

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.959 ^a	.919	.919	11223.628

a. Predictors: (Constant), potongan ongkir, produk terjual

Gambar 12. Uji Koefisien Determinasi (R2)

Berdasarkan hasil Model Summary pada gambar 12, nilai *Adjusted R Square* menunjukkan angka sebesar 91,9%. Hal ini mengindikasikan bahwa variabel independen, yaitu produk terjual (X1) dan potongan biaya pengiriman (X2), memiliki pengaruh sebesar 91,9% terhadap variabel dependen, yaitu penjualan (Y).

3.5.3 Hasil Uji Regresi Linear Berganda

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4806.341	1439.806		3.338	.001
	produk terjual	55514.120	869.992	.929	63.810	.000
	potongan ongkir	.263	.066	.058	4.010	.000

a. Dependent Variable: penjualan

Gambar 13. Uji Regresi Linear Berganda

Dari gambar 13, maka hasil regresi dapat dijelaskan sebagai berikut:

- a. *Intercept (Constant)*: Nilai intercept sebesar 4806.341 menunjukkan bahwa jika tidak ada produk terjual dan tidak ada potongan ongkir maka penjualan diprediksi sebesar 4806.341.
- b. Koefisien Produk Terjual : Nilai koefisien sebesar 55514.120 menunjukkan bahwa setiap peningkatan satu unit produk terjual akan meningkatkan penjualan sebesar 55514.120, dengan asumsi variabel lainnya konstan.
- c. Koefisien Potongan Ongkos Kirim : Nilai koefisien sebesar 0.263 menunjukkan bahwa setiap peningkatan satu unit potongan ongkir akan meningkatkan penjualan sebesar 0.263, dengan asumsi variabel lainnya konstan.

Persamaan Regresi Dari hasil regresi diatas didapatkan persamaan regresi yang dimana dari hasil regresi tersebut dapat diolah menjadi prediksi penjualan *marketplace* yaitu :

$$\begin{aligned}
 Y &= a + b_1X_1 + b_2X_2 + \dots + b_nX_n \\
 Y &= 4806.341 + 55514.120 X_1 + 0,263 X_2
 \end{aligned}
 \tag{4}$$

Setelah persamaan regresi linear didapatkan, maka untuk melihat estimasi penjualan *marketplace* Aufar Grosir Jeans dapat dilakukan dengan memasukkan nilai X1 dan X2 sesuai dengan frekuensi atau banyaknya produk terjual dan potongan biaya pada periode bulan Mei-September 2024 pada gambar dibawah ini:

Sum of jumlah produk	Sum of	Count of potongan
----------------------	--------	-------------------

terjual	penjualan	ongkir
738	46211370	504

Tabel 3. Frekuensi *Marketplace*

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

$$Y = 4806.341 + 55493,584 X_1 + 0,263 X_2$$

$$Y = 4806.341 + 55493,584 (723) + 0,263 (504)$$

$$Y = 4806.341 + 40,146,266.632 + 132.552$$

$$Y = 40,126,800.125$$

Maka hasil dari estimasi (perkiraan) total penjualan atau total pembayaran *marketplace* AUFAR Grosir *Jeans* adalah sebanyak 40,126,800,125 rupiah. Terjadi penurunan penjualan dari bulan sebelumnya sebanyak 6,084,569.88.

3.6 Asses

Sebelum melakukan evaluasi performa model yaitu dengan MSE, RMSE, dan MAPE, perlu dicari nilai prediksi (forecast) nya terlebih dahulu. Berikut adalah tabel forecast yang didapat dengan menggunakan tools python:

	jumlah produk terjual	potongan ongkir	Actual	Forecast
173	2	16500	119594	120941.790017
274	2	17500	97975	121184.849207
489	1	14000	65272	64013.625186
72	1	14000	62058	64013.625186
305	4	24000	243016	235405.767653
76	1	25000	76197	66687.276268
475	1	12000	64800	63527.506808
140	2	16625	89384	120972.172416
469	1	14500	57994	64135.154781
498	2	34000	100195	125195.325829
218	1	30000	73746	67902.572214
9	2	16000	125384	120820.260423
479	1	20000	58497	65471.980322
78	1	20000	63320	65471.980322
322	1	15000	62800	64256.684375
312	2	14000	128608	120334.142044
124	1	16000	68905	64499.743565
384	1	18000	64038	64985.861943
195	1	20000	60230	65471.980322
447	2	40000	139944	126653.680965

Tabel 4. Nilai Prediksi (*Forecast*)

3.6.1 Mean Squared Error (MSE)

Mean Squared Error: 136820412.26052144
 Root Mean Squared Error: 11697.025786947785
 Mean Absolute Percentage Error (MAPE): 8.99387261572891 %

Gambar 14. Hasil Pengukuran *Error*

Terlihat pada gambar 14, dengan bantuan *tools python* didapatkan MSE nya adalah 136.820.412 Nilai MSE memperlihatkan rata-rata kesalahan kuadrat antara nilai ramalan dan nilai asli. Karena nilai MSE besar, ini menandakan ada kesalahan yang cukup besar pada prediksi, tetapi sulit untuk

menginterpretasikan angka ini secara langsung karena satuannya adalah kuadrat dari satuan target.

3.6.2 Root Mean Squared Error (RMSE)

Hasil untuk perhitungan RMSE didapatkan nilai 11,697.02. RMSE adalah akar kuadrat dari MSE, sehingga hasilnya berada dalam skala yang sama dengan target. Nilai RMSE sebesar 11,697.02 menunjukkan bahwa rata-rata prediksi model memiliki deviasi sekitar 11,697.02 dari nilai aktual. target adalah penjualan dalam rupiah, maka prediksi model memiliki rata-rata kesalahan sekitar 11,697 rupiah, yang lebih mudah dipahami dibandingkan MSE.

3.6.3 Mean Absolute Percentage Error (MAPE)

MAPE memberikan gambaran tentang kesalahan rata-rata dalam bentuk persentase relatif terhadap nilai aktual. Dengan MAPE sebesar 8,99%, ini berarti prediksi model rata-rata memiliki kesalahan sekitar 8,99% dari nilai aktualnya. Hasil ini cukup baik jika kesalahan prediksi di bawah 10% membuktikan bahwa model yang digunakan cukup akurat.

4. Kesimpulan

Melihat dari hasil dan pembahasan penelitian diatas, maka Kesimpulan yang dapat diambil adalah :

1. Algoritma regresi linear berganda berhasil diterapkan untuk memprediksi penjualan *marketplace* berdasarkan produk terjual dan potongan biaya pengiriman. Model ini menunjukkan bahwa variabel produk terjual berpengaruh secara signifikan dan positif terhadap penjualan, diikuti oleh kontribusi kecil namun tetap signifikan dari potongan biaya pengiriman. Model regresi linear berganda yang digunakan memiliki akurasi tinggi dengan nilai Adjusted R Square sebesar 91,9% dan MAPE sebesar 8,99%, sehingga dapat diandalkan untuk mendukung pengambilan keputusan dalam meningkatkan penjualan di *marketplace* AUFAR Grosir Jeans. Dengan regresi linear berganda juga didapatkan prediksi penjualan AUFAR Grosir Jeans ditahun selanjutnya dengan frekuensi yang sama dengan sample adalah sebesar 40,126,800,125 rupiah.
2. Metode SEMMA telah digunakan secara efektif dalam analisis data, mulai dari tahap sample, eksplorasi, pembersihan, seleksi, hingga pemodelan dan evaluasi. Hasil analisis menunjukkan bahwa produk terjual dan potongan biaya pengiriman secara simultan memiliki pengaruh yang signifikan terhadap penjualan di *marketplace*. Keputusan tersebut juga didukung oleh hasil uji F dengan nilai signifikansi $0,000 < 0,005$ dan $F_{hitung} = 2832,640 > F_{tabel} = 3,013$ yang menunjukkan variabel independent yaitu produk terjual dan potongan biaya pengiriman berpengaruh secara simultan terhadap variabel dependen yaitu penjualan. Serta uji T dengan nilai signifikansi $0,000 < 0,05$ dan $t_{hitung} = - 54,271 > t_{tabel} = - 1964$ yang menunjukkan adanya hubungan yang signifikan dari variabel-variabel independen.

Daftar Pustaka

- [1] Ahdiat Adi, "5 E-Commerce dengan Pengunjung Terbanyak Sepanjang 2023," databokskatadata.co.id.
- [2] E. Triyanto, H. Sismoro, and A. D. Laksito, "IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 4, no. 2, pp. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.
- [3] R. Badzlin *et al.*, "Implementasi Data Mining Prediksi Penjualan Produk Semen Menggunakan Metode Linear Regression (Studi Kasus PT. Toyo Mortar Indonesia)," 2024. [Online]. Available: <https://journal.stmiki.ac.id>
- [4] Ferdinan, "Prediksi Jumlah Penjualan Tahun 2024 Menggunakan Metode Regresi Linier," 2024, [Online]. Available: <https://djournals.com/resolusi>
- [5] A. Al-Fadhilah Nur Wahyudin, A. Primajaya, and A. Susilo Yuda Irawan, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu Implementation of Double Linear Regression Algorithm On Sales Estimation of Astra Isuzu Car," 2020.
- [6] N. Eka Pratiwi, L. Suryadi, F. Ardhy, and P. Riswanto, "PENERAPAN DATA MINING PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR(K-NN) (STUDI KASUS : TOKO ZERITA MEUBEL)," 2022.
- [7] A. Selay *et al.*, "SISTEM INFORMASI PENJUALAN," 2023.
- [8] M. Evi Yuliana and S. Isti Kharisma Auliya Alamsyah, "Communication Development in Marketplace Business," *International Journal of Progressive Sciences and Technologies*

- [9] (IJPSAT, vol. 27, no. 1, pp. 60–66, 2021, [Online]. Available: <http://ijpsat.ijsh-journals.org>
S. Rahman *et al.*, “PYTHON : DASAR DAN PEMROGRAMAN BERORIENTASI OBJEK TAHTA MEDIA GROUP.”
- [10] A. Fitri, R. Rahim, and Nurhayati, “Anisa , Buku Dasar-dasar Statistika untuk Penelitian,” 2023, Accessed: Oct. 23, 2024. [Online]. Available: <https://repository.unugiri.ac.id:8443/id/eprint/4882/1/Anisa%20,%20Buku%20%20Dasar-dasar%20Statistika%20untuk%20Penelitian.pdf>
- [11] A. Anwar, “STATISTIKA UNTUK PENELITIAN PENDIDIKAN,” Kediri, 2020.
- [12] A. Novebrian Maharadja, I. Maulana, and B. Arif Dermawan, “Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi,” 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [13] Soecahyadi, “ANALISA STATISTIK DENGAN APLIKASI SPSS,” 2019. [Online]. Available: <https://www.researchgate.net/publication/332269530>