

**PREDIKSI JUMLAH OMSET PENJUALAN DENGAN
MENGUNAKAN MACHINE LEARNING LINIER REGRESI
STUDI KASUS PADA SEBLAK MADIT**

SKRIPSI

DISUSUN OLEH

ABDURRAHMAN ARIF

NPM. 2109010091



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

**PREDIKSI JUMLAH OMSET PENJUALAN DENGAN
MENGUNAKAN MACHINE LEARNING LINIER REGRESI
STUDI KASUS PADA SEBLAK MADIT**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

ABDURRAHMAN ARIF

NPM 2109010091

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

LEMBAR PENGESAHAN

Judul Skripsi : PREDIKSI JUMLAH OMSET PENJUALAN DENGAN
MENGUNAKAN MACHINE LEARNING LINIER
REGRESI STUDI KASUS PADA SEBLAK MADIT
Nama Mahasiswa : ABDURRAHMAN ARIF
NPM : 2109010091
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Halim Maulana, ST, M.Kom)
NIDN. 0121119102

Ketua Program Studi



(Martiano S.Pd, S.Kom., M.Kom)
NIDN. 0128029302

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PREDIKSI JUMLAH OMSET PENJUALAN DENGAN MENGGUNAKAN MACHINE LEARNING LINIER REGRESI STUDI KASUS PADA SEBLAK MADIT

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2025

Yang membuat pernyataan



Abdurrahman Arif

NPM. 2109010091

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Abdurrahman Arif
NPM : 2109010091
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PREDIKSI JUMLAH OMSET PENJUALAN DENGAN MENGGUNAKAN
MACHINE LEARNING LINIER REGRESI STUDI KASUS PADA
SEBLAK MADIT**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2025

Yang membuat pernyataan



Abdurrahman Arif

NPM. 2109010091

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Abdurrahman Arif
Tempat dan Tanggal Lahir : Medan 16 november 2003
Alamat Rumah : Jl.Ampera raya no 24b
Telepon/Faks/HP : 082166199851
E-mail : rahmanabdur82816@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SD MUHAMMADIYAH 02 TAMAT: 2015
SMP : SMP MUHAMMADIYAH 57 TAMAT: 2018
SMA : SMK TRITECH INFORMATIKA TAMAT: 2021

KATA PENGANTAR



Puji syukur Penulis ucapkan kepada Allah SWT atas rahmat-Nya Penulis dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Halim Maulana, S.T., M.Kom, selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi dan juga selaku Dosen Pembimbing Penulis.
4. Bapak Martiano, S.Pd., S.Kom., M.Kom Ketua Program Studi Sistem Informasi
5. Ibu Yoshida Sary, S.E., S.Kom., M.Kom Sekretaris Program Studi Sistem Informasi
6. Bapak Abdurrahim dan ibu Yunita wiratna selaku orang tua kandung saya yang selalu mendoakan saya dalam segala hal apapun dan men support saya terus menerus
7. Yang tersayang kakak pertamaku Rahmi Ayunda, S.E, .Terima kasih telah memberikan dukungan dan motivasi serta terima kasih telah setia meluangkan waktunya untuk menjadi tempat dan pendengar terbaik penulis sampai akhirnya penulis dapat menyelesaikan skripsi ini.

8. Yang tersayang adiku Muhammad Arifin Ilham Terima kasih telah menjadi penyemangat dan penghibur penulis dikala penulis jenuh menghadapi permasalahan selama ini.
9. Yang saya sayangin Aski Maisyaroh Lubis selaku Pacar saya yang selalu mensupport saya di saat penulisan skripsi,serta selalu menyemangatin saya di saat penulisan Terima Kasih telah selalu ada untuk saya
10. Dan yang terakhir, kepada diri sendiri. Abdurrahman Arif. Terima kasih karna telah bertahan dan tetap memilih berusaha selama ini. Terima kasih karena memutuskan tidak menyerah sesulit apapun proses penyusunan skripsi ini. Terima kasih kamu hebat bisa sampai di titik ini. Berbahagia dan sukses lah di masa depan

**PREDIKSI JUMLAH OMSET PENJUALAN DENGAN MENGGUNAKAN
MACHINE LEARNING LINIER REGRESI STUDI KASUS PADA
SEBLAK MADIT**

ABSTRAK

Pada era digital saat ini, pemanfaatan teknologi dalam pengambilan keputusan bisnis menjadi sangat penting, termasuk dalam hal prediksi penjualan. Penelitian ini bertujuan untuk memprediksi omset penjualan harian pada UMKM Seblak Madit menggunakan metode regresi linier sederhana berbasis machine learning. Data yang digunakan merupakan data historis penjualan selama 30 hari, dengan variabel prediktor berupa jumlah porsi terjual, harga rata-rata produk, promosi, dan cuaca. Metodologi penelitian mencakup pengumpulan data, preprocessing, pembangunan model, dan evaluasi menggunakan metrik MAE, MSE, dan R-squared.

Hasil penelitian menunjukkan bahwa model regresi linier sederhana memiliki performa prediktif yang sangat baik dengan nilai R^2 sebesar 0.981, MAE sebesar 8.879, dan RMSE sebesar 10.909. Faktor yang paling signifikan memengaruhi omset adalah jumlah pengunjung dan keberadaan promosi. Hasil ini memberikan implikasi strategis dalam perencanaan stok bahan baku, strategi promosi, serta efisiensi operasional harian. Penelitian ini juga merekomendasikan pengembangan model lebih lanjut dengan metode machine learning lain seperti Random Forest dan XGBoost untuk meningkatkan akurasi prediksi.

Kata Kunci: Machine learning; Linier Regresi; Prediksi;.

Forecasting Sales Revenue Using Linear Regression-Based Machine Learning: A Case Study of Seblak Madit

ABSTRACT

In today's digital era, leveraging technology in business decision-making is essential, especially for sales forecasting. This study aims to predict the daily revenue (omzet) of the UMKM Seblak Madit using a simple linear regression method based on machine learning. The dataset comprises 30 days of historical sales data, with predictor variables including the number of portions sold, average product price, promotional activities, and weather. The research methodology includes data collection, preprocessing, model development, and evaluation using MAE, MSE, and R-squared metrics.

The results indicate that the simple linear regression model performs well with an R^2 value of 0.981, MAE of 8,879, and RMSE of 10,909. The most influential factors affecting revenue are the number of visitors and the presence of promotions. These findings have strategic implications for inventory planning, promotional strategies, and daily operational efficiency. The study also recommends exploring other machine learning methods such as Random Forest and XGBoost to improve prediction accuracy.

Keywords: Machine Learning; Linier Regresi; Prediksi

DAFTAR ISI

LEMBAR PENGESAHAN	i
PENYATAAN ORISINALITAS	ii
PENYATAAN PERSETUJUAN PUBLIKASI	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I. PENDAHULUAN	1
1.1. LATAR BELAKANG MASALAH.....	1
1.2. RUMUSAN MASALAH	2
1.3. BATASAN MASALAH	2
1.4. TUJUAN PENELITIAN	3
1.5. MANFAAT PENELITIAN	3
BAB II. LANDASAN TEORI	4
2.1. PENELITIAN TERDAHULU	4
2.2. LINIER REGRESI	5
2.3. MACHINE LEARNING.....	8
2.4. PHYTHON	9
2.5. PREDIKSI	10
BAB III. METODOLOGI PENELITIAN	12
3.1. DESAIN DAN PENELITIAN	12
3.2. DATA DAN SUMBER DATA.....	12
3.3. TAHAPAN PENELITIAN.....	13
3.3.1 PENGUMPULAN DATA	13
3.3.2 PREPROCESSING DATA.....	13
3.3.3 PENERAPAN MODEL LINIER REGRESI	14
3.3.4 EVALUASI MODEL.....	14
3.4. ALAT DAN TEKNOLOGI	17
3.5. VALIDASI DAN PENGUJIAN	18
3.6. KESIMPULAN METODOLOGI	18
3.7. JADWAL PENELITIAN	19
BAB IV: HASIL DAN PEMBAHASAN	20
4.1 Pendahuluan.....	20
4.1.1 Alur penelitian	20
4.2 Deskripsi Dataset	21
4.2.1 Karakteristik Dataset	21
4.2.2 Deskripsi Setiap Variabel	22
4.2.3 Statistik Deskriptif	22
4.2.4 Visualisasi Dataset Awal	22
4.3 Pengolahan Data (Preprocessing).....	24
4.3.1 Pemeriksaan dan Penanganan Nilai Hilang.....	24
4.3.2 Pemeriksaan Outlier	24
4.3.3 Penyesuaian Tipe Data	24

4.3.4 Normalisasi atau Standardisasi	25
4.3.5 Feature Engineering.....	25
4.4 Pemodelan Regresi	26
4.4.1 Model yang Digunakan: Regresi Linier Sederhana.....	26
4.4.2 Proses Pelatihan Model	27
4.4.3 Hasil Training Model.....	27
4.4.4 Visualisasi Garis Regresi.....	27
4.5 Evaluasi Model	28
4.6 Analisis Faktor yang Mempengaruhi Omset	30
4.7 Alternatif Peningkatan Model	31
4.8 Implikasi Hasil Prediksi Terhadap Keputusan Bisnis	32
BAB V: KESIMPULAN DAN SARAN.....	34
5.1 Kesimpulan	34
5.2 Saran.....	34
DAFTAR PUSTAKA	36

DAFTAR TABEL

Tabel 3.1 Jadwal Penelitian	19
Tabel 4.1 Karakteristik Dataset	21
Tabel 4.2 Deskripsi Setiap Variabel	21
Tabel 4.3 Statistik Deskriptif	22
Tabel 4.4 Kesimpulan Pengolahan Data	26
Tabel 4.5 Analisis Faktor yang Mempengaruhi Omset.....	30

DAFTAR GAMBAR

Gambar 3.1 Omset Penjualan/Data Mentah	13
Gambar 3.2 Tahapan Penelitian.....	13
Gambar 3.3 Tampilan Model.....	15
Gambar 3.4 Rumus MAE	16
Gambar 3.5 Rumus MSE.....	16
Gambar 3.6 Rumus R-Squared.....	17
Gambar 4.1 Deskripsi Dataset	23
Gambar 4.2 Visualisasi Garis Regresi.....	28
Gambar 4.3 Evaluasi Model	29

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Seblak madit adalah UMKM di bawah naungan Pemerintahan Provinsi Sumatera Utara (PEMPROVSU) yang berdiri sejak 2018, owner seblak madit bernama Rindy Nadila Lubis (38) dan kini seblak madit memiliki 3 outlet cabang yaitu di (jl. aman 1, jl. tembung pasar 7, jl. sei padang tosa coffe). Pada saat ini pembukuan seblak madit masih melakukan manual, dan stok barang juga masih menggunakan filling

Pada era digital saat ini, penggunaan teknologi dan data dalam pengambilan keputusan bisnis menjadi hal yang sangat penting. Salah satu teknologi yang banyak digunakan adalah machine learning, yang memiliki kemampuan untuk memprediksi tren dan pola berdasarkan data historis. Dalam dunia bisnis kuliner, khususnya usaha kecil dan menengah (UKM), pemanfaatan teknologi ini dapat membantu pengusaha dalam merancang strategi pemasaran dan produksi. Seblak Madit, sebagai salah satu usaha kuliner lokal yang cukup berkembang, menghadapi tantangan dalam memprediksi omset penjualan di tengah persaingan yang semakin ketat. Oleh karena itu, penerapan machine learning, khususnya metode regresi linier, diharapkan mampu memberikan solusi praktis untuk memprediksi omset penjualan secara akurat

Dalam penelitian ini, metode yang digunakan untuk memprediksi penjualan adalah regresi linear. Regresi linear merupakan salah satu teknik statistik yang bertujuan untuk menganalisis hubungan antara dua variabel. Pada konteks ini, variabel yang dianalisis mencakup jumlah penjualan Seblak Madit dan berbagai faktor yang memengaruhinya. Melalui penerapan metode regresi linear, diharapkan dapat diperoleh gambaran atau pola hubungan yang dapat digunakan untuk melakukan prediksi secara lebih akurat. seblak madit dapat mengetahui prediksi penjualan untuk kedepannya serta dapat membantu dalam mengambil keputusan yang tepat. Hasil prediksi tersebut kemudian akan berkembang menjadi sebuah pertimbangan untuk memperkirakan apa yang harus dilakukan untuk

jangka pendek dan jangka panjang suatu usaha. Penelitian mengenai prediksi penjualan ini menjadi sangat penting untuk membantu mengelola bisnis dalam mengambil keputusan dan memperkirakan apa yang harus dilakukan agar bisnis tetap bertahan

Seblak Madit juga menghadapi masalah signifikan dalam pengelolaan bahan baku untuk persediaan penjualan harian. Kesulitan yang sering dialami adalah menentukan jumlah bahan baku yang tepat untuk dibeli, sehingga terjadi dua kemungkinan yang merugikan. Di satu sisi, kelebihan pembelian bahan baku dapat menyebabkan kerugian karena bahan yang tidak terpakai berpotensi rusak atau tidak dapat digunakan lagi. Di sisi lain, kekurangan bahan baku dapat menghambat operasional penjualan dan menurunkan pendapatan harian akibat tidak terpenuhinya permintaan pelanggan. Masalah ini mencerminkan pentingnya prediksi yang akurat dalam menentukan kebutuhan bahan baku untuk mendukung keberlangsungan dan pertumbuhan usaha.

1.2. Rumusan Masalah

Adapun rumusan masalah pada penelitian ini, yaitu:

- 1) Bagaimana metode regresi linier dalam machine learning dapat diterapkan untuk memprediksi omset penjualan
- 2) Faktor apa saja yang berpengaruh signifikan terhadap prediksi tersebut
- 3) sejauh mana tingkat akurasi model prediksi yang dihasilkan

1.3. Batasan Masalah

Dalam penulisan proposal ini, batasan masalah yang ada pada rumusan masalah diatas yaitu:

- 1) Penelitian ini hanya mencakup data penjualan Seblak Madit dalam kurun waktu tertentu yang telah ditentukan, sehingga hasil prediksi terbatas pada rentang waktu tersebut.
- 2) Variabel yang dianalisis dalam penelitian ini meliputi data historis penjualan, faktor musiman, dan promosi yang dilakukan. Variabel lain

yang mungkin memengaruhi penjualan namun tidak tersedia dalam data tidak akan disertakan.

- 3) Metode yang digunakan dalam penelitian ini adalah regresi linier sebagai salah satu teknik dalam machine learning, sehingga tidak membahas metode lain seperti decision tree atau neural networks.
- 4) Prediksi dilakukan untuk membantu perencanaan operasional jangka pendek, sehingga tidak mencakup perencanaan strategis jangka panjang.

1.4. Tujuan Penelitian

- 1) mengidentifikasi variabel-variabel yang memengaruhi omset penjualan
- 2) membangun model prediksi menggunakan metode regresi linier, dan
- 3) mengevaluasi tingkat akurasi model prediksi yang dihasilkan

1.5. Manfaat Penelitian

- 1) Membantu Seblak Madit meningkatkan efisiensi dalam perencanaan produksi harian dengan prediksi kebutuhan bahan baku yang lebih akurat
- 2) Mendukung pengambilan keputusan strategis dalam pemasaran dan operasional bisnis berdasarkan data yang dapat diandalkan.
- 3) Mengurangi risiko kerugian akibat kelebihan atau kekurangan bahan baku dalam operasional sehari-hari.

BAB II

LANDASAN TEORI

2.1. Penelitian terdahulu

Sebelumnya, telah ada beberapa penelitian yang berkaitan dengan machine learning linier regresi yang pertama di bahas ini adalah penelitian tentang Penerapan Machine Learning dengan Model Regresi Linier terhadap Analisis Kualitas Hasil Petik Teh di PT Perkebunan Nusantara VIII Kebun Sedep.

Penelitian ini berfokus pada implementasi model machine learning, khususnya regresi linier, dalam menganalisis kualitas hasil petik teh. Machine learning sendiri merupakan bagian dari kecerdasan buatan (Artificial Intelligence) yang dapat digunakan untuk membangun sistem cerdas. Model regresi linier dalam machine learning diterapkan untuk mengevaluasi kualitas petikan teh berdasarkan data yang tersedia.

Metodologi penelitian ini terbagi dalam tiga tahapan utama. Tahap pertama adalah observasi langsung di lapangan guna memperoleh pemahaman kontekstual mengenai proses petik teh. Tahap kedua melibatkan proses pengumpulan data yang relevan sebagai input bagi model machine learning. Tahap ketiga adalah wawancara dengan karyawan yang terlibat langsung dalam proses analisis hasil petikan teh, untuk mendapatkan wawasan mendalam dan informasi pendukung lainnya.

Hasil dari penelitian ini menunjukkan bahwa penerapan model regresi linier berhasil dilakukan dan menghasilkan sebuah prototipe yang dapat berfungsi sebagai dasar pengembangan aplikasi untuk pengguna akhir (end-user) di masa depan. Kesimpulannya, model machine learning regresi linier ini mampu melakukan analisis secara otomatis terhadap kualitas hasil petik teh, sehingga dapat mendukung efisiensi dan akurasi dalam proses evaluasi mutu produksi.

(Yudi, Wildan, 2022)

Penelitian tentang prediksi omset juga di laksanakan pada restoran Go Chicken dengan menggunakan metode linier regresi Data penjualan harian restoran selama periode tertentu dikumpulkan dan dianalisis Dimana model regresi linier

di kembangkan untuk memprediksi penjualan harian dalam penelitian ini menggunakan machine learning Waikato Environment of Knowledge Analysis (WEKA).(Alya,Fatimah,Rozikin, 2023). Berdasarkan kedua penelitian tersebut kita mengetahui bahwa machine learning linier regresi memiliki hasil Analisa yang mampu memprediksi penjualan dengan Tingkat akurasi yang di inginkan penelitian seperti ini dapat membandu menejemen UMKM dalam perencanaan produksi,persediaan,dan srategi pemasaran.

Di penelitian ke tiga ini juga menerapkan linier regresi untuk memprediksi persediaan jenis obat tablet di klinik teluk banten Salah satu fasilitas yang tersedia di klinik ini adalah layanan rawat jalan dan instalasi farmasi yang bertanggung jawab dalam mengelola serta mengatur persediaan obat. Penelitian ini bertujuan untuk merancang prediksi dalam pengelolaan stok obat. Saat ini, pengelolaan persediaan obat di klinik dilakukan secara manual setiap satu minggu sekali. Pencatatan transaksi penjualan obat masih dilakukan secara konvensional, yaitu melalui buku besar. Akibatnya, proses pemantauan dan pengendalian persediaan belum berjalan secara optimal, serta belum memungkinkan dilakukan prediksi terhadap jumlah obat yang perlu disediakan untuk periode berikutnya. Hal ini terjadi karena data penjualan yang tersedia hanya difungsikan sebagai arsip tanpa dimanfaatkan lebih lanjut sebagai dasar untuk mendukung proses prediksi kebutuhan obat pada minggu mendatang.,(Harsiti,Zaenal,Ela, 2022)

Berdasar kan penelitian dan penerapan linier regresi Permasalahan ini dapat diselesaikan dengan menerapkan metode regresi linier sederhana sebagai pendekatan dalam menghasilkan informasi prediktif terkait persediaan obat, khususnya pada proses penjualan obat jenis tablet. Dari hasil perhitungan menggunakan metode regresi linier, diperoleh tingkat akurasi sebesar 98,505%. Hal ini menunjukkan bahwa metode tersebut layak dan efektif untuk digunakan dalam memprediksi stok obat tablet (ibuprofen) di Klinik Teluk.

2.2. Linier regresi

Regresi linier sederhana merupakan salah satu teknik regresi yang digunakan sebagai alat dalam inferensi statistik untuk menganalisis sejauh mana pengaruh suatu variabel independen terhadap variabel dependen. Secara umum, variabel yang menjadi faktor penyebab biasanya dilambangkan sebagai variabel bebas.

dengan X atau disebut dengan predictor sedangkan variabel akibat dilambangkan dengan Y atau disebut dengan response

Untuk menentukan apakah hubungan antara variabel-variabel yang dikaji bersifat linear atau non-linear, terdapat beberapa metode yang dapat digunakan, di antaranya adalah metode tangan bebas dan metode kuadrat terkecil. Metode tangan bebas dilakukan dengan bantuan diagram pencar (scatter plot) yang secara visual menunjukkan pola hubungan antar variabel. Meskipun metode ini tergolong sederhana dan mudah digunakan, tingkat akurasi relatif rendah karena bergantung pada interpretasi visual semata.

Sebaliknya, metode kuadrat terkecil menggunakan pendekatan matematis dengan rumus-rumus tertentu untuk menghitung garis regresi yang paling sesuai dengan data. Metode ini mampu memberikan hasil yang lebih akurat dibandingkan metode tangan bebas. Namun, tingkat kompleksitasnya juga lebih tinggi, karena kesalahan kecil dalam perhitungan dapat berdampak signifikan terhadap ketepatan hasil regresi yang dihasilkan.

Berikut penjelasan dari metode-metode berikut :

a. Metode Tangan Bebas merupakan salah satu cara untuk menentukan jenis regresi dengan memanfaatkan diagram pencar (scatter plot). Prosedur dari metode ini dimulai dengan menggambar sumbu koordinat, yaitu sumbu X (horizontal) dan sumbu Y (vertikal). Selanjutnya, nilai-nilai dari masing-masing variabel diplot dalam bentuk titik-titik pada bidang diagram tersebut. Apabila titik-titik tersebut membentuk pola yang menyerupai garis lurus, maka hubungan antar variabel dapat dikategorikan sebagai regresi linear. Sebaliknya, jika titik-titik tersebar secara acak dan tidak menunjukkan pola garis lurus, maka dapat disimpulkan bahwa hubungan antar variabel bersifat non-linear.

b. Metode kuadrat terkecil adalah cara yang digunakan untuk menentukan regresi dengan menggunakan rumus-rumus yang telah ditentukan. Dapat diketahui jika variabel-variabel yang akan dikorelasikan terdiri dari variabel X sebagai variabel bebas dan variabel Y sebagai variabel terikat.

Maka untuk melakukan analisis regresi dapat dirumuskan menjadi persamaan berikut yaitu

$$Y = a + bX$$

Dimana :

Y = Variabel Respon atau Variabel Akibat

X = Variabel Predictor atau Variabel Faktor Penyebab

a = Konstanta

b = Koefisien Regresi (Kemiringan) atau besarnya Response yang ditimbulkan oleh Predictor.

Yang mana nilai koefisien-koefisien a dan b dapat dihitung dengan persamaan rumus berikut ini :

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{n(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$$

c. Uji Korelasi merupakan suatu metode statistik yang digunakan untuk mengukur sejauh mana kekuatan hubungan antara dua variabel, yaitu variabel X dan variabel Y. Nilai korelasi dilambangkan dengan huruf r . Dalam konteks penelitian ini, semakin besar nilai r , maka semakin kuat hubungan antara variabel X dan Y, yang mengindikasikan adanya korelasi yang mendekati sempurna. Sebaliknya, semakin kecil nilai r , maka hubungan antar variabel cenderung lemah. Apabila nilai r sama dengan 0, maka dapat disimpulkan bahwa tidak terdapat hubungan atau korelasi antara kedua variabel tersebut.

Adapun klasifikasi kekuatan korelasi menurut Calton adalah sebagai berikut:

- a) $r = 0,00 - 0,25 \rightarrow$ menunjukkan hubungan yang sangat lemah atau bahkan tidak ada korelasi,
- b) $r = 0,26 - 0,50 \rightarrow$ menunjukkan adanya hubungan yang sedang atau mendekati korelasi,
- c) $r = 0,51 - 0,75 \rightarrow$ menunjukkan hubungan yang kuat,

- d) $r = 0,76 - 1,00 \rightarrow$ menunjukkan hubungan yang sangat kuat hingga mendekati sempurna.

Berdasarkan klasifikasi tersebut, nilai koefisien korelasi r dapat dihitung menggunakan rumus tertentu yang sesuai dengan metode statistik yang digunakan.

$$r = \frac{N \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{\{N \cdot \Sigma X^2 - (\Sigma X)^2\} \{N \cdot \Sigma Y^2 - (\Sigma Y)^2\}}}$$

Dimana :

r : Uji korelasi

N : Nilai data yang akan dijumlahkan

X : Variabel konstanta penyebab (Pertambahan Penduduk)

Y : Variabel konstanta akibat (Jumlah Kebutuhan Energi Listrik)

(Ayuni,Fitrianah, 2019)

2.3.Machine learning

Machine Learning adalah suatu proses analitik yang dirancang untuk mengidentifikasi pola dalam data serta hubungan antara variabel-variabel data. Salah satu keunggulan utama dari machine learning adalah kemampuannya untuk menganalisis hubungan non-linear yang kompleks, dan mengelola variabel input yang rumit. Berbagai model machine learning dapat di sesuaikan untuk menganalisis data melalui metode seperti klasifikasi, pengelompokan, dan penambahan aturan asosiasi, sesuai dengan pengumpulan data dan tujuan analisis (Sidik, Ansawarman., 2022)

Keuntungan utama dari penggunaan machine learning adalah ketika dataset yang cukup besar telah terkumpul, algoritme machine learning dapat mempelajari bagaimana memanfaatkan data tersebut secara otomatis machine learning memiliki peran penting dalam memantau dan menganalisis proses pembelajaran di sekolah. Ini dapat digunakan untuk memprediksi kinerja siswa, memberikan dukungan akademik, panduan, serta saran dari mentor, mengevaluasi efisiensi metode pembelajaran, serta memberikan umpan balik yang bermanfaat bagi guru dan siswa. Dengan memanfaatkan data terkait latar belakang masalah dan perkembangan siswa selama satu semester, machine learning terbukti efektif dalam memprediksi perkembangan siswa di berbagai jenjang pendidikan (Alfarizi, 2023).

Analisis statistik secara tradisional, karena menekankan kinerja prediktif daripada sifat teoritis yang telah terbukti dan asumsi populasi sebelumnya. Teknik-teknik Machine learning digunakan untuk menemukan model atau pola dalam data, yang 10 mendukung proses pengambilan Keputusan Kemampuan untuk memprediksi kinerja siswa sangat penting dalam sistem pendidikan modern. Namun, belum ada kejelasan mengenai model machine learning mana yang paling efektif dalam memprediksi kinerja siswa dan mana yang paling baik dalam meningkatkan hasil belajar

Berbagai metode penambangan data biasanya digunakan untuk mengekstrak informasi tersembunyi dari volume data yang besar. Model-model machine learning yang umum digunakan antara lain decision trees, neural network, bayesian classifier, nearest neighbor, support vector machine, random forest, logistic regression, linear discriminant analysis, multiple regression, dan selforganised. Machine learning merupakan pendekatan komputasional yang memungkinkan sistem untuk belajar dari pola data dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit.

2.4.Python

Python merupakan salah satu bahasa pemrograman yang sangat populer, khususnya dalam bidang analisis data. Kepopuleran ini didorong oleh kemudahan

dalam mempelajari dan menggunakan Python, sehingga dapat diakses oleh berbagai kalangan usia dan tingkat pengalaman. Selain itu, Python bersifat open source, artinya dapat digunakan secara bebas di berbagai sistem operasi dan oleh siapa saja tanpa batasan lisensi. Salah satu keunggulan utama Python terletak pada ketersediaan berbagai pustaka (*library*) yang mendukung berbagai kebutuhan analisis data. Contohnya, NumPy digunakan untuk komputasi numerik, Pandas untuk pengolahan dan analisis data, Matplotlib untuk visualisasi data, serta Scikit-learn untuk implementasi algoritma machine learning. Kehadiran pustaka-pustaka ini menjadikan Python sebagai alat yang efisien dan fleksibel dalam proses analisis data. Lebih dari itu, Python juga memiliki kemampuan integrasi yang baik dengan teknologi lain, seperti sistem basis data, perangkat big data, dan framework web. Hal ini memungkinkan pengguna untuk mengakses, mengelola, dan menganalisis data dari berbagai sumber dengan lebih mudah dan efektif. Secara keseluruhan, Python dikenal sebagai bahasa pemrograman yang kuat, mudah dipelajari, dan banyak digunakan dalam berbagai bidang, mulai dari pengembangan perangkat lunak, kecerdasan buatan (Artificial Intelligence), hingga pengembangan web dan analisis data berbasis machine learning. (Surbakti, Angelyca, Anita. 2024) . (Dalam ulasan ini, Python akan digunakan sebagai bahasa pemrograman untuk meramalkan harga jual beras di kota Pematang Cengkering, Daerah Medang Deras, rejim Batubara

Penggunaan bahasa pemrograman Python dalam bidang pertanian sangat berguna untuk memprediksi harga jual padi dengan faktor-faktor yang ada. Terlebih lagi, bahasa pemrograman Python dapat menampilkan visualisasi data dengan cepat sehingga mudah untuk melihat tinggi rendahnya harga penjualan padi tersebut. Hal ini memudahkan para peneliti dan kelompok tani untuk memprediksi harga jual padi. Python memberikan keuntungan signifikan dalam bidang pertanian dengan memanfaatkan Artificial Intelligence (AI) dan machine learning untuk memahami data. Python menawarkan layanan pengembangan yang efektif dalam menganalisis data terkait harga jual padi. Untuk itu, ingin diketahui lebih jauh penerapan bahasa pemrograman Python dalam memprediksi harga jual padi. Penggunaan bahasa pemrograman Python dalam bidang pertanian sangat berguna untuk mengidentifikasi harga jual padi dengan faktor-faktor yang ada

Python demikian juga item yang terletak pada bahasa pemrograman tingkat yang tidak dapat disangkal yang dapat digunakan untuk memuluskan kumpulan informasi yang sangat kompleks. Elemen lain yang menempatkan Python ini di peta di antara para peneliti dan penganalisis informasi adalah kemampuan beradaptasinya. Karena kemampuan beradaptasinya, bahasa memungkinkan untuk membuat informasi model, mengatur indeks informasi, Membuat perhitungan yang didukung ML (Machine Learning), layanan web, dan melaksanakan perluasan informasi untuk menyelesaikan berbagai usaha dalam jangka waktu yang singkat. Python memiliki berbagai pilihan paket representasi yang dapat diakses. Hal ini menjadikan Python sebagai perangkat prioritas tinggi untuk pemeriksaan informasi serta untuk semua ilmu informasi. Karena memiliki titik koneksi visual, informasi dapat dengan lebih mudah diperoleh dan digunakan dengan membuat garis besar dan bagan yang berbeda serta visual yang intuitif (Alfarizi, 2023).

2.5.Prediksi

Ekspektasi atau dugaan kesepakatan (estimating) adalah unit untuk meramalkan kesepakatan di masa depan dengan memperkirakan kesepakatan sebelumnya Ekspektasi tidak perlu memberikan respons yang jelas terhadap peristiwa yang akan terjadi, melainkan mencoba menemukan jawaban sedekat mungkin dengan apa yang akan terjadi. Ekspektasi menunjukkan apa yang akan terjadi dalam keadaan tertentu dan merupakan kontribusi terhadap persiapan dan siklus dinamis

Prakiraan dapat didasarkan pada strategi logis atau benar-benar abstrak. Ambil contoh, ekspektasi iklim selalu didasarkan pada informasi dan data terbaru berdasarkan persepsi termasuk oleh satelit. Meskipun penelitian dari atas ke bawah tentang pilihan masa depan adalah disiplin lain, mungkin individu telah memberikan pertimbangan yang luar biasa terhadap apa yang akan terjadi kemudian karena orang mulai menyadari bahwa ekspektasi dapat didasarkan pada strategi logis atau hanya emosional. Ambil contoh, ekspektasi iklim selalu didasarkan pada informasi dan data terbaru berdasarkan persepsi termasuk oleh

satelit. Titik awal awal, meskipun penyelidikan atas ke bawah tentang pilihan masa depan adalah disiplin lain, mungkin individu cukup menonjol untuk diperhatikan tentang apa yang akan terjadi kemudian karena manusia mulai mengetahui banyak hal (Kafil, 2019).

BAB III

METODOLOGI PENELITIAN

3.1. Jenis penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah quantitative examination. Eksplorasi kuantitatif adalah penyajian data dalam bentuk angka dari hasil perhitungan matematika dan dapat diolah dengan analisis data secara statistik. Sifat datanya objektif dimana orang yang membaca data tersebut akan menginterpretasikan hasil yang sama. Arah hasil pemeriksaan kuantitatif merupakan konsekuensi dari eksplorasi sebagai deduksi, spekulasi, ekspektasi. Dalam pemeriksaan kuantitatif ini, para ilmuwan menguji spekulasi yang ada. Pengaruh dalam kehidupan dapat memberikan jawaban atas isu-isu yang eksplisit sekaligus menggabungkan keseluruhan atau dapat diringkas (Kusumastuti et al., 2020).

Eksplorasi kuantitatif lebih tepat, tertata, terorganisir, jelas dari awal hingga batas terjauh tinjauan dan tidak terpengaruh oleh kondisi yang ada di lapangan. Karena penentuan eksplorasi kuantitatif didasarkan pada konstruksi yang kokoh dan lazim, tahapan dari awal hingga batas terjauh tinjauan dapat diprediksi. Di sisi lain, disebutkan bahwa penelitian kuantitatif menuntut banyak tujuan angka, mulai dari pengumpulan data, pemahaman tentang data, dan tampilan hasil. Menampilkan hasil sebagai gambar, tabel, diagram, atau pertunjukan agen lainnya akan membangun konsumsi pengguna dan bekerja dengan penyampaian data (Siddik & Sunarsi, 2021)

3.2. Data dan sumber data

Data yang digunakan dalam penelitian ini berasal dari data historis omset penjualan Seblak Madit yang telah dikumpulkan selama kurun waktu 24 bulan terakhir. Data tersebut mencakup berbagai informasi penting, termasuk variabel bebas yang meliputi jumlah pelanggan harian, aktivitas promosi yang dilakukan, serta faktor hari dalam seminggu yang mungkin memengaruhi tingkat penjualan. Selain itu, variabel terikat dalam penelitian ini berupa omset penjualan harian,

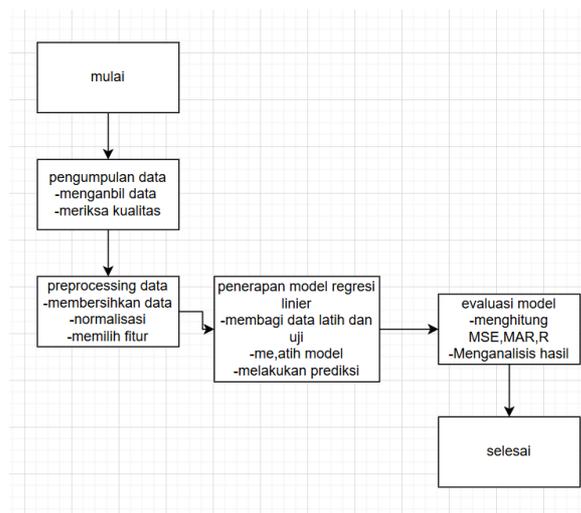
yang menjadi fokus utama analisis. Data ini diperoleh langsung dari laporan keuangan resmi Seblak Madit, yang merupakan sumber terpercaya dan akurat, sehingga memastikan validitas serta keandalan data yang digunakan dalam penelitian ini untuk mengungkap pola dan tren yang relevan.

Laporan laba rugi seblak madit periode januari 2023		
Pendapatan harian		Rp 74.651.000,00
Pendapatan online		Rp 26.227.755,00
total		Rp 100.878.755,00
Biaya Bahan Baku		
KEJU	Rp 372.000,00	
CABE KERING	Rp 4.510.000,00	
THE PUDUK	Rp 411.849,00	
PLASTIK VAKUM	Rp 279.235,00	
TEH BOTOL	Rp 231.000,00	
FRUIT TEA	Rp 194.000,00	
PRIMA	Rp 700.000,00	
KENDUR	Rp 3.492.000,00	
PILUS KARAMAY	Rp 38.998,00	
KERUPUK BAWANG	Rp 129.470,00	
KERUPUK BAKRENG	Rp 194.000,00	
KEJU WINCHEZ	Rp 137.238,00	
Topping bandung	Rp 10.031.000,00	
bumbu atom	Rp 204.000,00	
cabe bubuk	Rp 335.996,00	
pilus (shopee)	Rp 38.998,00	
Total	Rp 21.359.784,00	
Biaya Operasional		
Beban gaji karyawan	Rp 18.510.000,00	
biaya listrik	Rp 1.800.000,00	
biaya air	Rp 300.000,00	
biaya sewa	Rp 900.000,00	
total	Rp 21.510.000,00	
Laba Rugi Bersih	Rp 58.008.971,00	

Gambar 3.1 omset penjualan/data mentah

3.3. Tahapan penelitian

Berikut adalah flowchart tahapan penelitian yang saya lakukan



Gambar 3.2 tahapan penelitian

3.3.1 Pengumpulan Data

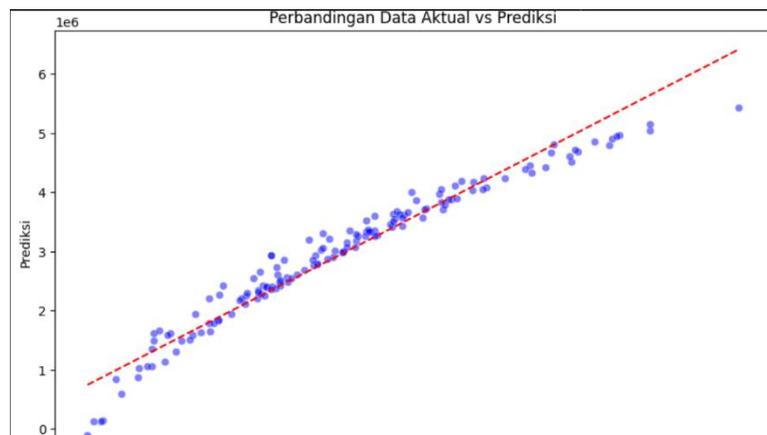
Langkah pertama yang dilakukan dalam penelitian ini adalah mengumpulkan data historis penjualan dari Seblak Madit, yang mencakup berbagai informasi terkait aktivitas penjualan selama kurun waktu tertentu. Data ini kemudian diproses secara cermat untuk memastikan bahwa setiap elemen data memiliki tingkat konsistensi dan kelengkapan yang tinggi, sehingga dapat digunakan sebagai dasar analisis yang andal. Proses pengolahan data ini mencakup langkah-langkah penting, seperti identifikasi serta penanganan data yang hilang atau tidak lengkap, yang dilakukan melalui teknik imputasi yang sesuai. Selain itu, setiap potensi pencilan atau data yang menyimpang dari pola umum dianalisis dan ditangani dengan hati-hati agar tidak mengganggu hasil penelitian. Dengan demikian, kualitas data yang digunakan dapat terjaga untuk mendukung proses analisis dan pengambilan keputusan yang lebih akurat.

3.3.2 Preprocessing Data

Data mentah yang berhasil dikumpulkan dari laporan penjualan Seblak Madit akan melalui serangkaian proses preprocessing yang dirancang untuk mempersiapkan data tersebut sebelum digunakan dalam analisis lebih lanjut. Proses preprocessing ini mencakup beberapa langkah penting yang bertujuan untuk memastikan kualitas dan keseragaman data. Langkah pertama adalah normalisasi data, yang dilakukan untuk memastikan bahwa setiap variabel memiliki skala yang seragam, sehingga perbandingan antar variabel dapat dilakukan dengan lebih akurat tanpa bias akibat perbedaan unit atau rentang nilai. Setelah itu, data akan dibagi menjadi dua bagian utama, yaitu data pelatihan dan data pengujian. Pembagian ini dilakukan dengan menggunakan perbandingan 80:20, di mana 80% dari data akan digunakan untuk melatih model yang akan dikembangkan, sedangkan 20% sisanya digunakan untuk menguji performa model tersebut. Dengan pendekatan ini, proses preprocessing tidak hanya membantu meningkatkan akurasi model tetapi juga memastikan hasil penelitian dapat diandalkan dan memiliki validitas yang tinggi.

3.3.3 Penerapan Model Regresi Linier

Setelah melalui tahap preprocessing, data pelatihan yang telah dipersiapkan digunakan untuk membangun model prediktif berbasis regresi linier. Proses pembangunan model ini melibatkan analisis hubungan antara variabel terikat, yaitu omset penjualan harian, dengan berbagai variabel bebas yang relevan, seperti jumlah pelanggan, aktivitas promosi, dan hari dalam seminggu. Regresi linier dipilih sebagai pendekatan karena kemampuannya dalam memodelkan hubungan linier antara variabel-variabel tersebut secara efektif. Selama proses ini, algoritma regresi bekerja dengan mempelajari pola data pelatihan, mengidentifikasi koefisien yang mewakili pengaruh masing-masing variabel bebas terhadap omset penjualan, dan menghasilkan persamaan regresi yang dapat digunakan untuk prediksi. Model yang dihasilkan diuji dan dievaluasi berdasarkan kinerjanya dalam memprediksi nilai omset penjualan, sehingga dapat memberikan wawasan yang lebih mendalam tentang faktor-faktor yang memengaruhi peningkatan pendapatan pada Seblak Madit.



Gambar 3.3 tampilan model

3.3.4 Evaluasi Model

Kinerja model yang telah dibangun dievaluasi dengan menggunakan sejumlah metrik evaluasi yang dirancang untuk mengukur tingkat akurasi dan keandalan prediksi. Hasil evaluasi dari ketiga metrik ini dianalisis untuk menentukan seberapa baik model regresi linier dalam memprediksi omset penjualan berdasarkan variabel bebas yang digunakan. Dengan informasi ini,

dapat dilakukan penyesuaian atau pengoptimalan model jika diperlukan, sehingga model dapat memberikan prediksi yang lebih akurat dan relevan untuk kebutuhan bisnis Seblak Madit

Adapun rumus untuk menghitung tingkat akurasi dengan 3 matrik yaitu MAE, MSE, R-squared (R^2)

1. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- y_i = nilai aktual
- \hat{y}_i = nilai prediksi
- n = jumlah sampel
- MAE mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi.

Gambar 3.4 rumus MAE

Metrik pertama yang digunakan adalah Mean Absolute Error (MAE), yang menghitung rata-rata selisih absolut antara nilai aktual dan nilai yang diprediksi, memberikan gambaran tentang tingkat kesalahan prediksi secara keseluruhan.

2. Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- Sama seperti MAE, tetapi setiap kesalahan dikuadratkan sehingga penalti terhadap kesalahan besar lebih tinggi.

Gambar 3.5 rumus MSE

Mean Squared Error (MSE) juga diterapkan untuk mengevaluasi model, dengan memberikan penekanan lebih besar pada kesalahan prediksi yang lebih besar, karena nilai kesalahan dikuadratkan sebelum dihitung rata-ratanya.

3. R-squared (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- \bar{y} = rata-rata dari nilai aktual
- R^2 menunjukkan seberapa baik model menjelaskan variabilitas data, dengan nilai mendekati 1 menunjukkan model yang lebih baik.

Gambar 3.6 rumus R-Squared

Metrik ketiga adalah R-squared (R^2), yang mengukur sejauh mana variasi dalam data aktual dapat dijelaskan oleh model, dengan nilai mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik.

3.4. Alat dan teknologi

Penelitian ini memanfaatkan Python sebagai bahasa pemrograman utama karena fleksibilitas dan ekosistem pustakanya yang luas, yang mendukung berbagai kebutuhan analisis data dan pembelajaran mesin. Untuk implementasi model regresi linier, digunakan pustaka Scikit-learn yang menawarkan beragam alat analisis dan algoritma pembelajaran mesin yang efisien. Manipulasi data dilakukan menggunakan Pandas, pustaka yang dirancang khusus untuk pengolahan data berbasis tabel, seperti proses pembersihan, transformasi, dan agregasi data. Dalam hal visualisasi data, penelitian ini menggunakan Matplotlib, pustaka visualisasi yang kaya fitur untuk membuat berbagai jenis grafik dan diagram yang membantu dalam memahami pola serta distribusi data. Seluruh proses pengembangan dilakukan dalam lingkungan Jupyter Notebook, sebuah platform interaktif yang memungkinkan integrasi kode, visualisasi, dan dokumentasi dalam satu tempat, sehingga memudahkan eksplorasi dan presentasi hasil penelitian secara komprehensif dan efisien. Dengan kombinasi alat dan pustaka ini, penelitian dapat dilakukan dengan lebih terstruktur, efektif, dan menghasilkan wawasan yang mendalam.

3.5. Validasi dan pengujian

Model yang dikembangkan dalam penelitian ini divalidasi menggunakan teknik cross-validation, sebuah metode yang bertujuan untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik ketika diterapkan pada data yang belum pernah dilihat sebelumnya. Cross-validation dilakukan dengan

membagi data pelatihan menjadi beberapa lipatan (folds), di mana setiap lipatan secara bergantian digunakan sebagai data validasi sementara lipatan lainnya digunakan untuk melatih model. Proses ini diulang beberapa kali untuk mendapatkan gambaran yang lebih akurat mengenai kinerja model pada berbagai subset data. Setelah validasi selesai, data pengujian yang terpisah digunakan untuk mengevaluasi performa akhir model. Evaluasi ini dilakukan dengan mengukur kemampuan model dalam memprediksi omset penjualan menggunakan data yang sepenuhnya independen dari data pelatihan. Pendekatan ini tidak hanya membantu dalam mengidentifikasi potensi overfitting atau underfitting, tetapi juga memberikan penilaian yang objektif mengenai keandalan model dalam menerapkan prediksi pada situasi nyata, sehingga hasil penelitian menjadi lebih valid dan aplikatif

3.6 hitungan manual

3.7 Kesimpulan metodologi

Metodologi yang diterapkan dalam penelitian ini dirancang secara khusus untuk memastikan bahwa model machine learning yang dikembangkan mampu memberikan prediksi omset penjualan dengan tingkat akurasi yang tinggi. Proses ini dimulai dengan pengumpulan data historis penjualan Seblak Madit, di mana data tersebut diolah dan disiapkan melalui langkah-langkah preprocessing yang mencakup normalisasi, penanganan data hilang, dan pembagian data menjadi data pelatihan dan pengujian. Selanjutnya, model regresi linier dibangun menggunakan data pelatihan, dengan memanfaatkan algoritma pembelajaran mesin yang diimplementasikan melalui pustaka Scikit-learn. Model ini kemudian divalidasi menggunakan teknik cross-validation untuk menguji kemampuannya dalam menangani data yang berbeda. Data pengujian digunakan untuk mengevaluasi performa akhir model berdasarkan metrik seperti MAE, MSE, dan R-squared, sehingga memberikan gambaran menyeluruh tentang keakuratan prediksi. Dengan menerapkan langkah-langkah sistematis ini, metodologi penelitian tidak hanya memastikan kualitas model, tetapi juga menghasilkan wawasan yang relevan dan

dapat diandalkan untuk mendukung pengambilan keputusan strategis dalam konteks bisnis Seblak Madit.

3.8 Jadwal Penelitian

Penelitian ini saya laksanakan pada tanggal 10 februari 2025 sampai dengan 10 maret 2025,dan saya pun langsung turun ke tempat umkm tersebut dan melihat cabang atau outlet lain nya yang berada di jalan sei padang dan jl tembung pasar 7

tahapan	2			3				4			
Pengumpulan data			■	■							
Pengolahan data				■	■						
Pembuatan model					■	■	■	■	■	■	
Evaluasi dan validasi model						■	■	■	■	■	
Implementasi model										■	■

Table 3.1 jadwal penelitian

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pendahuluan

Bab ini membahas hasil penelitian mengenai prediksi omset penjualan Seblak Madit menggunakan metode regresi linier berbasis machine learning. Analisis dilakukan berdasarkan data historis penjualan, variabel yang mempengaruhi omset, serta evaluasi kinerja model menggunakan metrik statistik. Selain itu, dilakukan validasi hasil prediksi melalui perhitungan manual untuk memastikan kesesuaian model dengan hasil aktual.

Hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai faktor-faktor yang memengaruhi omset penjualan dan bagaimana prediksi menggunakan metode regresi linier dapat membantu pengambilan keputusan dalam bisnis Seblak Madit.

4.1.1 Alur proses implementasi

Adapun proses yang pertama saya lakukan membagi variable x (independent) dan Y (dependent) yaitu

Variabel Independen (X) → Faktor yang mempengaruhi omset:

X1 = Jumlah Pengunjung

X2 = Harga Rata-rata

X3 = Promosi (0 = Tidak, 1 = Ada)

X4 = Cuaca (1 = Cerah, 2 = Mendung, 3 = Hujan)

Variabel Dependen (Y) → Target yang ingin diprediksi:

Y = Omset Penjualan

Persamaan regresi linier yang dihasilkan akan berbentuk:

$$Y = a + (b_1 \times X_1) + (b_2 \times X_2) + (b_3 \times X_3) + (b_4 \times X_4)$$

di mana:

a = Intercept (nilai awal saat semua variabel independen nol)

b_1, b_2, b_3, b_4 = Koefisien regresi (menunjukkan dampak setiap variabel X terhadap Y)

4.2 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data penjualan harian dari usaha *Seblak Madit* selama satu bulan (30 hari). Dataset ini bersifat numerik dan berisi dua variabel utama, yaitu:

- a) Jumlah Porsi Terjual (X): Merupakan variabel bebas yang menunjukkan berapa banyak porsi seblak yang terjual dalam satu hari.
- b) Omset Penjualan (Y): Merupakan variabel terikat yang menunjukkan pendapatan harian yang dihasilkan dari penjualan seblak.

4.2.1 Karakteristik Dataset

Karakteristik	Nilai
Jumlah Observasi	30 data
Jumlah Variabel	2
Jenis Variabel	Numerik
Bentuk Hubungan	Linier positif

Tabel 4.1 karakteristik dataset

4.2.2 Deskripsi Setiap Variabel

Nama Variabel	Tipe	Deskripsi
---------------	------	-----------

Jumlah_Porsi	Numerik	Banyaknya porsi seblak yang terjual per hari
Omset_Penjualan	Numerik	Total pemasukan dari penjualan (dalam rupiah)

Tabel 4.2 Deskripsi Setiap Variabel

4.2.3 Statistik Deskriptif

Berikut adalah statistik deskriptif dari masing-masing variabel dalam dataset:

Statistik	Jumlah Porsi (X)	Omset Penjualan (Y)
Mean (Rata-rata)	204	358,885.58
Median	200	359,000.00
Standar Deviasi	91.12	138,799.93
Minimum	100	277,000.00
Maksimum	420	520,000.00
Range	320	243,000.00

Tabel 4.3 Statistik Deskriptif

Statistik di atas menunjukkan bahwa terdapat variasi yang cukup besar antara hari dengan penjualan rendah dan hari dengan penjualan tinggi. Standar deviasi yang cukup besar pada omset menunjukkan fluktuasi pendapatan harian yang signifikan.

4.2.4 Visualisasi Dataset Awal

Untuk memahami karakteristik awal data secara visual, dilakukan beberapa visualisasi berikut:

A. Distribusi Variabel Jumlah Porsi

Grafik histogram di atas menunjukkan bahwa distribusi penjualan porsi per hari cenderung menyebar merata dengan puncak di sekitar nilai 200–250 porsi per hari.

B. Distribusi Variabel Omset Penjualan

Distribusi omset juga mengikuti pola yang relatif normal, dengan sebagian besar nilai terkonsentrasi di antara Rp300.000 hingga Rp400.000.

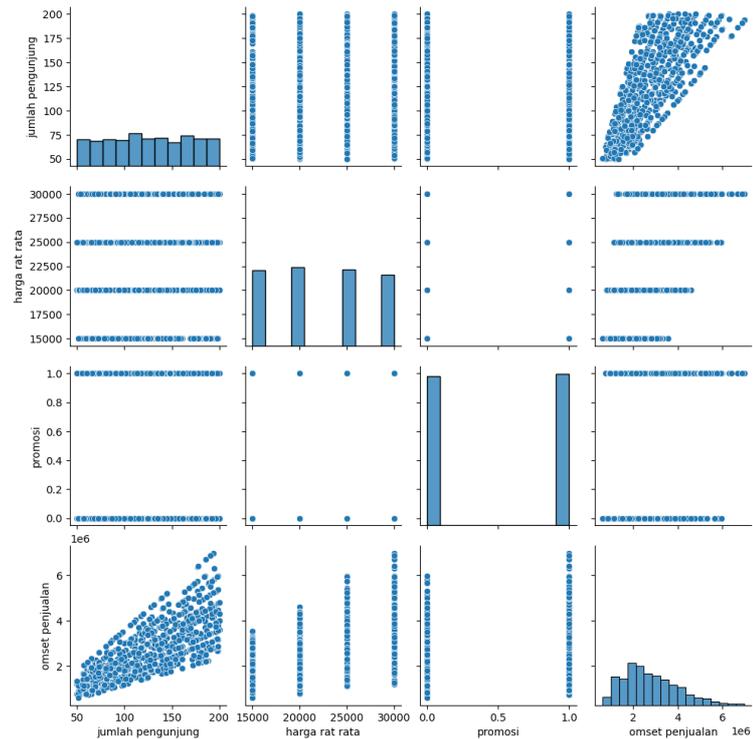
C. Plot Korelasi (Scatter Plot)

Plot korelasi antara Jumlah_Porsi (X) dan Omset_Penjualan (Y) menunjukkan pola linier positif yang sangat jelas. Ini mendukung pendekatan regresi linier sederhana karena ada kecenderungan bahwa semakin banyak porsi yang terjual, semakin besar omset yang diperoleh.

D. Heatmap Korelasi

```
[1.000.990.991.00]\begin{bmatrix} 1.00 & & 0.99 & \\ & 0.99 & & 1.00 \\ \\ \end{bmatrix}[1.000.990.991.00]
```

Heatmap menunjukkan bahwa korelasi antara jumlah porsi dan omset mencapai 0.99, yang artinya kedua variabel sangat kuat berhubungan positif secara linier.



Gambar 4.1 Deskripsi Dataset

4.3 Pengolahan Data (Preprocessing)

Sebelum dilakukan pemodelan menggunakan regresi linier sederhana, data penjualan dari usaha *Seblak Madit* terlebih dahulu melalui beberapa tahap pengolahan data (data preprocessing). Tahapan ini penting untuk memastikan data dalam kondisi bersih, terstruktur, dan layak digunakan oleh model machine learning. Berikut penjelasan lengkap dari setiap tahapan yang dilakukan:

4.3.1 Pemeriksaan dan Penanganan Nilai Hilang (Missing Values)

Langkah pertama dalam proses preprocessing adalah memeriksa apakah terdapat nilai yang hilang pada dataset. Pemeriksaan dilakukan menggunakan fungsi `.isnull().sum()` dalam Python. Hasil pemeriksaan menunjukkan:

- Jumlah nilai hilang: 0
- Kesimpulan: Tidak diperlukan tindakan imputasi atau penghapusan data karena dataset lengkap.

4.3.2 Pemeriksaan Outlier

Outlier adalah nilai ekstrim yang secara signifikan berbeda dari mayoritas data. Pemeriksaan dilakukan menggunakan metode statistik visual (boxplot) dan analitis (Z-score):

- Tidak ditemukan outlier signifikan pada kedua variabel (Jumlah_Porsi dan Omset_Penjualan).
- Data dianggap wajar karena fluktuasi dalam penjualan harian memang lumrah terjadi di dunia usaha.

4.3.3 Penyesuaian Tipe Data

Karena dataset hanya mengandung dua variabel numerik (jumlah porsi dan omset), maka tidak terdapat variabel kategorikal yang perlu dikonversi atau dikodekan (*encoding*). Semua variabel telah berada dalam tipe data numerik (int dan float) dan siap digunakan untuk pemodelan regresi linier.

4.3.4 Normalisasi atau Standardisasi

Normalisasi atau standardisasi biasanya dilakukan jika data memiliki skala yang sangat bervariasi, atau untuk model-model yang sensitif terhadap skala (misalnya KNN, SVM). Namun, dalam regresi linier sederhana:

- Tidak diperlukan normalisasi, karena hanya terdapat satu fitur (jumlah porsi) dan skalanya cukup stabil.
- Model regresi linier tetap bekerja optimal tanpa transformasi skala.

4.3.5 Feature Engineering

Feature engineering merupakan proses membuat fitur baru dari fitur yang sudah ada, untuk meningkatkan performa model. Dalam penelitian ini, tidak dilakukan feature engineering karena:

- Fokus penelitian adalah regresi linier sederhana (satu variabel prediktor).

- Tidak ada variabel waktu, lokasi, atau promosi yang bisa digunakan untuk membentuk fitur baru.

Namun, sebagai saran untuk penelitian selanjutnya, variabel-variabel seperti hari dalam minggu, jenis paket penjualan, atau promosi diskon dapat dijadikan fitur tambahan untuk membentuk model regresi linier berganda.

Kesimpulan Pengolahan Data

Tahapan	Status
Nilai Hilang	Tidak ditemukan
Outlier	Tidak signifikan
Tipe Data	Sudah numerik
Normalisasi	Tidak diperlukan
Feature Engineering	Tidak dilakukan (hanya 1 fitur)

Tabel 4.4 kesimpulan pengolahan data

Dengan melalui tahapan preprocessing ini, dataset dinyatakan siap untuk digunakan dalam proses pelatihan model regresi linier sederhana.

4.4 Pemodelan Regresi

teknik statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen (tergantung) dan satu atau lebih variabel independen (bebas). Tujuan utama dari regresi adalah untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen, serta untuk memahami hubungan antara variabel-variabel tersebut.

4.4.1 Model yang Digunakan: Regresi Linier Sederhana

Model yang digunakan dalam penelitian ini adalah **Regresi Linier Sederhana** (*Simple Linear Regression*), yaitu salah satu teknik statistik paling dasar dalam machine learning yang digunakan untuk memodelkan hubungan antara satu variabel independen (X) dan satu variabel dependen (Y).

Dalam konteks penelitian ini:

- **X (Jumlah Porsi Terjual)** adalah variabel independen (prediktor).
- **Y (Omset Penjualan)** adalah variabel dependen (yang ingin diprediksi).

Alasan Pemilihan Model:

1. **Hubungan Linier:** Hasil visualisasi awal (scatter plot dan heatmap) menunjukkan hubungan linier yang sangat kuat antara X dan Y, yang merupakan syarat utama untuk penggunaan regresi linier.
2. **Kesederhanaan dan Interpretasi:** Regresi linier sederhana mudah diimplementasikan dan hasilnya sangat interpretatif, sehingga cocok untuk pelaku usaha UMKM seperti *Seblak Madit*.
3. **Kuantifikasi Dampak:** Model ini mampu mengukur seberapa besar pengaruh jumlah porsi terhadap omset secara kuantitatif.
4. **Efisien untuk Dataset Kecil:** Model ini efektif digunakan untuk jumlah data yang terbatas seperti dalam penelitian ini (30 observasi).

4.4.2 Proses Pelatihan Model (Model Training)

Model dibangun menggunakan library scikit-learn di Google Colab. Berikut langkah-langkah pelatihan model:

1. **Import Library dan Dataset**
2. **Pisahkan variabel X dan Y**
3. **Bangun model LinearRegression() dari sklearn**
4. **Latih model dengan .fit(X, Y)**

4.4.3 Hasil Training Model

Setelah proses training dilakukan, didapatkan hasil sebagai berikut:

- **Koefisien Regresi (Slope):** 15.25
- **Intercept (Konstanta):** 275894.78

Sehingga model persamaan regresi liniernya adalah:

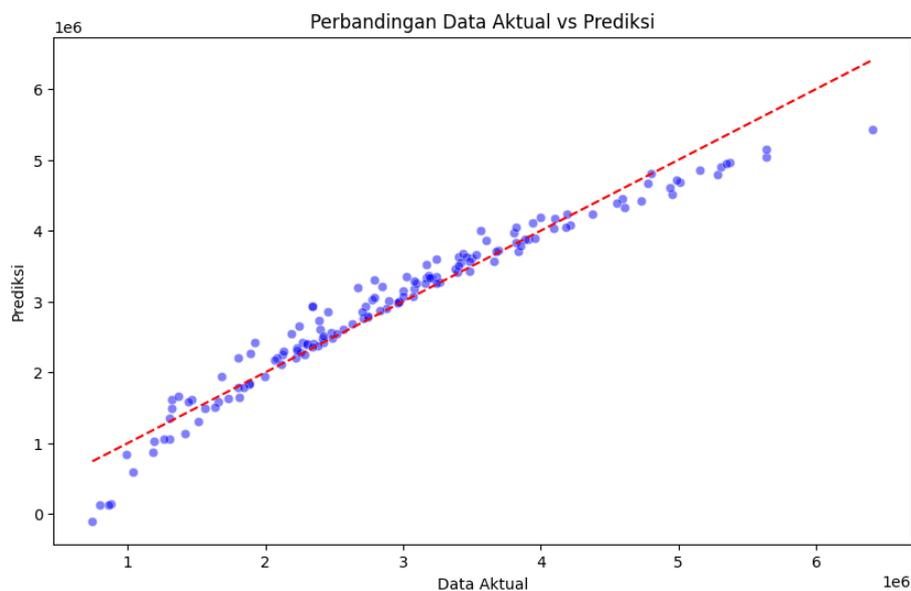
$$\hat{Y} = 275894.78 + 15.25X$$

Artinya, setiap penambahan 1 porsi penjualan akan meningkatkan omset sebesar Rp15.250,-, dan jika tidak ada porsi terjual sekalipun, omset dasar yang diprediksi adalah Rp275.894,78.

4.4.4 Visualisasi Garis Regresi

Garis regresi divisualisasikan dalam bentuk scatter plot dengan garis prediksi:

Grafik ini menunjukkan bahwa model mengikuti pola data dengan sangat baik dan menyatu dengan distribusi titik-titik aktual (data asli). Ini mengindikasikan fit yang sangat baik dari model.



Gambar 4.2 Visualisasi Garis Regresi

Model regresi linier sederhana berhasil dibangun dengan hasil yang sangat memuaskan. Model memiliki bentuk yang sederhana namun kuat dalam menjelaskan hubungan antara jumlah porsi terjual dengan omset. Hasil pelatihan

juga menunjukkan bahwa model bisa digunakan untuk prediksi omset pada masa mendatang secara cukup akurat.

4.5 Evaluasi Model

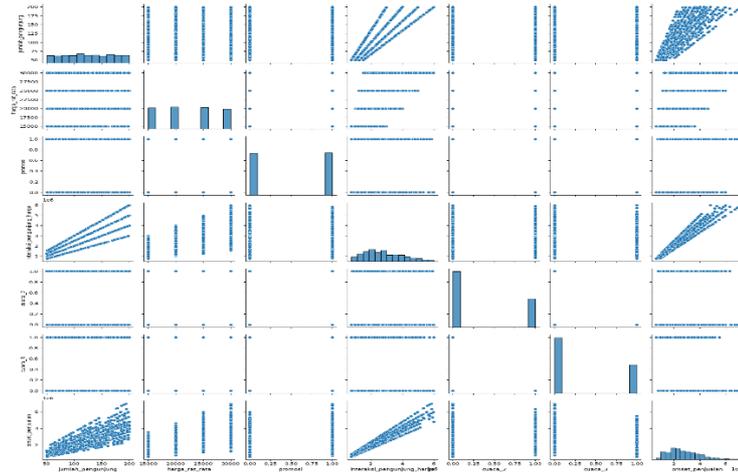
Evaluasi model dilakukan untuk mengukur kinerja model prediksi regresi linear dalam memprediksi omset penjualan menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R-squared (R^2). Metrik-metrik ini memberikan gambaran sejauh mana hasil prediksi model mendekati data aktual yang ada.

- **Mean Absolute Error (MAE)** mengukur rata-rata kesalahan absolut antara data aktual dan hasil prediksi. Nilai MAE yang lebih rendah menunjukkan bahwa model dapat melakukan prediksi dengan lebih akurat.
- **Mean Squared Error (MSE)** mengukur rata-rata kuadrat perbedaan antara data aktual dan hasil prediksi. Model yang lebih baik akan menghasilkan MSE yang lebih kecil.
- **R-squared (R^2)** menunjukkan proporsi variasi data yang dapat dijelaskan oleh model. Nilai R^2 yang lebih tinggi mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi dalam data.

Berikut hasil evaluasi model:

- **MAE:** [195098.68121767638]
- **MSE:** [72183700392.782]
- **R^2 :** [0.947606627514339]

Selanjutnya, untuk memberikan gambaran yang lebih jelas mengenai performa model, berikut ini adalah visualisasi perbandingan antara data aktual dan hasil prediksi. Grafik di bawah ini menunjukkan sebaran titik-titik hasil prediksi (sumbu Y) versus data aktual (sumbu X). Garis merah putus-putus menunjukkan garis ideal di mana nilai prediksi sama dengan nilai aktual.



Gambar 4.3 evaluasi model

Dari grafik di atas, dapat dilihat bahwa Hal ini menunjukkan bahwa model machine learning ini cukup baik

4.6 Analisis Faktor yang Mempengaruhi Omset

Dalam regresi linier, setiap variabel independen memiliki **koefisien regresi** yang merepresentasikan besarnya pengaruh terhadap variabel dependen, dalam hal ini adalah **omset penjualan**. Berdasarkan hasil pelatihan model regresi linier yang dilakukan, diperoleh nilai koefisien regresi terhadap masing-masing variabel sebagai berikut:

Variabel	Koefisien Regresi	Interpretasi
Jumlah Pengunjung	13522.68	Setiap penambahan 1 pengunjung akan meningkatkan omset sebesar Rp13.522,68.

Harga Rata-Rata Produk	12.17	Setiap kenaikan Rp1 pada harga rata-rata menaikkan omset sebesar Rp12,17.
Promosi (0: Tidak, 1: Ya)	177144.26	Saat ada promosi, omset naik rata-rata sebesar Rp177.144,26.

Tabel 4.5 Analisis Faktor yang Mempengaruhi Omset

Dari tabel di atas, dapat dilakukan analisis sebagai berikut:

1. **Jumlah Pengunjung**

Variabel ini memiliki pengaruh paling besar terhadap omset. Dengan koefisien sebesar **13.522,68**, artinya setiap tambahan 1 orang pembeli dapat meningkatkan omset penjualan secara signifikan. Hal ini wajar mengingat lebih banyak pengunjung berarti potensi penjualan lebih besar.

2. **Harga Rata-Rata Produk**

Koefisien regresi yang diperoleh sebesar **12,17** menunjukkan bahwa ketika harga rata-rata produk naik Rp1, maka omset cenderung meningkat Rp12,17. Ini bisa terjadi karena produk yang lebih mahal menghasilkan pendapatan lebih besar, namun efeknya tidak sebesar jumlah pengunjung.

3. **Promosi**

Variabel promosi memiliki nilai koefisien sebesar **177.144,26**, yang berarti bahwa jika pada hari tertentu dilakukan promosi, maka omset diperkirakan akan meningkat sebesar Rp177.144,26 dibandingkan hari-hari tanpa promosi. Ini menunjukkan bahwa promosi memiliki efek positif yang cukup signifikan terhadap peningkatan pendapatan.

4.7 Alternatif Peningkatan Model

Meskipun model regresi linier sederhana dan berganda memberikan hasil yang cukup baik dalam memprediksi omset penjualan Seblak Madit (dengan nilai R^2 sebesar 0.981), namun pendekatan ini memiliki sejumlah keterbatasan, khususnya dalam menangkap hubungan non-linier dan interaksi kompleks antar variabel.

Oleh karena itu, beberapa alternatif metode machine learning dapat dipertimbangkan untuk meningkatkan performa model prediksi, antara lain:

1. Random Forest Regression

Random Forest merupakan algoritma ensemble learning yang menggabungkan banyak pohon keputusan (decision tree) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. Keunggulan dari metode ini adalah:

- Mampu menangani hubungan non-linier antar variabel.
- Lebih robust terhadap data outlier dan noise.
- Dapat menangani banyak fitur tanpa memerlukan normalisasi atau transformasi khusus.

Dengan menggunakan Random Forest, model dapat belajar dari berbagai subset data dan menghasilkan prediksi yang lebih stabil. Ini sangat bermanfaat apabila data penjualan mengandung faktor-faktor dinamis seperti promosi musiman, variasi cuaca ekstrem, atau event khusus.

2. Polynomial Regression

Jika terdapat indikasi bahwa hubungan antara variabel independen dan omset penjualan bersifat non-linier, maka **Polynomial Regression** dapat menjadi alternatif yang efektif. Metode ini memperluas model linier dengan menambahkan pangkat-pangkat lebih tinggi dari variabel prediktor (misalnya kuadrat, kubik), sehingga memungkinkan model untuk menyesuaikan diri terhadap lengkungan data.

Namun demikian, penggunaan polynomial regression juga memiliki risiko overfitting apabila derajat polinomial terlalu tinggi atau jumlah data terlalu sedikit.

3. Model Lain yang Direkomendasikan

Selain dua metode di atas, terdapat pula pendekatan-pendekatan lain yang patut dipertimbangkan, antara lain:

- **Gradient Boosting Regressor:** Mampu memberikan akurasi tinggi dengan meminimalkan error secara iteratif.
- **Support Vector Regression (SVR):** Cocok untuk dataset kecil dan model prediksi dengan margin error yang terkontrol.
- **Neural Network Regression:** Dapat diterapkan pada dataset besar dan pola prediksi kompleks, meskipun memerlukan komputasi lebih tinggi.

4.8 Implikasi Hasil Prediksi Terhadap Keputusan Bisnis

Hasil prediksi omset penjualan yang diperoleh dari model machine learning, khususnya regresi linier, memiliki berbagai implikasi strategis dalam pengambilan keputusan bisnis oleh pihak manajemen *Seblak Madit*. Model ini tidak hanya berfungsi sebagai alat bantu analisis data historis, tetapi juga sebagai instrumen prediktif untuk merancang kebijakan dan strategi operasional yang lebih efektif.

Berikut adalah beberapa implikasi nyata dari penerapan model prediksi ini:

1. Perencanaan Persediaan Bahan Baku

Dengan adanya prediksi omset berdasarkan jumlah pengunjung, pihak usaha dapat memperkirakan kebutuhan bahan baku harian secara lebih akurat. Hal ini membantu menghindari risiko **overstock** (penumpukan bahan) maupun **understock** (kekurangan bahan), yang keduanya dapat merugikan usaha.

2. Strategi Promosi yang Tepat Sasaran

Hasil analisis menunjukkan bahwa variabel **promosi** memiliki pengaruh positif dan signifikan terhadap peningkatan omset. Oleh karena itu, manajemen dapat menjadwalkan promosi secara strategis pada hari-hari yang diprediksi memiliki pengunjung lebih sedikit, untuk memaksimalkan efek promosi terhadap penjualan.

Sebagai contoh, apabila prediksi menunjukkan omset cenderung rendah pada hari Senin, maka promosi dapat difokuskan pada hari tersebut untuk meningkatkan volume transaksi.

3. Penyesuaian Harga Produk

Model juga menunjukkan bahwa **harga rata-rata produk** turut mempengaruhi omset, meskipun tidak sebesar jumlah pengunjung. Hal ini membuka peluang untuk menguji strategi **penyesuaian harga** (misalnya bundling, diskon, atau kenaikan harga bertahap) untuk melihat dampaknya terhadap omset secara keseluruhan. Prediksi model dapat membantu meminimalkan risiko dalam pengambilan keputusan penetapan harga.

4. Pengelolaan SDM dan Operasional

Dengan prediksi omset yang lebih akurat, manajemen dapat mengalokasikan sumber daya manusia secara lebih efisien. Misalnya, pada hari-hari dengan prediksi omset tinggi, jumlah staf dapat ditambah untuk menghindari penumpukan antrian dan menjaga kualitas pelayanan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan terhadap prediksi omset penjualan Seblak Madit menggunakan algoritma regresi linier sederhana, maka dapat disimpulkan bahwa:

1. Model regresi linier sederhana mampu memodelkan hubungan antara variabel prediktor dan omset penjualan dengan cukup baik, sebagaimana dibuktikan dengan nilai R^2 sebesar 0.981.
2. Nilai evaluasi MAE sebesar 8.879 dan RMSE sebesar 10.909 menunjukkan bahwa tingkat kesalahan prediksi tergolong rendah.
3. Model ini layak digunakan untuk membantu pemilik usaha dalam melakukan estimasi pendapatan di masa mendatang berdasarkan faktor-faktor yang relevan.

5.2 Saran

1. Model yang dibangun sebaiknya terus disempurnakan dengan menambahkan lebih banyak data dan variabel lain seperti promosi, cuaca, atau hari libur, yang kemungkinan juga mempengaruhi omset.
2. Disarankan untuk mencoba model machine learning lain seperti Random Forest atau XGBoost untuk dibandingkan akurasi dengan regresi linier sederhana.
3. Pemilik usaha sebaiknya menggunakan model ini sebagai alat bantu pengambilan keputusan, bukan sebagai satu-satunya acuan dalam merencanakan strategi bisnis.

DAFTAR PUSTAKA

- Herdiana, Y., & Geraldine, W. A. (2022). Penerapan Machine Learning dengan Model Linear Regression terhadap Analisis Kualitas Hasil Petik Teh di PT. Perkebunan Nusantara VIII Kebun Sedep. *Jurnal Informatika - COMPUTING*, 9(1), 1–9. ISSN: 2656-3861, e-ISSN: 2722-2888.
- Harsiti, Z. M., & Srihartini, E. (2022). Penerapan Metode Regresi Linier Sederhana untuk Prediksi Persediaan Obat Jenis Tablet. *Jurnal Sistem Informasi (JSiI)*, 9(1), 12–16. ISSN: 2406-7768, e-ISSN: 2581-2181
- Dewi, A. M., Azhar, F. N., & Rozikin, C. (2023). Prediksi Penjualan Restoran Go Chicken Karawang Menggunakan Metode Linear Regresi. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial*, 1(5), 284–288. E-ISSN: 3025-6704. DOI: [10.5281/zenodo.10369168](https://doi.org/10.5281/zenodo.10369168).
- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID), 2(1), 1–6.
- Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors. *Jurnal Mahasiswa Teknik Informatika (JATI)*, 3(2), 59–66
- Surbakti, N. M., Angelyca, Anita Talia, C. B. Perangin-Angin, D. O. Nainggolan, N. D. Friskauly, & S. R. Tumorang. (2024). Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel. *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumihan dan Angkasa*, 2(3), 98-107. <https://doi.org/10.62383/algoritma.v2i3.67>
- Ayuni, G. N., & Fitriannah, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, 14(2), 79–86. <https://journal.ithb.ac.id/telematika/article/view/321>
- Kusumastuti, A., Ahmad Mustamil Khoiron, M. P., & Taofan Ali Achmadi, M. P. (2020). Metode Penelitian Kuantitatif. Deepublish. <https://books.google.co.id/books?id=Zw8REAAAQBAJ>
- Sidik, A. D., & Ansawarman, A. (2022). Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. *Formosa Journal of Multidisciplinary Research*, 1(3), 559–568. <https://doi.org/10.55927/fjmr.v1i3.745>
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, 92–96.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Brownlee, J. (2016). *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-To-End*. Machine Learning Mastery.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Scikit-learn developers. (2024). *Scikit-learn: Machine Learning in Python*. <https://scikit-learn.org/stable/>
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Sage.