

# 2022 - A Comparison of Supervised (SIMANTEC)

*by* Heri Supriyanto

---

**Submission date:** 07-Jul-2022 11:06AM (UTC+0700)

**Submission ID:** 1867570398

**File name:** 22.\_Heri\_Supriyanto\_SIMANTEC\_-\_artikel.pdf (442.7K)

**Word count:** 3508

**Character count:** 21134

## PERBANDINGAN METODE *SUPERVISED LEARNING* UNTUK PERAMALAN *TIME SERIES* PADA KUNJUNGAN PASIEN RAWAT JALAN

### *A COMPARISON OF SUPERVISED LEARNING METHODS FOR FORECASTING TIME SERIES IN OUTSIDE PATIENT VISITS*

Heri Supriyanto<sup>1)</sup>

<sup>1</sup>Prodi Sistem Informasi, Fakultas Teknik dan Desain, Universitas Hayam Wuruk  
Perbanas

Jl. Wonorejo Utara No. 16, Rungkut, Surabaya

E-mail : <sup>1)</sup>[heri.supriyanto@hayamwuruk.ac.id](mailto:heri.supriyanto@hayamwuruk.ac.id)

#### ABSTRAK

Data merupakan sesuatu yang penting karena dapat digunakan untuk membantu mengambil sebuah keputusan atau kebijakan dalam sebuah organisasi. Salah satu bentuk keluaran dari pemanfaatan data adalah menghasilkan peramalan dimasa yang akan datang. Organisasi yang membutuhkan peramalan salah satunya adalah rumah sakit. Data yang dapat digunakan untuk peramalan adalah data kunjungan pasien. Tujuan dari penelitian ini membandingkan beberapa metode *supervised learning* pada kasus peramalan data kunjungan pasien rawat jalan, dengan menghasilkan model hasil dari proses eksperimen metode KNN, SVR, *Decision Tree*, *Random Forest* dan Regresi Linear. Dari data *Plot Time Series* kunjungan rawat jalan diperoleh hasil autokorelasi menggunakan metode ACF yang memiliki tingkat signifikansi yaitu pada lag 1 = 0,797, lag 2 = 0,688, dan lag 3 = 0,579. Sehingga pembentukan *dataset* didapatkan bahwa pada data sekarang ( $Y_t$ ) memiliki pengaruh terhadap data satu bulan sebelumnya ( $X_{t-1}$ ), dua bulan sebelumnya ( $X_{t-2}$ ), dan tiga bulan sebelumnya ( $X_{t-3}$ ). Hasil model peramalan yang dilakukan menghasilkan bahwa metode *random forest* yang memiliki model terbaik dengan nilai evaluasi model RMSE sebesar 204,43 dan nilai MAPE sebesar 12%. Berdasarkan kriteria nilai MAPE model yang sudah dibuat memiliki kategori baik.

**Kata kunci :** ACF, MAPE, RMSE, *Supervised Learning*, *Time Series*.

#### ABSTRACT

*Data is something important because it can be used to help make a decision or policy in an organization. One form of output from the use of data is to produce forecasts in the future. One of the organizations that need forecasting is the hospital. Data that can be used for forecasting is patient visit data. The purpose of this study is to compare several supervised learning methods in the case of forecasting outpatient visit data, by producing a model result from the experimental process of KNN, SVR, Decision Tree, Random Forest and Linear Regression methods. From the Plot Time Series data for outpatient visits, autocorrelation results using the ACF method have a significance level, namely at lag 1 = 0.797, lag 2 = 0.6, and lag 3 = 0.579. So that the formation of the dataset is found that the current data ( $Y_t$ ) has an influence on the data one month earlier ( $X_{t-1}$ ), the previous two months ( $X_{t-2}$ ), and the previous three months ( $X_{t-3}$ ). The results of the forecasting model carried out resulted that the random forest method had the best model with an evaluation value of the RMSE model of 204.43 and the MAPE value of 12%. Based on the criteria for the MAPE value, the model that has been made has a good category.*

**Keywords:** ACF, MAPE, RMSE, *Supervised Learning*, *Time Series*.

## PENDAHULUAN

Pada era digital saat ini, data merupakan hal yang sangat penting karena dapat digunakan untuk membantu mengambil sebuah keputusan atau kebijakan dalam sebuah organisasi. Salah satu bentuk keluaran dari pemanfaatan data tersebut adalah dengan menghasilkan data peramalan dimasa yang akan datang. Peramalan ini penting untuk membantu sebuah organisasi dalam mengevaluasi dan memutuskan sebuah arah kebijakan yang tepat untuk masa yang akan datang.

Organisasi yang membutuhkan data peramalan salah satunya adalah fasilitas kesehatan seperti rumah sakit. Pada data rumah sakit, terdapat data medis yang dapat dibuat sebagai bahan pertimbangan dalam pengambilan sebuah kebijakan yaitu data kunjungan berobat pasien [1]. Di rumah sakit, data kunjungan pasien sering kali hanya sebagai data laporan yang dibuat pihak rekam medis sebagai laporan bulanan kepada Top Manajemen dan Kementerian Kesehatan dalam bentuk laporan rekapitulasi bulanan. Secara bisnis, data jumlah kunjungan pasien dapat memiliki pengaruh terhadap perencanaan rumah sakit dimasa mendatang [2].

Seperti pada rumah sakit pada kelas Tipe D Swasta, selain pemasukan yang berasal dari unit Apotek/Farmasi dan Rawat Inap, pemasukan mayoritas berasal dari unit pelayanan kesehatan kepada pasien khususnya dari instalasi rawat jalan [3]. Jumlah kunjungan pasien yang tidak menentu menjadi kendala bagi pihak rumah sakit dalam menentukan arah kebijakannya seperti manajemen fasilitas sarana dan prasarana dan sumber daya manusia yang dibutuhkan [2]. Berdasarkan beberapa kasus tersebut, maka perlu dilakukan prediksi/peramalan mengenai jumlah kunjungan pasien tersebut. Peramalan pada kasus ini bertujuan untuk dapat memberikan informasi penting terkait data kunjungan pasien pada masa yang akan datang sebagai bahan dalam perencanaan dan

pengambilan keputusan bagi manajemen rumah sakit.

Pada kasus peramalan kunjungan ini secara luas dapat diselesaikan dengan Metode Kuantitatif yaitu metode yang memodelkan data dengan fungsi matematis [4]. Salah satu cara yang digunakan yaitu dengan pendekatan analisis Regresi [5]. Analisis regresi sebagai pendekatan model statistika yang berfungsi untuk mencari keterkaitan variabel satu dengan variabel yang lain [6]. Selain itu dapat diproses menggunakan pendekatan *machine learning* yaitu dengan membuat model untuk melakukan proses pembelajaran data, yang salah satu jenis data tersebut adalah data deret waktu (*time series*) [7]. Data deret waktu ini didasarkan pada data masa lalu yang bisa difungsikan untuk memperkirakan data yang akan datang [8].

Dalam literatur, studi analisis deret waktu (*time series*) telah diterapkan di beberapa sektor bidang seperti analisis data ekonomi dan medis [9]. Beberapa penelitian dengan topik peramalan kunjungan harian pasien dengan *time series analysis* [9][10] dilakukan dengan pendekatan metode yang dapat menyelesaikan task regresi seperti *Support Vector Regression* (SVR), *Multi-Layer Perceptron Regression* (MLPR), *Radial Basis Function Regression* (RBF), dan Regresi Linear [5]. Dari hasil eksperimen perbandingan metode-metode tersebut diperoleh metode SVR memiliki performa model terbaik diukur dari nilai *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE) yang paling rendah. Metode task regresi lainnya yaitu dengan menggunakan kategori model non linear dengan perbandingan metode yaitu linear-SVR, Radial-SVR, Random Forest, dan AdaBoost [11]. Beberapa penelitian sebelumnya ini menunjukkan bahwa metode-metode data mining dapat melakukan proses peramalan dengan hasil yang cukup baik.

Berdasarkan kasus dan penelitian terdahulu, penelitian ini bertujuan untuk membandingkan dan menganalisis

beberapa metode task regresi (*supervised learning*) pada kasus peramalan data kunjungan pasien rawat jalan, dengan menghasilkan model hasil dari proses eksperimen metode kNN, SVR, Decision Tree, Random Forest dan Regresi Linear. Sedangkan untuk pengujian evaluasi model pada penelitian ini menggunakan *Mean Absolute Percentage Error (MAPE)* dan *Root Mean Squared Error (RMSE)*.

Data yang digunakan pada penelitian ini menggunakan data primer dengan jenis data deret waktu (*time series*) dengan durasi bulanan dimulai dari bulan Januari 2015 sampai dengan bulan Desember 2021.

### METODE

Pada penelitian ini terdapat beberapa tahapan dalam proses pembuatan model peramalan kunjungan rawat jalan.

Langkah pertama yaitu pengambilan data, pada penelitian ini adalah data primer dari salah satu rumah sakit swasta tipe Kelas D. Data peramalan pada penelitian ini berasal dari laporan rekapitulasi (RL 5.2 Kunjungan Rawat Jalan) yang dilaporkan bulanan kepada Kementerian Kesehatan melalui Aplikasi SIRS Online (Sistem Informasi Rumah Sakit) [12]. Data plot pada penelitian ini berbentuk data deret waktu (*time series*).

Langkah kedua yaitu melakukan seleksi fitur. Seleksi fitur pada penelitian ini bertujuan untuk mencari korelasi data sekarang dengan data sebelumnya, tujuannya adalah untuk menghasilkan dataset model regresi. Proses seleksi ini diproses dengan menggunakan metode *Autocorrelation Function (ACF)*. ACF dapat memberikan informasi tentang tren dari sekumpulan data historis yang mengacu pada tingkat korelasi variabel yang sama antara dua interval waktu yang berurutan. Nilai autokorelasi berkisar antara -1 sampai 1. Nilai antara -1 dan 0 menunjukkan autokorelasi negatif. Nilai antara 0 dan 1 menunjukkan autokorelasi positif [13].

Berikut ini persamaan dari metode ACF yang digunakan.

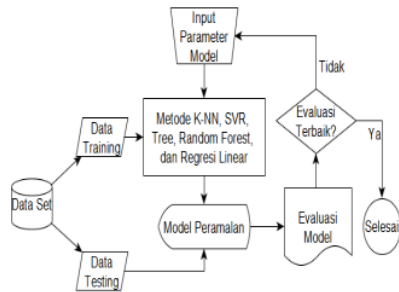
$$\text{Corr}(Y_t, Y_{t-k}), k = 1, 2, \dots \quad (1)$$

$$r_{Y_t Y_{t-k}} = \frac{\sum(Y_t - \bar{Y})(Y_{t-k} - \bar{Y}_{t-k})}{\sqrt{\sum(Y_t - \bar{Y})^2 \sum(Y_{t-k} - \bar{Y}_{t-k})^2}} \quad (2)$$

Nilai  $k$  merupakan jeda waktu (*time gap*) yang dipertimbangkan yang biasa disebut sebagai lag. Sebagai contoh bahwa Autokorelasi lag 1 (yaitu,  $k = 1$ ) adalah korelasi antara nilai-nilai yang terpisah pada satu periode waktu. Lebih umum, nilai  $k$  pada lag autokorelasi adalah korelasi antara nilai-nilai yang terpisah di periode waktu tertentu [14]. Berdasarkan persamaan yang dihasilkan diatas, maka menghasilkan plot data *time series* dengan membentuk variabel input dan output model regresi yang dibuat sebagai dataset pada kasus penelitian ini.

Langkah ketiga dari dataset yang dihasilkan, maka akan dibagi dua bagian yaitu sebagai data training sebesar 80% dengan sejumlah 65 data, sedangkan untuk data testing sebesar 20% yaitu dengan sejumlah 16 data. Pembentukan data training dimulai dari bulan April 2015 sampai dengan Agustus 2020, sedangkan data testing dari bulan September 2020 sampai dengan Desember 2021.

Langkah keempat adalah pembuatan model peramalan yang dilakukan dengan melakukan eksperimen dengan aplikasi Orange 3.26 [15] dengan tujuan memperoleh model terbaik dari metode-metode *supervised learning* yang dapat memproses task regression seperti KNN, SVR, Decision Tree, Random Forest, dan Regresi Linear. Gambar 1 berikut merupakan siklus proses eksperimen yang dilakukan.



Gambar 1. Siklus eksperimen model

Berdasarkan Gambar 1 diatas, pada data training setiap metode dilakukan uji parameter masukan, sampai pada data testing menghasilkan nilai evaluasi model terbaik yaitu dengan melihat nilai RMSE terkecil. Adapun penjelasan teori dari metode-metode *Supervised Learning* tersebut.

*K-Nearest Neighbor* (KNN) dapat digunakan untuk task klasifikasi maupun regresi. Pada task regresi metode ini biasa disebut *KNN Regression*. Prinsip kerja dari metode KNN Regresi adalah hasil regresi diperoleh dari rata-rata nilai *K* tetangga (*neighbor*) dengan jarak yang paling dekat, dimana output yang diprediksi biasanya dalam bentuk nilai kontinyu [16]. Untuk fungsi jarak yang digunakan pada eksperimen ini menggunakan *Chebyshev*. Berikut rumus [17] pada fungsi jarak tersebut.

$$d_c(i, j) = \max|x_i - x_j| \quad (3)$$

$d_c(i, j)$  adalah jarak Chebyshev dari data ke-*i* ke data ke-*j*. Untuk  $x_i$  adalah data ke-*i* sedangkan  $x_j$  adalah data ke-*j*.

*Support Vector Regression* (SVR) memiliki dua buah model yaitu klasifikasi dan regresi. Metode ini dapat digunakan untuk memperoleh data training terbaik dan memiliki performa yang cukup handal terkait penyelesaian kasus task regresi [18]. Pada metode SVR ini memiliki beberapa parameter yaitu *epsilon* ( $\epsilon$ ), *cost* (*C*), dan *gamma* ( $\gamma$ ) untuk menentukan nilai yang optimal [19]. Kemudian kernel yang pakai dalam model SVR ini menggunakan *Radial*

*Basis Function* (RBF). Fungsi kernel adalah untuk menemukan *hyperplane* yang optimal. Berikut persamaan [20] dari kernel RBF.

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \quad (4)$$

Nilai  $\gamma$  adalah parameter kernel, sedangkan  $x_i$  dan  $x$  mengacu pada vektor dari produk di ruang fitur.

*Decision Tree* adalah suatu metode klasifikasi yang dapat digunakan juga untuk regresi dengan menggunakan contoh pohon, yang dinyatakan sebagai *node* yang menggambarkan tiap atribut, Selanjutnya daun menggambarkan tiap kelas, dan setiap cabangnya memperlihatkan data nilai dari setiap kelas [21]. Konsep dari pohon keputusan ini adalah menerjemahkan suatu data menjadi pohon keputusan serta mempunyai aturan dalam penentuan keputusan [22].

*Random Forest* merupakan metode yang dapat digunakan untuk task klasifikasi dan regresi [23]. Konsep dari metode random forest ini adalah dengan membuat beberapa pohon keputusan yang berkaitan terhadap semua pohon keputusan yang berperan sebagai model *ensemble*. Pada setiap pohon keputusan menempatkan kelas peramalan dimana hasil model terbaik berdasarkan hasil yang paling maksimum [24]. Perhitungan dalam membangun pohon keputusan dengan informasi *gain* yang mengilustrasikan pada pemilihan atribut yang dipakai setiap *node*. Contoh nilai *N* merupakan *node* untuk membagi data disetiap kelas sesuai dengan atribut pada suatu data yang di inisialisasi sebagai *D*. maka pemecahan (*split*) *node* diproses berdasarkan atribut yang mempunyai nilai informasi *gain* tertinggi, berikut rumus untuk mencari nilai informasi *gain* [25].

$$Gain(A) = Info(D) - Info_A(D) \quad (5)$$

$$Info(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (6)$$

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (7)$$

Nilai  $Info(D)$  didapatkan berdasarkan persamaan (6) sedangkan untuk mendapatkan nilai  $Info_A(D)$  berdasarkan persamaan (7). Jika pada nilai informasi  $gain$  dengan bertipe kontinyu diperlukan nilai pemecahan (*split point*) terbaik untuk menghimpunkan nilai.

*Regresi Linear* merupakan metode yang dapat meramalkan kebutuhan masa depan berdasarkan data sebelumnya/masa lalu dengan melihat korelasi/keterkaitan antara variabel *Independent* (variabel yang mempengaruhi) dengan variabel *dependent* (variabel yang dipengaruhi) [6]. Pada penelitian ini menggunakan metode regresi linear berganda karena menggunakan 3 variabel independent. Berikut ini persamaan pada regresi linear berganda pada penelitian ini.

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 \quad (8)$$

$$A = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 & \sum X_1X_3 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 & \sum X_2X_3 \\ \sum X_3 & \sum X_1X_3 & \sum X_2X_3 & \sum X_3^2 \end{bmatrix} \quad (9)$$

$$H = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \\ \sum X_3Y \end{bmatrix} \quad (10)$$

$$a = \frac{\det A_1}{\det A} \quad (11)$$

$$b_1 = \frac{\det A_2}{\det A} \quad (12)$$

$$b_2 = \frac{\det A_3}{\det A} \quad (13)$$

$$b_3 = \frac{\det A_4}{\det A} \quad (14)$$

Dari persamaan diatas, untuk menghitung nilai  $a$ ,  $b_1$ ,  $b_2$ , dan  $b_3$  menggunakan persamaan menggunakan persamaan (11), (12), (13), (14).

Selanjutnya model peramalan yang dihasilkan berdasarkan hasil eksperimen, selanjutnya dilakukan evaluasi model menggunakan dua metode yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). RMSE adalah akar kuadrat dari *Mean Square Error* dari semua kesalahan [26]. Jika nilai RMSE semakin mendekati nilai 0 maka model dikatakan semakin baik. Sedangkan MAPE merupakan rata-rata absolut persentase antara nilai prediksi yaitu  $\hat{Y}_{t-1}$  untuk ramalan satu langkah kedepan dan nilai sebenarnya [27]. Kriteria nilai MAPE yaitu jika nilai <10% maka sangat baik, 10% - 20% baik, 20% - 50% cukup, dan 50% buruk [28]. Berikut ini adalah rumus dari evaluasi model RMSE dan MAPE yang di tampilkan pada persamaan 3 dan 4.

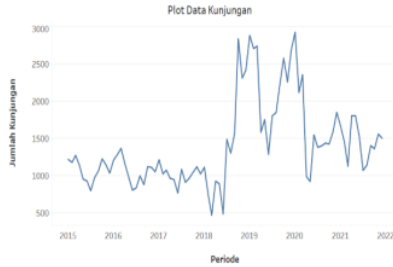
$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - Y'_t)^2} \quad (15)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - Y'_t}{Y_t} \right| \times 100\% \quad (16)$$

Nilai  $Y_t$  merupakan data aktual data asli dari periode  $t$ . Nilai  $Y'_t$  merupakan data hasil peramalan pada periode  $t$ , sedangkan nilai  $n$  adalah jumlah data atau banyaknya data yang terdapat pada dataset.

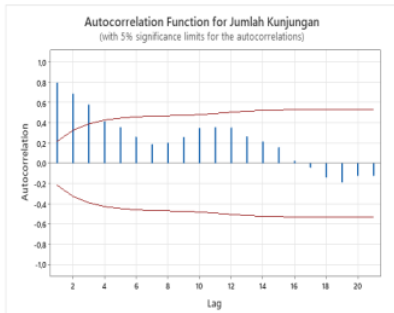
**HASIL DAN PEMBAHASAN**

Periode data yang digunakan dimulai dari bulan Januari 2015 sampai dengan bulan Desember 2021. Gambar 2 berikut ini merupakan data *plot time series* kunjungan pasien rawat jalan.



**Gambar 2.** Plot data *time series* kunjungan pasien rawat jalan

Berdasarkan data *time series* diatas menunjukkan data nonstasioner, artinya tidak memiliki rata-rata konstan disetiap periode. Sehingga diperlukan untuk membuat dataset dengan model regresi yaitu dengan mencari korelasi variabel dari data *time series* menggunakan metode ACM. Berikut hasil korelasi yang dihasilkan.



**Gambar 3.** Grafik ACF jumlah kunjungan

Berdasarkan Gambar 3 diatas, terdapat 3 lag yang keluar dari garis merah/selang kepercayaan (*confidence interval*) yang merupakan batas signifikansi dari autokorelasi. Berdasarkan grafik ACF tersebut, pada lag 1 memiliki nilai autokorelasi positif sebesar 0,797, kemudian pada lag 2 juga memiliki nilai autokorelasi positif sebesar 0,688, dan lag 3 sebesar 0,579, ketiga nilai lag tersebut memiliki nilai

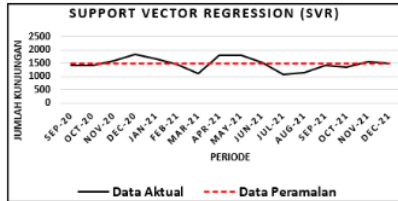
lebih besar dari nilai kepercayaan minimal yaitu 0,5 [29]. Hal ini menunjukkan berdasarkan data *plot time series* jumlah kunjungan rawat jalan bahwa data saat ini ( $Y_t$ ) memiliki pengaruh signifikan dengan data satu bulan sebelumnya ( $X_{t-1}$ ), kemudian juga memiliki pengaruh pada dua bulan sebelumnya ( $X_{t-2}$ ), dan 3 bulan sebelumnya ( $X_{t-3}$ ). Sehingga pembentukan dataset ini terdiri dari 3 variabel bebas (*independent*). Sebagai contoh jika pada data  $Y_t$  adalah Januari 2022, maka nilai variabel  $X_{t-1}$  adalah Desember 2021, selanjutnya pada variabel  $X_{t-2}$  adalah Nopember 2021, dan variabel  $X_{t-3}$  adalah Oktober 2021.

Pembentukan dataset ini selanjutnya dilakukan proses pemodelan peramalan dengan melakukan eksperimen dengan cara memasukkan parameter pada setiap model. Berikut ini masukan parameter model yang dilakukan.

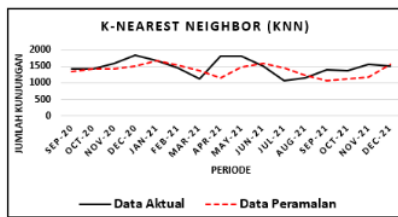
**Tabel 1.** Masukan/Input Parameter Model.

Model	Parameter
K-NN	<i>number of neighbors</i> = 19; <i>metric</i> = Chebyshev; <i>weight</i> = distance.
SVR	<i>kernel</i> = rbf; <i>cost</i> (C) = 0,10; <i>regression less epsilon</i> ( $\epsilon$ ) = 1,00; <i>iteration limit</i> = 12.
Tree	<i>Min. number of instances in leaves</i> : 1; <i>do not split subsets smaller than</i> = 66; <i>limit the maximal tree depth to</i> = 100.
Random Forest	<i>number of trees</i> = 17; <i>replicable training</i> = true; <i>growth control</i> ( <i>do not split subsets smaller than</i> = 43).
Regresi Linear	<i>No regularization</i> = true ( <i>multivariate linear regression</i> ).

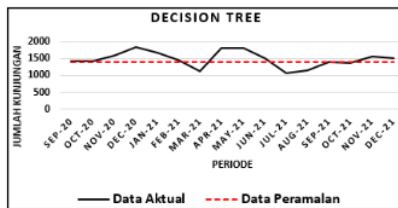
Parameter-parameter pada model tersebut menghasilkan peramalan dari masing-masing model yang digambarkan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 4 sampai gambar 8 di bawah ini.



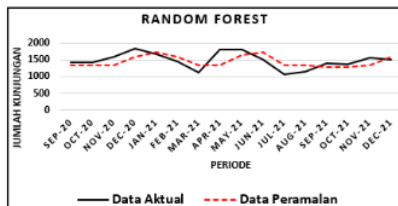
Gambar 4. Hasil Model Peramalan SVR



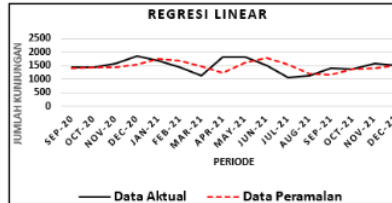
Gambar 5. Hasil Model Peramalan KNN



Gambar 6. Hasil Model Peramalan Decision Tree



Gambar 7. Hasil Model Peramalan Random Forest



Gambar 8. Hasil Model Peramalan Regresi Linear

Hasil peramalan berasal dari data testing yaitu dari bulan September 2020 sampai dengan Desember 2021 dengan sebanyak 16 baris. Berdasarkan hasil grafik menunjukkan pada data hasil peramalan setiap model memiliki hasil yang baik yang dapat ditunjukkan memiliki pola/trend yang sama dengan data aktual. Pada perhitungan manual metode regresi linear berganda yang digunakan diperoleh persamaan seperti berikut :

$$Y_t = 228 + 0,673 X_{t-1} + 0,213 X_{t-2} - 0,048 X_{t-3} \quad (17)$$

Dari hasil masing-masing model peramalan yang dihasilkan diatas, selanjutnya dilakukan evaluasi model dengan menggunakan metode RMSE dan MAPE yang ditunjukkan pada Tabel 2 dibawah ini.

Tabel 2. Evaluasi Model.

Model	RMSE	MAPE
K-NN	274,32	15%
SVR	228,94	13%
Tree	244,68	13%
Random Forest	204,43	12%
Regresi Linear	259,29	14%

Dari hasil evaluasi model tersebut, model terbaik adalah random forest dengan nilai RMSE terkecil yaitu 204,43 dan nilai MAPE 12%. Berdasarkan nilai tersebut maka model dikategorikan baik.

### SIMPULAN

Dari data plot *time series* kunjungan rawat jalan diperoleh hasil autokorelasi yang memiliki tingkat signifikansi yaitu pada lag 1 = 0,797, lag



$2 = 0,688$ , dan  $\text{lag } 3 = 0,579$ . Sehingga pembentukan dataset didapatkan bahwa pada data sekarang ( $Y_t$ ) memiliki pengaruh terhadap data satu bulan sebelumnya ( $X_{t-1}$ ), dua bulan sebelumnya ( $X_{t-2}$ ), dan tiga bulan sebelumnya ( $X_{t-3}$ ).

Hasil model peramalan yang dilakukan menghasilkan bahwa metode random forest yang memiliki model terbaik yang dilihat berdasarkan nilai evaluasi model RMSE sebesar 204,43 dan nilai MAPE sebesar 12%. Hal ini menunjukkan bahwa model yang sudah dibuat memiliki kategori baik.

### SARAN

Berdasarkan hasil evaluasi model bahwa nilai RMSE masih memiliki nilai yang cukup besar, pada penelitian selanjutnya bisa dilakukan pengembangan proses pembuatan model supaya mendapatkan nilai RMSE yang lebih kecil. Metode yang mungkin bisa digunakan untuk memperkecil nilai RMSE yaitu dengan pendekatan *Rolling Forecast*.

### DAFTAR PUSTAKA

- [1] D. M. Rani, B. N. Widyaningrum, N. Hasanah, "Analisis Trend Jumlah Kunjungan Pasien Saat Pandemi dengan Metode Trend Kuadrat Terkecil Di Rumah Sakit Panti Wilasa dr. Cipto Semarang". *Jurnal Rekam Medis dan Informasi Kesehatan Indonesia (Jurmiki)*, vol. 01, no.01, 2021.
- [2] M. Yunihati, Z. Basem, A. Setiawan, "Pengaruh Fasilitas Dan Tarif Pelayanan Terhadap Kunjungan Pasien Pada Klinik Dokter Junaidi Kecamatan Kampa". *Jurnal Riset Manajemen Indonesia*, vol. 2, no.4, 2020.
- [3] V. Andita, W. Hermawat, N. S. Hartati, "Pengaruh Jumlah Pelayanan Rawat Jalan, Rawat Igd Dan Rawat Inap Terhadap Tingkat Pendapatan Rumah Sakit Umum Daerah (RSUD) Cideres

Kabupaten Majalengka". *Jurnal Manajemen Sekolah Tinggi Ilmu Ekonomi Cirebon*, vol. 14, no.2, 2019.

- [4] Y. A. Ozcan, *Quantitative Methods in Health Care Management*, 2nd ed. San Fransisco: JohnWiley & Sons, 2009.
- [5] J. Thongkam, V. Sukmak, W. Mayusiri, "A comparison of regression analysis for predicting the daily number of anxiety-related outpatient visits with different time series data mining". *KKU Engineering Journal*, July – September 2015;42(3):243-249, 2015.
- [6] T. Khotimah dan R. Nindyasari, "Forecasting dengan metode regresi linier pada sistem penunjang keputusan untuk memprediksi jumlah penjualan batik (studi kasus: Kub Sarwo Endah Batik Tulis Lasem)". *J. Mantik Penusa*, vol. 1, no. 1, hlm. 71–92, 2017.
- [7] C. C. Anggarwal, *Data Mining The Textbook*, Springer International Publishing Switzerland, 2015.
- [8] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, 3rd ed. New York: Springer, 2011.
- [9] W-C. Juang, S-J. Huang, F-F Huang, et al, Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ Open* 2017;7:e018628. doi:10.1136/bmjopen-2017-018628. 2017.
- [10] H. Kinaci, M. G. Unsal, R. Kasap, "A close look at 2019 novel coronavirus (COVID 19) infections in Turkey using time series analysis & efficiency analysis". *Chaos, Solitons and Fractals*, <https://doi.org/10.1016/j.c>

- haos.2020.110583, 2021.
- [11] Y. Zhang, L. F. Z. R. Kong, J. Yang, et al, "Emergency patient flow forecasting in the radiology department. *Health Informatics Journal*, Article 1460458220901889, 2020.
- [12] Kemenkes, *Aplikasi RS Online*. Website: <https://sirs.kemkes.go.id/fo/>, diakses tanggal 27 Januari 2022.
- [13] C. F. Institute, *Autocorrelation*. Website: <https://corporatefinanceinstitute.com/resources/knowledge/other/autocorrelation/>, diakses tanggal 5 Februari 2022.
- [14] E. C. Science, *Applied Regression Modeling, 2nd edition*. Website: <https://online.stat.psu.edu/stat462/node/188/>, diakses tanggal 7 Februari 2022.
- [15] O. Center, *SpeedyWiki & BelajarWiki*, Website: <https://lms.onnocenter.or.id/wiki/index.php/Orange>, diakses tanggal 11 Februari 2021.
- [16] M. Cristian, "Average Monthly Rainfall Forecast in Romania by Using K-Nearest Neighbors Regression". *Annals of the „Constantin Brâncuși”*, 4, 5-12. 2018.
- [17] D. S. Seruni, M. T. Furqon, R. C. Wihandika, "Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression". *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 4, No. 4, hlm. 1075-1082, 2020.
- [18] L. Y Kurniawati, H. Tjandrasa, I. Arieshanti, Prediksi Pergerakan Harga Saham Menggunakan Support Vector Regression. *Jurnal Simantec*, Vol 4, No. 1, 2014.
- [19] C. W. Hsu, C. C. Chang, C. J. Lin, A Practical Guide to Support Vector Classification. Taipei : *Department of Computer Science National Taiwan University*. 2016.
- [20] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer. 1995.
- [21] Robianto, S. H. Sitorus, . U. Rustian, "Penerapan Metode Decision Tree Untuk Mengklasifikasikan Mutu Buah Jeruk Berdasarkan Fitur Warna Dan Ukuran". *Jurnal Komputer dan Aplikasi*. vol 09, No. 01, 2021.
- [22] A. Izyuddin, S. Wibisono "Jurnal Manajemen Informatika & Sistem Informasi. International". *Jurnal Manajemen Informatika & Sistem Informasi*, vol. 3, no. 2, 2020.
- [23] M. Rianto, R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest". *Paradigma*, vol. 23, no.1, 2021.
- [24] R. Kabir, F. B. Ashraf, F. R. Ajwad, "Analysis of Different Predicting Model for Online Shoppers". *International Conference on Computer and Information Technology (ICCIT)*. DOI:10.1109/ICCIT48885.2019.9038521. 2019.
- [25] M. L. Suliztia, "Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask", *S.Si. skripsi, Universitas Islam Indonesia*, Yogyakarta, 2020.
- [26] S. P. Neill, M. R. Hashemi. Ocean Modelling for Resource Characterization. *Fundamentals of Ocean Renewable Energy*, 193–235. doi:10.1016/b978-0-12-810448-4.00008-2, 2018.
- [27] D. C. Montgomery, C. L. Jennings, M. Kulahci. *Introduction To Time Series*

*Analysis And Forecasting.*  
Canada: John Wiley And Sons,  
Inc. 2015.

- [28] J. J. Montaña Moreno, A. Palmer Pol, A. Sesé Abad, and B. Cajal Blasco, "Using the R-MAPE index as a resistant measure of forecast accuracy," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013.
- [29] J.D. Cryer and K-S. Chan. *Time Series Analysis: With Application in R*, Second Edition. USA : Springer Science+Business Media, LLC . 2008.

# 2022 - A Comparison of Supervised (SIMANTEC)

---

ORIGINALITY REPORT

---

14%

SIMILARITY INDEX

13%

INTERNET SOURCES

4%

PUBLICATIONS

3%

STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

3%

★ id.scribd.com

Internet Source

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      On

# 2022 - A Comparison of Supervised (SIMANTEC)

---

GRADEMARK REPORT

---

FINAL GRADE

**/0**

GENERAL COMMENTS

**Instructor**

---

PAGE 1

---

PAGE 2

---

PAGE 3

---

PAGE 4

---

PAGE 5

---

PAGE 6

---

PAGE 7

---

PAGE 8

---

PAGE 9

---

PAGE 10

---