

Penerapan Model Regresi Linier untuk Memprediksi Jumlah Pasien Baru pada Klinik XYZ

Nova Tampati

Teknik Informatika, Universitas Yos Soedarso; nova.tampati@gmail.com

Nur Kurniasari

Teknik Informatika, Universitas Yos Soedarso; nurkurniasari.nia@gmail.com

ABSTRACT

Linear Regression is a statistical technique used to model the relationship between one dependent variable that we want to predict with one or more independent variables used to make the prediction. This model assumes that the relationship between these variables can be approximated by a straight line. This study explains how the number of patients affects new patients at the XYZ clinic. The results of the analysis obtained produced a MAPE of 10%. This proves that the Linear Regression model gets a very small prediction error, an indication of a value of 10% to 20% of the MAPE value indicates that the accuracy of the model obtained is very good

Keywords: *Data Mining, Linear Regression, Machine Learning, Mean Absolute Percentage Error (MAPE)*

ABSTRAK

Regresi Linier adalah sebuah teknik statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen yang ingin kita prediksi dengan satu atau lebih variabel independen yang digunakan untuk melakukan prediksi. Model ini mengasumsikan bahwa hubungan antara variabel-variabel tersebut dapat didekati dengan sebuah garis lurus. Pada penelitian ini menjelaskan bagaimana pengaruh jumlah pasien terhadap pasien baru yang berada pada pada klinik XYZ. Dari hasil Analisa yang didapat menghasilkan MAPE sebesar 10%. Ini membuktikan model Regresi Linier tersebut mendapat kesalahan prediksi yang sangat kecil, indikasi nilai 10% sampai 20% nilai MAPE menandakan akurasi model yang didapat sangat baik.

Kata kunci: *Data Mining, Regresi Linier, Pembelajaran Mesin, Mean Absolute Percentage Error (MAPE)*

PENDAHULUAN

Latar Belakang

Klinik memiliki peran yang sangat penting dalam peningkatan kesehatan masyarakat. Sebagai ujung tombak pelayanan kesehatan tingkat pertama, klinik berperan dalam berbagai aspek, mulai dari pencegahan hingga pengobatan penyakit. Klinik merupakan salah satu pilar penting dalam sistem pelayanan kesehatan. Dengan berbagai peran yang dimilikinya, klinik berkontribusi secara signifikan dalam meningkatkan kesehatan masyarakat. Oleh karena itu, keberadaan dan pengembangan klinik perlu terus didukung untuk mencapai tujuan kesehatan masyarakat yang lebih baik.

Informasi jumlah pasien pada sebuah klinik memiliki peran yang sangat krusial, baik untuk kepentingan internal klinik maupun untuk analisis eksternal. Informasi jumlah pasien adalah aset berharga bagi sebuah klinik. Dengan memanfaatkan data ini secara efektif, klinik dapat meningkatkan kinerja, mengembangkan strategi bisnis yang tepat, dan memberikan pelayanan kesehatan yang lebih baik bagi Masyarakat. Dengan mempunyai data informasi jumlah pasien sebuah klinik dapat menganalisa untuk meningkatkan kualitas pelayanan. Indikator sebuah klinik dapat berkembang karena banyaknya pasien baru dang datang. Dengan semakin berkembangnya layanan kesehatan di masyarakat, menjadikan klinik untuk bersaing dalam meningkatkan kualitas layanan Kesehatan[1].

Management klinik membutuhkan informasi prediksi jumlah pasien yang akan datang berkunjung di klinik [2]. Prediksi jumlah pasien baru di sebuah klinik merupakan hal yang sangat krusial dalam perencanaan dan pengelolaan klinik. Dengan adanya prediksi yang akurat, klinik dapat mengambil langkah-langkah yang tepat untuk meningkatkan efisiensi, kualitas pelayanan, dan juga pendapatan.

Di dalam Data Mining terdapat beberapa metode untuk memprediksi jumlah pasien yang umum diterapkan, diantaranya adalah metode Regresi Linear. Metode ini kerap digunakan kerap digunakan dibidang lainnya.

Beberapa penelitian terdahulu menggunakan metode Regresi Linier menghasilkan nilai MAPE yang bervariasi. Pada tahun 2018 menghasilkan nilai MAPE 31% [3] Di tahun 2019 [4] menghasilkan prediksi dengan beberapa kriteria nilai error MAPE, dimana terdapat 26 model prediksi regresi linier yang memiliki nilai error kurang dari 20% artinya mempunyai akurasi sebesar 80%. Di tahun 2020 peneliti mendapatkan nilai MAPE yang layak sebesar 31,6% [5].

Berdasarkan literatur yang ada, penelitian ini bertujuan untuk menganalisa jumlah pasien yang klinik XYZ yaitu dengan memprediksi jumlah pasien dengan menggunakan model Regresi Linier dengan harapan hasil Analisa pada penelitian ini dapat dimanfaatkan oleh pihak klinik sehingga klinik dapat meningkatkan pelayanannya.

METODE

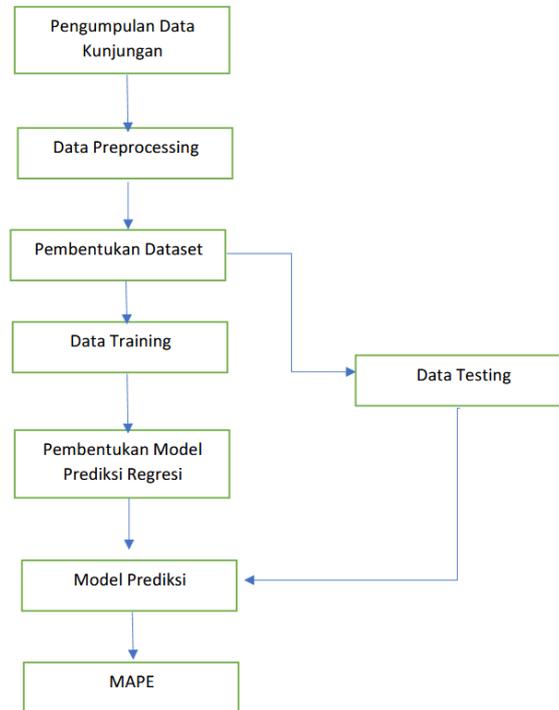
Dalam penelitian ini menggunakan pemrograman Python dengan Library untuk Analisa data. Python sangat baik untuk Analisa data karena sintak yang digunakan sangat mudah, Python tidak terlalu rumit, sehingga Anda bisa cepat mulai melakukan analisis data. Data diambil dari jumlah pasien yang datang dan pasien baru pada tahun 2020 sampai 2023. Sebelum data diproses ke sistem, data harus mengalami tahapan preprocessing data. Data Preprocessing bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya [6]. Dimana preprocessing data menyiapkan data mentah dan berguna dalam memperbaiki presisi maupun kinerja data sehingga menghasilkan hasil analisa yang lebih akurat. Dengan adanya proses preprocessing data maka data yang digunakan semakin mudah diolah.

Tabel 1. Contoh Dataset yang digunakan

Bulan	Total Pasien	Jumlah Pasien baru	Bulan	Total Pasien	Jumlah Pasien baru
Jan-20	808	118	Jan-21	767	110
Feb-20	763	116	Feb-21	724	120
Mar-20	469	74	Mar-21	909	158
Apr-20	363	63	Apr-21	879	141
May-20	518	85	May-21	797	107
Jun-20	678	98	Jun-21	861	116
Jul-20	679	88	Jul-21	515	76
Aug-20	714	123	Aug-21	693	86
Sep-20	693	134	Sep-21	752	90
Oct-20	720	126	Oct-21	792	121
Nov-20	733	128	Nov-21	615	91
Dec-20	701	92	Dec-21	541	69

Pada penelitian ini menggunakan data primer yaitu data kunjungan dari klinik XYZ yang sudah tercatat di Sistem Informasi Manajemen milik klinik XYZ. Pada proses pembentukan dataset. Data yang digunakan yaitu jumlah pasien tiap bulan dan jumlah pasien baru tiap bulan yang dilakukan selama 4 tahun dari data 2020 sampai dengan data 2023. Data tersebut akan dipecah 80% data digunakan untuk data training dan 20% data digunakan untuk data testing. Data yang sudah dipisah digunakan untuk membuat model Regresi Linier. Data pasien klinik XYZ meliputi data pasien rawat jalan yang sudah tercatat pada sistem informasi manajemen.

Setelah dataset sudah terbentuk maka dataset tersebut digunakan untuk membentuk model prediksi. Untuk melihat tingkat kesalahan prediksi. Penelitian ini menggunakan MAPE (*Mean Absolute Percentage Error*) sebagai tolak ukur untuk kecukupan model yang didapat. MAPE mudah diimplementasikan untuk melihat nilai prediksi. Berikut ini adalah alur proses prediksi menggunakan Regresi Linier.



Gambar 1. Alur Penelitian

A. Regresi Linier

Regresi linier adalah sebuah metode statistik yang digunakan untuk memodelkan hubungan linier antara satu variabel dependen (Y) dengan satu atau lebih variabel independen (X). Sederhananya, regresi linier mencoba mencari garis lurus terbaik yang dapat menggambarkan hubungan antara variabel-variabel tersebut. Untuk persamaannya dapat ditulis sebagai berikut.

$$Y = a + bX$$

Y: Variabel dependen (yang ingin kita prediksi)

X: Variabel independen (yang digunakan untuk memprediksi Y)

a: Intersep (nilai Y ketika X = 0)

b: Koefisien regresi (menunjukkan seberapa besar perubahan Y ketika X berubah satu satuan)

B. Data Data Preprocessing

Data Preprocessing adalah proses persiapan data proses mengubah data mentah menjadi bentuk yang lebih mudah dipahami dan diolah. Ini melibatkan pembersihan, transformasi, dan integrasi data untuk meningkatkan kualitas dan kesesuaiannya dengan tugas analisis data[7]. Tujuan utama data preprocessing adalah meningkatkan kualitas data, mempersiapkan data untuk analisis dan meningkatkan akurasi model.

C. Data Training dan Data Testing

Pembentukan data training dan data testing sangat penting karena data training digunakan untuk mempelajari pola dan hubungan dalam data training untuk membuat prediksi atau klasifikasi pada model yang digunakan. Sedangkan data testing digunakan untuk membantu mengidentifikasi kekurangan dalam model dan mengoptimalkan kinerja.

D. MAPE

MAPE (*Mean Absolute Percentage Error*) adalah salah satu metrik yang umum digunakan untuk mengukur akurasi suatu model prediksi, terutama dalam bidang peramalan atau prediksi. MAPE merupakan rata-rata diferensiasi absolut antara nilai peramalan dan aktual, yang dinyatakan sebagai

presentase nilai actual [8]. MAPE membantu kita membandingkan kinerja berbagai model peramalan. Model dengan MAPE terendah umumnya dianggap sebagai model terbaik. Dalam bisnis, MAPE dapat digunakan untuk membuat keputusan penting seperti perencanaan produksi, pengelolaan inventori, dan strategi pemasaran. Untuk perhitungan MAPE dapat dicari dengan menggunakan rumus berikut.

$$MAPE = \frac{100}{n} \sum \frac{|At - Ft|}{At}$$

dimana:

At = nilai aktual pada data t

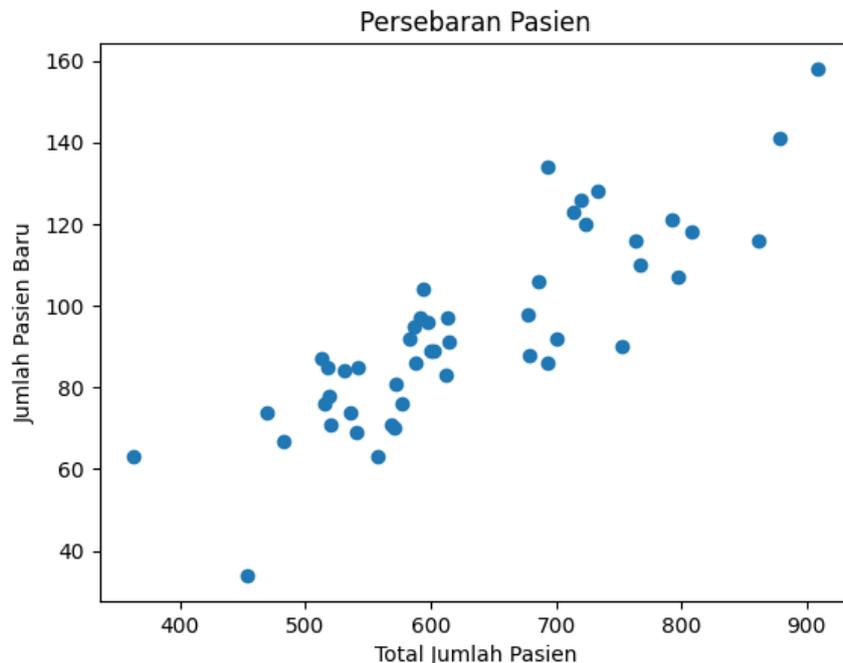
Ft = nilai peramalan pada data t

n = jumlah periode data

MAPE (*Mean Absolute Percentage Error*) memberikan hasil dalam bentuk persentase, sehingga mudah diinterpretasikan. MAPE mengukur kesalahan relatif terhadap nilai aktual, sehingga lebih relevan untuk data dengan skala yang sangat berbeda. MAPE adalah alat yang sangat berguna dalam mengevaluasi akurasi model peramalan. Nilai MAPE yang ideal adalah mendekati 0. Semakin kecil nilai MAPE, semakin akurat model peramalan. Kelebihan menggunakan MAPE antara lain mudah dipahami karena hasil yang digunakan dalam bentuk persentase.

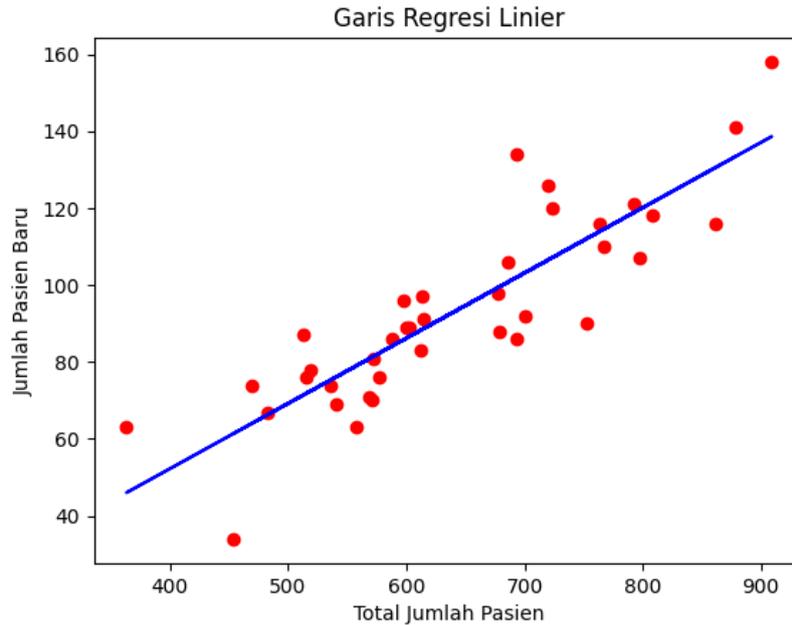
HASIL

Pada penelitian ini untuk menganalisa data menggunakan pemrograman python. Untuk melihat persebaran datanya, maka dilakukan visualisa dengan menggunakan library python yaitu library matplotlib. Data jumlah total pasien sebagai sumbu x dengan data pasien baru dengan sumbu y yang mempunyai peran yang sangat penting menggambarkan hubungan antar variabel. Untuk melihat persebaran data dapat dilihat pada gambar 2 berikut ini.



Gambar 2. Visualisasi persebaran data

Dari persebaran data tersebut dilakukan proses pemisahan data dengan menggunakan library python yaitu sklearn dengan memisahkan data training sebesar 80% dan data testing sebesar 20%. Data tersebut digunakan untuk membuat model Regresi Linier. Model Regresi Linier yang sudah terbentuk nantinya digunakan untuk memprediksi data baru.



Gambar 3. Visualisasi Garis Regresi Linier

Dari hasil Analisa model Regresi Linier yang didapat nilai dengan melihat kesalahan prediksi dengan menggunakan MAPE (*Mean Absolute Percentage Error*). Hasil prediksi dengan data testing didapat nilai MAPE sebesar 10%, Semakin kecil nilai MAPE maka semakin kecil kesalahan hasil prediksi, sebaliknya semakin besar nilai MAPE maka semakin besar kesalahan hasil prediksi. itu berarti dengan kisaran nilai antara 10% sampai 20% menandakan model yang didapat sangat baik. Model Regresi Linier yang didapat bagus dan digunakan untuk memprediksi data 5 bulan kedepan.

Tabel 2. Hasil Prediksi dengan Data Baru

No	Jumlah Total Pasien	Jumlah Pasien Baru
1	600	86
2	540	76
3	720	106
4	490	67
5	660	96

Pada table 2 jumlah total pasien digunakan untuk memprediksi jumlah pasien baru. Pada tabel adalah hasil prediksi dengan menggunakan data baru selama 5 bulan ke depan. Nilai yang diprediksi adalah nilai jumlah pasien baru selama 5 bulan kedepan. Jumlah total pasien mempunyai hubungan dengan jumlah pasien baru. Dengan adanya model Regresi Linier yang sangat baik makan klinik XYZ dapat memperkirakan jumlah pasien yang akan datan selanjutnya.

KESIMPULAN

Regresi linier adalah salah satu teknik statistik yang paling dasar namun sangat kuat dalam analisis data. Metode ini digunakan untuk memodelkan hubungan linear antara satu variabel dependen (variabel yang ingin kita prediksi) dengan satu atau lebih variabel independen (variabel prediktor). Banyak problem-problem yang ada masih bisa digunakan untuk model Regresi Linier. Pada penelitian ini dengan nilai MAPE yang didapat sebesar 10% berarti mengindikasikan model Regresi Linier masih cukup baik untuk memprediksi jumlah pasien baru pada klinik XYZ. Jadi nilai jumlah total pasien dengan jumlah pasien baru mempunyai korelasi.

DAFTAR PUSTAKA

- [1] A. Wijaya, P. Souldoni Akbar, H. Sarma Sangkot, and E. Sri Dewi Hastuti Suryandari, "Analisis Informasi Dan Pengembangan Dalam Penerapan Aplikasi Customer Relationship Management Klinik Sebagai Pendukung Edukasi Penyakit Tidak Menular," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, pp. 308–318, 2023, doi: 10.37792/jukanti.v6i2.1044.
- [2] A. Wibowo, D. Iskandar, and W. A. S. Wibowo, "Data Mining Dalam Prediksi Jumlah Pasien Dengan Regresi Linear Dan Exponential Smoothing," *J. Sist. Inf. dan Sains Teknol.*, vol. 5, no. 1, pp. 1–8, 2023.
- [3] M. W. Alauddin, M. Anshori, A. S. Wicaksono, and F. Utaminingrum, "Comparative Study based on Error Calculation in Multiple Linear Regression Coefficient for Forest Fires Prediction," *3rd Int. Conf. Sustain. Inf. Eng. Technol. SIET 2018 - Proc.*, pp. 115–120, 2018, doi: 10.1109/SIET.2018.8693226.
- [4] W. M. Baihaqi, M. Dianingrum, K. Aswin, and N. Ramadhan, "7067-23787-1-Pb," vol. 5, no. 2, pp. 86–93, 2019, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/coreit/article/view/7067>
- [5] X. Chen, M. Yi, and J. Huang, "Application of a PCA-ANN Based Cost Prediction Model for General Aviation Aircraft," *IEEE Access*, vol. 8, pp. 130124–130135, 2020, doi: 10.1109/ACCESS.2020.3008442.
- [6] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [7] A. Agung, A. Daniswara, I. Kadek, and D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," *J. Informatics Comput. Sci.*, vol. 05, pp. 97–100, 2023.
- [8] B. Putro, M. Tanzil Furqon, and S. H. Wijoyo, "Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing (Studi Kasus: PDAM Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4679–4686, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>