

**Peramalan Kebutuhan Obat Menggunakan XGBoost Studi Kasus pada Rumah Sakit XYZ****Muhammad Dzul Asmi Alhamdi<sup>1</sup>, Herman<sup>2</sup>, Wistiani Astuti<sup>3</sup>**<sup>1</sup>[alhamdidzul@gmail.com](mailto:alhamdidzul@gmail.com), <sup>2</sup>[herman@umi.ac.id](mailto:herman@umi.ac.id), <sup>3</sup>[wistiani.astuti@umi.ac.id](mailto:wistiani.astuti@umi.ac.id)

Fakultas Ilmu Komputer, Universitas Muslim Indonesia, Makassar, Indonesia

**Informasi Artikel**

Diterima : 7 Ags 2023  
Direview : 15 Sep 2023  
Disetujui : 28 Okt 2023

**Kata Kunci**

Machine Learning,  
Regresi,  
Peramalan Deret Waktu,  
XGBoost

**Abstrak**

Obat memiliki peran yang penting dalam pelayanan farmasi di rumah sakit, dari menyelamatkan nyawa hingga menyembuhkan pasien, namun perencanaan obat masih dilakukan secara manual menggunakan metode manual sehingga menghambat proses perencanaan obat, penelitian ini menggunakan XGBoost untuk melakukan peramalan *time series* pada penggunaan obat. Data yang digunakan adalah data perbulan penggunaan obat pada kategori *vital* dan *essential* dari tahun 2017 hingga 2022, penelitian ini menggunakan data cuaca sebagai fitur eksternal untuk membantu model bekerja. Hasil penelitian menunjukkan model XGBoost memiliki skor rata-rata skor RMSE dan MAE yang lebih rendah pada obat *vital* dibanding ketika melatih obat *essential*, sehingga model yang dilatih masih perlu perbaikan dalam menggunakan model XGBoost untuk meningkatkan performa model.

**Keywords**

Machine Learning,  
Regression,  
Time Series Forecasting,  
XGBoost,

**Abstrak**

*Pharmaceutical drugs play a crucial role in pharmaceutical services in hospitals, from saving lives to healing patients. However, drug planning is still done manually using manual methods, which hinders the drug planning process. This research utilizes XGBoost to perform time series forecasting on drug usage. The data used consists of monthly drug usage in the vital and essential categories from 2017 to 2022. The study incorporates weather data as an external feature to assist the model's performance. The research findings indicate that the XGBoost model yields lower average RMSE and MAE scores for vital drugs compared to when training on essential drugs. However, the model still requires improvement when using XGBoost to enhance its performance.*

## A. Pendahuluan

Perkembangan teknologi membawa manfaat besar di berbagai bidang, termasuk kesehatan di rumah sakit. Salah satu bidang yang penting untuk ditingkatkan adalah pelayanan farmasi, yang memiliki peran yang penting dalam merawat pasien. Ketersediaan obat yang tepat sesuai permintaan pasien menjadi salah satu faktor dalam memberikan pelayanan farmasi berkualitas. Namun, di Rumah Sakit XYZ, perencanaan penyediaan obat masih dilakukan secara manual, hal ini menghambat proses perencanaan kebutuhan obat di masa depan, karena itu diperlukan metode peramalan yang diharapkan dapat membantu dalam merencanakan penyediaan obat yang lebih efisien di rumah sakit XYZ.

Peramalan adalah teori yang menggunakan data dari masa lalu atau saat ini untuk memprediksi peristiwa di masa depan [1]. Dalam konteks institusi kesehatan, peramalan digunakan untuk mendukung pengambilan keputusan [2]. Salah satu metode peramalan yang sering digunakan adalah peramalan *time series* dimana peramalan yang menggunakan kumpulan data pengamatan yang teratur, di mana setiap pengamatan terekam dalam interval waktu yang berurutan [3].

Machine Learning memungkinkan mesin untuk belajar dari data dan mengenali pola tanpa perlu diprogram secara langsung [4]. Hal ini membantu dalam menghadapi data secara efisien dan membuat prediksi untuk masa depan, salah satu algoritma yang dapat digunakan adalah algoritma XGBoost untuk peramalan *time series*.

XGBoost adalah algoritma Machine Learning yang berbasis pada *gradient tree boosting*. Algoritma ini dapat digunakan dalam time series forecasting karena kemampuannya menangani permasalahan skala besar dalam Machine Learning [5].

Beberapa penelitian telah menggunakan XGBoost untuk peramalan *time series*, termasuk peramalan pesanan produk farmasi [6] di mana XGBoost bekerja lebih baik daripada model SARIMA, Prophet, Random Forest, Linear Regression, LSTM, dan LSTM Multi-output untuk memprediksi produk Top 300 dengan RMSE 9,05. peramalan tingkat air tanah di Selangor, Malaysia [7], pada penelitian tersebut XGBoost unggul dengan skor  $R^2$  0,92 dibanding dengan model yang menggunakan *Support Vector Regression*, dan *Artificial Neural Network*. Penelitian tentang peramalan penjualan produk [8] di mana XGBoost bekerja lebih baik daripada LSTM, dan peramalan kecelakaan di Bali [9] di mana XGBoost memiliki kinerja yang baik dalam memprediksi jumlah orang meninggal akibat kecelakaan dengan skor RMSE 4,92 dan jumlah orang dengan luka berat dengan skor RMSE 4,11.

Penelitian ini bertujuan untuk menerapkan teknologi machine learning dengan menggunakan algoritma XGBoost untuk meramalkan kebutuhan obat di rumah sakit. Dengan adanya penelitian ini, diharapkan dapat membantu dalam perencanaan penyediaan obat di rumah sakit tersebut.

## B. Metode Penelitian

Tahapan pada penelitian ini terbagi menjadi beberapa langkah yaitu pengumpulan dataset, preprocessing dataset, *exploratory data analysis*, *dataset split*, *feature engineering*, *hyperparameter tuning* dan *fitting* model XGBoost, dan evaluasi kinerja model menggunakan *evaluation metrics*.

## 1. Pengumpulan *dataset*

Data yang digunakan pada penelitian ini terbagi menjadi 2 yaitu:

### a. Data obat

Data obat yang dikumpul ialah data penggunaan obat *vital* dan *essential* perbulan dari tahun 2017 hingga tahun 2022 dari Rumah Sakit XYZ yang dikumpulkan dari sistem yang digunakan rumah sakit, data yang diambil terdiri atas 12 produk obat *vital* dan 15 produk obat *essential*.

### b. Data cuaca

Data cuaca yang dikumpul adalah data cuaca lokal periode 2017-2022, data cuaca tersebut terdiri atas nilai temperatur rata-rata, temperatur maksimum, temperatur minimum, dan kecepatan angin.

## 2. Preprocessing data

Data yang telah dikumpul akan diproses untuk memudahkan langkah berikutnya, pada tahap ini *missing value* yang terdapat pada data yang dikumpul akan dibersihkan.

## 3. Exploratory data analysis

Pada tahap ini, data yang telah dipreprocess akan dianalisa untuk memahami data yang akan digunakan. pada penelitian ini data akan dipahami melalui visualisasi dengan cara ini dapat mengetahui pola pada data dan mengetahui hubungan antara data fitur dan data target.

## 4. Dataset split

Pada tahap ini, data akan dibagi menjadi *training set* dan *testing set*, penelitian ini menggunakan *time series split* untuk membagi data menjadi beberapa lipatan atau *fold* yang ditentukan, tiap *fold* terdiri atas training data dan testing data, kemudian *fold* berikutnya data terbaru akan menjadi testing data dan data pada *fold* sebelumnya akan menjadi training data, proses ini akan berlanjut hingga jumlah *fold* yang ditentukan terpenuhi.

## 5. Feature engineering

*Feature engineering* adalah proses untuk membuat dan memilih atribut yang akan digunakan dalam memprediksi dengan tujuan untuk meningkatkan hasil prediksi.

## 6. Hyperparameter tuning dan fitting model XGBoost

*Hyperparameter tuning* adalah proses untuk menemukan *hyperparameter* terbaik untuk model XGBoost dengan tujuan untuk mendapatkan kinerja yang optimal ketika melakukan prediksi pada data obat, kemudian model XGBoost akan dilatih menggunakan data penggunaan obat.

## 7. Evaluasi model

Pada penelitian ini, model yang telah dilatih, dievaluasi dengan cara visualisasi perbandingan antara data asli dan data prediksi serta menggunakan *root mean square error* (RMSE) dan *mean absolute error* (MAE) untuk mengetahui kinerja model XGBoost yang telah dilatih menggunakan data penggunaan obat *vital* dan *essential*.

## C. Hasil dan Pembahasan

### 1. Data

Data yang digunakan adalah dataset penggunaan obat *vital* dan *essential* yang diambil dari sistem rumah sakit, tiap data merupakan rekaman penggunaan obat tiap bulannya dari tahun 2017 hingga tahun 2022. Data yang dikumpulkan dapat dilihat pada tabel 1.

**Tabel 1.** Data obat

| NO | Nama_Obat                    | Jumlah | Bulan  | Kelompok |
|----|------------------------------|--------|--------|----------|
| 1  | BRAXIDIN TAB                 | 1819   | 201701 | 2        |
| 2  | CETIRIZINE 10 MG TAB         | 625    | 201701 | 2        |
| 3  | CITICOLINE 125MG INJ/10      | 61     | 201702 | 2        |
| 4  | INDUXIN 1ML/10 (OXYTOCYN)    | 58     | 201701 | 1        |
| 5  | MORPHIN HCL 10MG INJEKSI /10 | 5      | 201702 | 1        |
| 6  | CIPROFLOXACIN 500 MG         | 150    | 202007 | 2        |

### 2. Data preprocessing

Data perbulan penggunaan obat dibagi menjadi 2 file *csv* berdasarkan dari kelompok obat yaitu file obat *vital* berisi data perbulan penggunaan obat dengan nomor kelompok 1 dan untuk file obat *essential* berisi data perbulan penggunaan obat dengan nomor kelompok 2. Setelah itu, data obat akan dilabel ulang dan dibersihkan dari *missing value* dan data cuaca yang telah *dipreprocess* akan diinput ke dalam dua file yang dibuat.

Dalam data obat yang digunakan terdapat periode di mana terjadinya pandemi COVID-19, hal ini mengakibatkan terjadinya peningkatan maupun penurunan drastis dalam penggunaan obat-obat tertentu. Berdasarkan dari [10], periode ini dapat dikategorikan sebagai *outlier*. Setelah berdiskusi dengan *domain expert* periode ini di mulai pada bulan April 2020 hingga bulan Desember 2021, pada penelitian ini data untuk periode pandemi tidak akan diubah.

Setelah data *dipreprocess*, file *csv* memiliki 72 baris dan 18 kolom untuk obat *vital* dan 72 baris dan 21 kolom untuk obat *essential*, Data yang telah diproses akan diinput ke dalam *Jupyter Notebook* dimana jenis data pada kolom tanggal diubah ke format *datetime*. Berikut tabel beberapa data obat *vital* dan *essential* yang telah diproses.

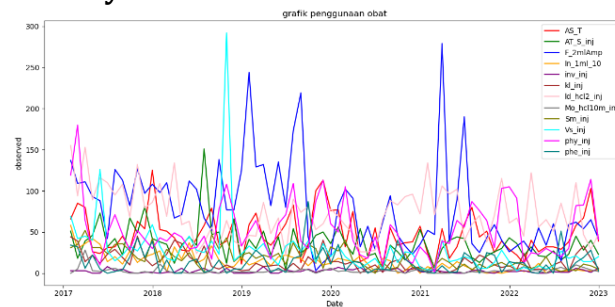
**Tabel 2.** *Preprocessed data obat vital*

| Tanggal   | phy_inj | phe_inj | tavg | tmin | tmax | wspd |
|-----------|---------|---------|------|------|------|------|
| 1/31/2017 | 119     | 34      | 26.6 | 23.6 | 30.9 | 6.6  |
| 2/28/2017 | 180     | 32      | 26.9 | 23.8 | 31.5 | 7.6  |
| 3/31/2017 | 64      | 6       | 27   | 23.9 | 31.7 | 6.4  |
| 4/30/2017 | 27      | 22      | 27.2 | 23.6 | 32.2 | 6.3  |

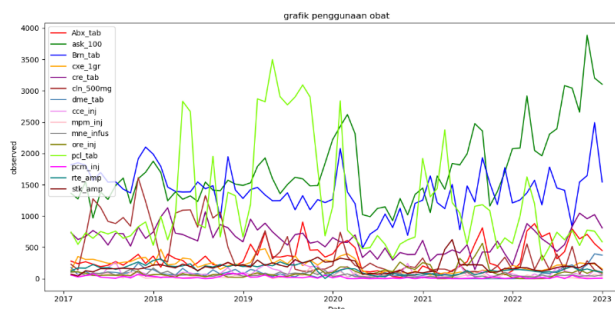
**Tabel 3.** *Preprocessed data obat essential*

| Tanggal   | rte_amp | stk_amp | tavg | tmin | tmax | wspd |
|-----------|---------|---------|------|------|------|------|
| 1/31/2017 | 123     | 72      | 26.6 | 23.6 | 30.9 | 6.6  |
| 2/28/2017 | 171     | 41      | 26.9 | 23.8 | 31.5 | 7.6  |
| 3/31/2017 | 167     | 127     | 27   | 23.9 | 31.7 | 6.4  |
| 4/30/2017 | 226     | 109     | 27.2 | 23.6 | 32.2 | 6.3  |

### 3. Exploratory data analysis



**Gambar 1.** Grafik penggunaan obat *vital*



**Gambar 2.** Grafik penggunaan obat *essential*

Dari grafik diatas dapat dilihat penggunaan obat *vital* dan *essential* perbulan dari tahun 2017 hingga 2022 memiliki penurunan drastis di awal tahun 2020 yang disebabkan oleh pandemi COVID-19 dengan obat *essential* yang lebih terpengaruh dibanding dengan obat *vital*.

#### 4. Dataset split

Pada penelitian ini data dibagi menggunakan *time series split*, dengan jumlah *split* sebanyak 2 *split* dan *test size* sebesar 24 data.

#### 5. Feature engineering

Pada penelitian ini, fitur yang digunakan adalah fitur waktu yang terdiri atas bulan dan tahun, dan fitur cuaca yang terdiri atas *tavg*, *tmin*, *tmax*, dan *wspd*. Masing- masing obat akan memiliki fitur yang berbeda, cara untuk menentukan fitur untuk tiap obat dengan menggunakan *heatmap* untuk melihat korelasi antar obat dan fitur.

#### 6. Hyperparameter tuning dan fitting model XGBoost

Pada penelitian ini *hyperparameter* yang akan digunakan dapat dilihat pada tabel 4. Setelah *hyperparameter* model telah ditentukan, data obat *vital* dan *essential* akan dilatih menggunakan model XGBoost.

**Tabel 4.** Hyperparameter XGBoost

| Nama <i>hyperparameter</i> | Nilai            |
|----------------------------|------------------|
| booster                    | gbtree           |
| Max_depth                  | 20               |
| objective                  | Reg:squarederror |
| N_estimators               | 4000             |
| Learning_rates             | 0.01             |
| Early_stopping_rounds      | 50               |
| subsample                  | 0.4              |
| Colsample_bytree           | 0.5              |

## 7. Evaluasi model

Hasil dari model XGboost untuk obat *vital* dan *essential* memiliki skor RMSE dan MSE yang bervariasi untuk obat *vital* dan *essential*. Skor RMSE dan MAE obat *vital* dan *essential* dapat dilihat pada tabel 1 dan 2.

**Tabel 5.** Skor evaluasi obat *vital*

| Obat <i>vital</i> | RMSE  | MAE   |
|-------------------|-------|-------|
| AS_T              | 24.86 | 18.68 |
| AT_S_inj          | 13.88 | 11.72 |
| F_2mlAmp          | 60.16 | 39.84 |
| In_1ml_10         | 5.00  | 4.10  |
| inv_inj           | 2.49  | 2.00  |
| kl_inj            | 8.77  | 6.58  |
| ld_hcl2_inj       | 25.93 | 21.94 |
| Mo_hcl10m_inj     | 2.83  | 2.27  |
| Sm_inj            | 7.35  | 6.12  |
| Vs_inj            | 10.62 | 8.75  |
| phy_inj           | 30.49 | 24.21 |
| phe_inj           | 15.75 | 10.16 |

**Tabel 6.** Skor evaluasi obat *essential*

| Obat <i>essential</i> | RMSE   | MAE    |
|-----------------------|--------|--------|
| Abx_tab               | 262.97 | 213.07 |
| ask_100               | 681.24 | 556.83 |
| Brn_tab               | 361.35 | 269.54 |
| cxe_1gr               | 77.28  | 66.41  |
| cre_tab               | 188.32 | 152.66 |
| cln_500mg             | 174.71 | 128.61 |
| dme_tab               | 68.94  | 51.84  |
| cce_inj               | 71.77  | 57.34  |
| mpm_inj               | 36.46  | 29.15  |
| mne_infus             | 25.09  | 20.68  |
| ore_inj               | 108.84 | 76.65  |
| pcl_tab               | 763.43 | 597.83 |
| pcm_inj               | 15.49  | 11.30  |
| rte_amp               | 56.70  | 49.34  |
| stk_amp               | 95.77  | 67.29  |

Berdasarkan dari tabel 5 dan 6 model yang telah dilatih memiliki kinerja yang lebih baik pada model yang dilatih menggunakan data penggunaan obat *vital* dibanding dengan model yang dilatih menggunakan data penggunaan obat *essential*, hal ini terjadi karena obat *essential* memiliki *range* penggunaan obat yang lebih banyak dibanding obat *vital* sehingga *margin of error*-nya lebih besar, alasan lain ialah karena *hyperparameter* yang digunakan bersifat merata untuk semua obat sehingga terdapat data obat yang cocok dengan *hyperparameter* yang digunakan dan terdapat yang tidak cocok.

#### D. Simpulan

Berdasarkan dari hasil model XGBoost yang dilatih dalam penelitian ini, XGBoost dapat digunakan untuk obat *vital* dengan rata-rata skor RMSE 17,34 dan MAE 13,03 namun skor tersebut dapat lebih rendah lagi, untuk obat berkategori *essential* masih membutuhkan perbaikan untuk mendapatkan performa model yang lebih optimal.

Saran untuk penelitian selanjutnya ialah menggunakan metode *cross validation* untuk memilih nilai pada *hyperparameter tuning*, mengumpulkan data yang lebih banyak, menggunakan fitur pola penyakit, dan merancang model *single-output* yang melatih masing-masing data penggunaan obat agar *hyperparameter* sesuai untuk tiap obat.

#### E. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada pihak rumah sakit yang telah bersedia untuk memberikan data penggunaan obatnya untuk digunakan dalam penelitian ini, serta terima kasih kepada Program Studi Sistem Informasi Universitas Muslim Indonesia yang telah memberi arahan untuk penelitian ini.

#### F. Referensi

- [1] A. N. Dzaky, "Implementation Of Unsupervised And Supervised Machine Learning To Determine Minimum Profit Margin In Agricultural Sector (Case Study: CV Rawin Agro Nusa)," 2022.
- [2] M. I. Anshory, Y. Priyandari, and Y. Yuniaristanto, "Peramalan Penjualan Sediaan Farmasi Menggunakan Long Short-term Memory: Studi Kasus pada Apotik Suganda," *Performa: Media Ilmiah Teknik Industri*, vol. 19, no. 2, Oct. 2020, doi: 10.20961/performa.19.2.45962.
- [3] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice, 3rd edition," *OTexts: Melbourne, Australia.*, 2021. OTexts.com/fpp3 (accessed May 07, 2023).
- [4] Sumpena, "Machine Learning Klasifikasi Penduduk Miskin Wilayah Desa Tarajau Kabupaten Tasikmalaya Dengan Menggunakan Support Vektor Machine," 2021.
- [5] A. N. Rachmi, "Implementasi Metode Random Forest Dan XGBoost Pada Klasifikasi Customer Churn," 2020.
- [6] B. K. Almentero, J. Li, and C. Besse, "Forecasting pharmacy purchases orders," *Proceedings of 2021 IEEE 24th International Conference on Information Fusion, FUSION 2021*, 2021, doi: 10.23919/FUSION49465.2021.9627017.
- [7] A. Ibrahim Ahmed Osman, A. Najah Ahmed, M. F. Chow, Y. Feng Huang, and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1545–1556, Jun. 2021, doi: 10.1016/j.asej.2020.11.011.
- [8] D. Swami, A. D. Shah, and S. K. B. Ray, "Predicting Future Sales of Retail Products using Machine Learning," Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.07779>
- [9] N. Nyoman, P. Pinata, M. Sukarsa, N. Kadek, and D. Rusjyanti, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python."

- 
- [10] Eurostat, "Guidance on time series treatment in the context of the COVID-19 crisis," 2020. Accessed: Aug. 05, 2023. [Online]. Available: [https://ec.europa.eu/eurostat/documents/10186/10693286/Time\\_series\\_treatment\\_guidance.pdf](https://ec.europa.eu/eurostat/documents/10186/10693286/Time_series_treatment_guidance.pdf)