

STORAGE SERVER DATABASE UTILIZATION FORECASTING USING HOLT-WINTERS AND ARIMA METHODS IN E-GOVERNMENT SYSTEM. STUDY AT KEMENKEU RI

Adinda Krida Wicaksono^{*1}, Teguh Prasetyo², Nazori Az³

^{1,2,3}Master of Computer Science, Faculty of Information Technology, Universitas Budi Luhur, Indonesia
Email: ¹2111601874@student.budiluhur.ac.id, ²2111602013@student.budiluhur.ac.id, ³nazori@budiluhur.ac.id

(Article received: June 19, 2023; Revision: August 24, 2023; published: Desember 23, 2023)

Abstract

Storage is a storage medium which is an important part of the system infrastructure. Storage utilization is one of the variables in determining the needs and performance of a system. If the storage is lacking or depleted it will cause the system to be hampered and the possibility of interference occurring. There were eight complaints or reports of system disturbances recorded in the Ministry of Finance's ticketing service desk application caused by anomalies in storage or disk capacity. Things like this can be anticipated if we know the storage capacity requirements for the future. In compiling ICT capacity analysis at Pusintek there has been no use of past data to plan future storage requirements. The use of data mining algorithms can be used to obtain forecast values from storage utilization. The use of the ARIMA model and Holt-Winters Exponential Smoothing as a method used to predict inventory is considered to be a fairly accurate model. Both algorithms are used to compare which accuracy is better in predicting storage requirements by measuring RMSE. Research data was obtained from monthly utilization reports from January 2021 to December 2022. From the evaluation results it can be concluded that forecasting storage server database utilization using the Holt-Winters method is better than the ARIMA method with the RMSE results for the Holt-Winters method being 14332.661717740748 and the RMSE method ARIMA is 20498.977982137938. The results of this forecasting can be utilized in planning database server storage needs.

Keywords: ARIMA, Capacity Planning, data mining, Holt-Winter, Prediction, Storage.

PERAMALAN UTILISASI STORAGE SERVER DATABASE MENGGUNAKAN METODE HOLT-WINTERS DAN ARIMA PADA SISTEM E-GOVERNMENT. STUDI PADA KEMENKEU RI

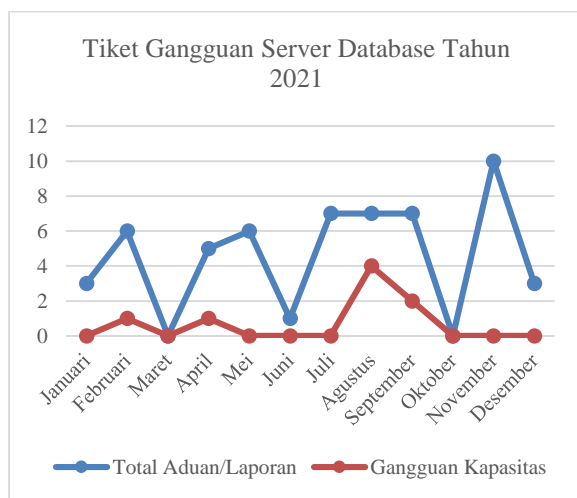
Abstrak

Storage adalah media penyimpanan yang merupakan bagian penting dari infrastruktur sistem. Utilisasi storage menjadi salah satu variabel dalam menentukan kebutuhan dan performa suatu sistem. Apabila storage kurang atau habis akan mengakibatkan sistem terhambat dan kemungkinan terjadi gangguan. Terdapat delapan aduan atau laporan gangguan sistem tercatat pada aplikasi ticketing service desk Kemenkeu yang disebabkan oleh anomali pada kapasitas storage atau disk. Hal seperti ini dapat diantisipasi apabila kita mengetahui kebutuhan kapasitas storage untuk masa depan. Dalam penyusunan analisis kapasitas TIK di Pusintek belum ada pemanfaatan data lampau yang digunakan untuk merencanakan kebutuhan storage di masa depan. Penggunaan data mining algoritma dapat digunakan untuk mendapatkan nilai peramalan dari utilisasi storage. Penggunaan model ARIMA dan Exponential Smoothing Holt-Winters sebagai metode yang digunakan untuk memprediksi persediaan dinilai menjadi model yang cukup akurat. Kedua algoritma digunakan untuk membandingkan akurasi mana yang lebih baik dalam memprediksi kebutuhan storage dengan mengukur RMSE. Data penelitian didapatkan dari laporan utilisasi setiap bulan dari Januari 2021 sampai Desember 2022. Dari hasil evaluasi dapat disimpulkan bahwa peramalan utilisasi storage server database menggunakan metode Holt-Winters lebih baik daripada metode ARIMA dengan hasil RMSE metode Holt-Winters adalah 14332,661717740748 dan RMSE metode ARIMA adalah 20498,977982137938. Hasil dari peramalan ini dapat dimanfaatkan dalam melakukan perencanaan kebutuhan storage server database.

Kata kunci: ARIMA, data mining, Holt-Winter, Penyimpanan, Peramalan, Perencanaan Kapasitas.

1. PENDAHULUAN

Dalam melaksanakan tugas dan fungsi penyusunan analisis kapasitas TIK yang tertuang pada visi organisasi yang tertuang pada Peraturan Menteri Keuangan Nomor 118/PMK.01/2021 tentang Organisasi dan Tata Kerja Kementerian Keuangan, diperlukan perencanaan yang terukur terkait dengan kapasitas atau utilisasi perangkat, salah satunya adalah perangkat storage server. Storage adalah media penyimpanan yang merupakan bagian penting dari infrastruktur sistem. Utilisasi storage menjadi salah satu variabel dalam menentukan kebutuhan dan performa suatu sistem. Apabila storage kurang atau habis akan mengakibatkan sistem terhambat dan kemungkinan terjadi gangguan.



Gambar 1 Aduan/Laporan Gangguan Server Database Tahun 2021

Pada Gambar 1, menampilkan laporan gangguan Server Database Tahun 2021. Data diperoleh dari hasil ekspor pada aplikasi pencatatan gangguan yaitu servicedesk.kemenkeu.go.id yang merupakan portal *ticketing* Service Desk Kemenkeu. Laporan atas gangguan terhadap suatu sistem di Kemenkeu bias dilaporkan melalui portal Service Desk Kemenkeu oleh semua pegawai Kemenkeu. Terdapat 8 (delapan) aduan atau laporan gangguan sistem tercatat yang disebabkan oleh anomali pada kapasitas storage atau disk server database. Salah satunya adalah laporan gangguan dengan nomor tiket 658448 tentang Nadine (Sistem Naskah Dinas Elektronik Kemenkeu) mengalami kelambatan dan juga gagal simpan naskah yang mengakibatkan gangguan layanan persuratan masal di seluruh lingkungan Kemenkeu. Dalam melaksanakan tugas dan fungsi pelaksanaan operasional TIK, Pusintek berkewajiban menjaga semua sistem Kemenkeu minim gangguan sesuai dengan *Service Level Agreement* yang disepakati dengan pengguna. Dalam melaksanakan tugas dan fungsi penyusunan analisis kapasitas TIK, hal seperti ini dapat diantisipasi apabila kita mengetahui kebutuhan kapasitas storage untuk masa depan.

Dalam merencanakan kebutuhan kapasitas storage yang merupakan bagian dari penyusunan analisis kapasitas TIK di Pusintek saat ini belum ada pemanfaatan data lampau yang digunakan untuk merencanakan kebutuhan storage di masa depan. Kondisi saat ini untuk penyusunan kebutuhan kapasitas storage tahun berikutnya, bagian perencanaan melakukan permintaan kebutuhan storage setiap tahunnya untuk digunakan sebagai bahan pertimbangan dalam perencanaan kapasitas storage. Untuk meningkatkan efisiensi dan efektifitas dalam perencanaan kebutuhan kapasitas storage diperlukan perhitungan terukur yang dapat menunjang dalam pengambilan keputusan di masa depan. Penggunaan data mining algoritma dapat digunakan untuk mendapatkan nilai peramalan dari utilisasi storage.

Penelitian sebelumnya yang bertujuan untuk menentukan peramalan utilisasi, kapasitas, dan persediaan diantaranya dilakukan dengan menggunakan metode Holt-Winters [1]–[9], ARIMA [4]–[8], [10]–[12] serta beberapa metode lain [1], [4], [13]. Metode Holt-Winters banyak digunakan untuk memprediksi utilisasi kapasitas dikarenakan cocok untuk data dengan time series yang memiliki pola seasonality (naik turun disekitar alur trend). Metode ARIMA (Autoregressive Integrated Moving Average) adalah metode prediktif yang dapat membuat peramalan berdasarkan sintesa pola data historis.

Beberapa penelitian terdahulu terkait prediksi utilisasi database dilakukan dengan berbagai pendekatan yang telah dilakukan antara lain:

1. Ahmad dan Bowman [14] melakukan penelitian prediksi *workload* dengan dataset utilisasi cpu dan transfer disk. Pada penelitian ini diasumsikan bahwa atribut penggunaan cpu dan transfer disk mempengaruhi penggunaan *workload* database. Metode yang digunakan adalah regresi linear. Lingkungan penelitian pada area development. Database yang digunakan Sybase sql anywhere.
2. Tirado dkk. [15] melakukan analisis *workload* pada *web* infrastruktur untuk melakukan adaptasi penggunaan *workload* secara dinamis dengan memanfaatkan data lampau log server database dan menggunakan metode ARIMA dalam peramalannya.
3. Roy, Dubey dan Gokhale [16] melakukan penelitian prediksi utilisasi dengan metode ARMA dengan rentang waktu dataset jam. Atribut yang digunakan adalah Utilisasi *Response Time Analysis Server*.
4. Mozafari dkk. [17] menggunakan database MySQL serta tools DBSeer untuk pengujian pada lingkungan testing dengan database OLTP. Dataset yang digunakan adalah log server database.
5. Ma dkk. [18] menggunakan database MySQL dan PostgreSQL pada lingkungan testing

dengan waktu peramalan selama satu minggu. Dataset yang digunakan adalah log workload (beban kerja) *query* database. Tujuan dari penelitian tersebut adalah untuk menentukan optimasi *query* yang sesuai berdasarkan hasil prediksi yang didapatkan.

6. Baldan dkk. [19] menggunakan metode ARIMA dan Pemulusan Eksponensial dalam melakukan prediksi utilisasi *workload* pada server *cloud*.

Pada penelitian ini, dilakukan peramalan utilisasi storage database dengan metode Holt-Winters dan ARIMA. Kedua algoritma digunakan untuk membandingkan akurasi mana yang lebih baik dalam memprediksi kebutuhan storage dengan mengukur RMSE. Data yang digunakan untuk penelitian ini adalah data bertipe *univariate* yang diekstrak dari laporan utilisasi setiap bulan untuk server database SQL Server yang dikelola oleh Pusintek Kemenkeu. Dari data set yang telah dikumpulkan, data set merupakan data set *time series* yang menunjukkan trend peningkatan utilisasi storage. Ukuran data akan selalu mengalami perubahan. Operasi Insert atau update bisa menambah ukuran data, sedangkan operasi delete bisa mengurangi ukuran data yang tersimpan.

Hasil dari penelitian penambangan data lampau utilisasi storage ini diharapkan dapat menjadi solusi untuk memprediksi kebutuhan storage server pada periode berikutnya yang bermanfaat dalam melakukan penyusunan analisis kapasitas TIK khususnya pada perencanaan kebutuhan storage server database.

2. METODE PENELITIAN

2.1. Tinjauan Studi

Data mining adalah teknik pengolahan data mentah dalam jumlah besar menjadi bentuk informasi yang berguna dalam pengambilan keputusan [20]. Penambangan data adalah kegiatan yang mengekstraksi informasi penting yang tersirat dan sebelumnya tidak diketahui dari data [21]. Penambangan data adalah kegiatan menemukan keteraturan, hubungan dan juga pola dari data historis yang telah dikumpulkan sebelumnya dalam sebuah kumpulan data yang besar [22]. Penambangan data adalah hasil penggalian pola atau informasi yang menarik dan kemungkinan bermanfaat dari sekumpulan data besar [23].

Penambangan data atau data mining adalah serangkaian proses yang menggabungkan ilmu statistik, matematika, artificial intelligence dan juga machine learning dalam melakukan ekstraksi dan identifikasi informasi yang berguna terkait basis data yang besar. Ketika informasi telah digali dan diubah menjadi informasi yang berguna, dan menjadi informasi, dapat digunakan sebagai dasar pengambilan keputusan dengan menggunakan informasi yang dihasilkan.

Peramalan adalah kegiatan memprediksi kejadian di masa depan dengan menggunakan dan mempertimbangkan data dari masa lalu. Prakiraan adalah alat penting untuk perencanaan yang efektif dan efisien. Perkiraan penggunaan dan tentukan kuantitas inventaris yang diperlukan untuk perencanaan kapasitas inventaris. Peramalan merupakan sebuah proses sistematis yang melakukan perkiraan nilai yang paling mungkin terjadi di masa depan berdasarkan data masa lalu dan sekarang dengan meminimalkan kesalahan [24]. Hasil dari nilai peramalan belum tentu pasti benar akan terjadi, tetapi bertujuan untuk mencari jawaban sedekat mungkin dengan apa yang akan terjadi.

Pemodelan prediktif (peramalan) data deret waktu dibagi menjadi linier dan nonlinier. Untuk model deret waktu linier, ini adalah eksponensial-Holt, Holt-Winters, ARIMA-Box-Jenkins, model intervensi, fungsi transfer (ARIMAX), dan VARIMA (VARIMAX). Model deret waktu nonlinier mencakup model jaringan saraf autoregresif nonlinier, model nonparametrik yang bergantung pada negara, dan model teori ekonomi.

Metode Holt-Winters atau metode pemulusan eksponensial tiga kali adalah salah satu metode dengan melakukan pemulusan tiga kali sebelum dilakukan prediksi. Metode ini adalah perluasan dari dua parameter Holt. Metode Holt-Winters merupakan metode peramalan deret waktu untuk data yang memiliki perilaku musiman.

Menurut Makridakis, et.al (1999) dalam jurnal I. R. Akolo (2019) [6], ada tiga persamaan yang digunakan dalam metode ini, yaitu:

- Pemulusan eksponensial data asli

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1 + \alpha) + (S_{t-1} + b_{t-1}) \quad (1)$$

- Pemulusan pola trend

$$b_t = \gamma (S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad (2)$$

- Pemulusan pola musiman

$$I_t = \delta \frac{X_t}{S_t} + (1 + \delta)I_{t-L} \quad (3)$$

- Ramalan m periode ke depan

$$\hat{Y}_{t-m} = (S_t + b_t m)I_{t-L+m} \quad (4)$$

Dengan I_t = nilai pemulusan musiman waktu ke- t , S_t = nilai pemulusan eksponensial waktu ke- t , X_t = data ke- t , α = konstanta pemulusan data asli ($0 < \alpha < 1$), δ = konstanta pemulusan pola musiman ($0 < \delta < 1$), γ = konstanta pemulusan pola trend ($0 < \gamma < 1$), L = panjang musiman, b_t = konstanta

pemulusan musiman pada waktu t , m = jumlah periode yang akan diramalkan, \hat{Y}_{t-m} = nilai peramalan untuk m periode ke depan.

Metode ARIMA (Autoregressive Integrated Moving Average) adalah metode prediktif atau prediktif yang dapat membuat peramalan berdasarkan sintesa pola data historis. ARIMA sering disebut sebagai metode deret waktu Box-Jenkins. ARIMA, atau Autoregressive Integrated Moving Average. Metode ini pertama kali dikemukakan oleh George Edward Pelham Box dan Gwilym Meirion Jenkins. Model ARIMA merupakan model univariat, sehingga model ini cocok digunakan ketika observasi deret waktu tidak berhubungan secara statistik. Apabila dilakukan komparasi dengan beberapa metode peramalan yang berbeda karakteristik seperti GARCH, VAR, dan CAPM, metode ARIMA merupakan metode yang lebih sesuai dan cocok dengan karakteristik data deret waktu [25].

Persyaratan awal yang harus dipenuhi dalam melakukan pemodelan dengan metode ARIMA adalah data harus bertipe stasioner dan memiliki error atau residual yang merupakan *white noise* (residu tidak berkorelasi otomatis dan terdistribusi secara normal). Untuk menguji stasioneritas data dapat dilakukan pengujian *augmented Dickey-Fuller*, sedangkan untuk melakukan pendeteksian *white noise* dan normalitas dari data yang ada dapat digunakan pengujian dengan Ljung-Box dan uji Kolmogorov-Smirnov [26].

Bentuk umum ARIMA (p,d,q) dengan d merupakan orde dari *differencing* adalah sebagai berikut.

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)\alpha_t \quad (5)$$

Dimana: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$
 $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

Dengan ϕ_i adalah parameter model *autoregressive* dan θ_i adalah parameter model *moving average*.

Untuk mengetahui apakah model yang telah dibuat optimal atau tidak, dilakukan evaluasi dengan mengukur akurasi atau jumlah kesalahan pada model yang kita buat. Berikut adalah kriteria evaluasi dan validasi model:

1. Akurasi

Menunjukkan seberapa baik model yang telah dibuat dengan mengkorelasikan hasil yang diperoleh dengan atribut data yang disediakan. Model data mining dapat dikatakan akurat apabila menghasilkan data yang mendekati data yang ada.

2. Keandalan

Evaluasi dengan melakukan pengukuran di mana model penambahan data diterapkan ke kumpulan data yang berbeda. Model data mining dapat dipercaya jika menghasilkan model umum yang sama terlepas dari data uji yang diberikan.

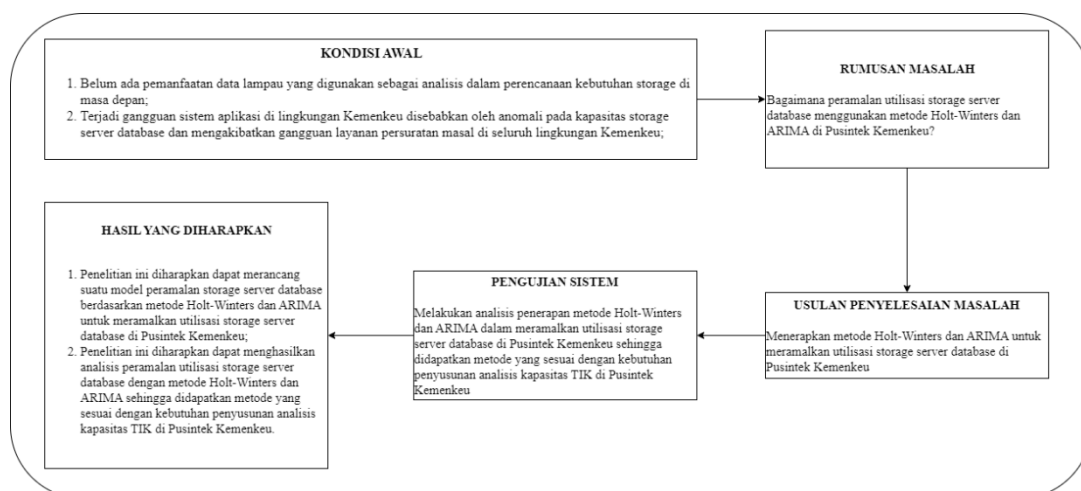
3. Kegunaan

Berisi beberapa metrik yang mengukur apakah model yang telah dibuat memberikan informasi yang berguna dan sesuai tujuan awal penambahan data.

Keseimbangan antara ketiganya diperlukan karena model yang akurat belum tentu andal, dan model yang andal maupun akurat belum tentu berguna. Banyak metode yang dapat digunakan untuk mengevaluasi hasil studi yang dilakukan, antara lain: RMSE (Root Mean Squared Error), Confusion Matrix dan Area Under the Curve (AUC) / ROC (Receiver Operating Characteristics), Silhouette Factor, Lift Ratio dan MAPE (Mean Absolute Percentage Error).

2.2. Kerangka Konsep

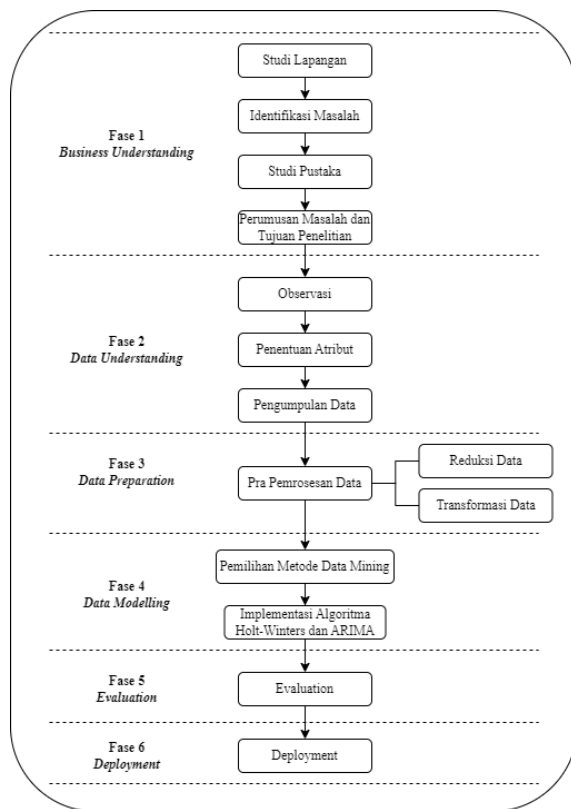
Kerangka konsep penelitian ini dapat dilihat pada Gambar 2, kerangka konsep diawali dengan melihat kondisi awal di lapangan atau objek penelitian, dilanjutkan dengan merumuskan masalah disertai dengan usulan penyelesaian masalah. Selanjutnya dilakukan pengujian sistem berdasar hasil usulan penyelesaian masalah sebelumnya sehingga didapatkan hasil yang diharapkan.



Gambar 2 Kerangka Konsep

2.3. Metodologi Penelitian

Langkah penelitian yang digunakan mengacu pada kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang memiliki enam fase [27] dan dikenalkan pertama tahun 1996 [28]. Tahapan penelitian yang dilakukan dalam penelitian ini ditampilkan pada Gambar 3.



Gambar 3 Tahapan Penelitian.

Tahapan dimulai dengan fase *Business Understanding*. Pada tahap ini dilakukan beberapa kegiatan untuk memahami latar belakang permasalahan yang dihadapi dengan lebih komprehensif sehingga menghasilkan pertanyaan penelitian dan tujuan data mining.

Setelah merumuskan masalah dan tujuan penelitian, fase selanjutnya adalah *Data Understanding*. Pada fase ini kegiatan diawali dengan observasi yang bertujuan mendapatkan pemahaman terhadap data serta didapatkan data yang sesuai untuk pemecahan masalah.

Data yang telah melalui tahap data understanding perlu dilakukan beberapa aktivitas sebelum digunakan untuk pemodelan. Hal ini akan dilakukan pada tahap *data preparation*. Pada tahap ini dilakukan seleksi dan pre-processing data awal sehingga menjadi data bersih yang sudah diseleksi dan dapat dijadikan input untuk proses prediksi.

Setelah mendapatkan file data set keseluruhan utilisasi disk server database Nadine, kegiatan dilanjutkan pada fase data modelling. Pada tahap ini dilakukan pemilihan metode data mining yang sesuai

dengan karakter dataset dan tujuan penelitian. Selanjutnya dilakukan pemodelan data menggunakan metode Holt-Winters dan ARIMA. Tahap ini bertujuan mengolah dan memanfaatkan data untuk menghasilkan informasi tersembunyi dari balik data tersebut.

Fase selanjutnya adalah dengan melakukan evaluasi. *Output* yang dihasilkan pada tahap ini adalah didapatkan hasil evaluasi berupa nilai RMSE dari hasil pemodelan peramalan utilisasi storage server database. Metode estimasi yang mempunyai RMSE lebih kecil dikatakan lebih akurat daripada metode estimasi yang mempunyai RMSE lebih besar. Dari hasil tersebut didapatkan hasil peramalan dengan RMSE terbaik

Fase terakhir adalah *Deployment*. Pada tahap ini dilakukan implementasi metode Holt-Winters dan ARIMA untuk meramalkan utilisasi storage server database di Pusintek Kemenkeu. Selanjutnya dilakukan analisis peramalan utilisasi storage server database dengan metode Holt-Winters dan ARIMA yang telah terbentuk sehingga didapatkan metode yang sesuai dengan kebutuhan penyusunan analisis kapasitas TIK di Pusintek Kemenkeu.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Dari hasil studi lapangan didapatkan permasalahan terkait perencanaan kapasitas storage server database yang belum optimal dibuktikan dengan terjadinya gangguan layanan sistem aplikasi di lingkungan Kemenkeu yang disebabkan oleh anomali pada kapasitas storage server database dan mengakibatkan gangguan layanan persuratan masal di seluruh lingkungan Kemenkeu. Gangguan layanan sistem tercatat pada aplikasi ticketing service desk Kemenkeu.

Permasalahan lain yang ditemukan adalah tidak adanya pemanfaatan data lampau yang digunakan sebagai analisis dalam perencanaan kebutuhan storage di masa depan. Utilisasi storage menjadi salah satu variabel dalam menentukan kebutuhan dan performa suatu sistem. Apabila storage kurang atau habis akan mengakibatkan sistem terhambat dan kemungkinan terjadi gangguan. Pemanfaatan data lampau dapat berguna untuk peramalan kebutuhan storage.

Hasil dari tahapan ini adalah terdefinisinya rumusan masalah dan tujuan penelitian. Rumusan masalah yang didapatkan adalah dengan melakukan perencanaan kapasitas berdasarkan peramalan utilisasi storage server database menggunakan metode Holt-Winters dan ARIMA di Pusintek Kemenkeu.

3.2. Data Understanding

Kegiatan dimulai dari observasi untuk mengetahui sifat data baik itu jenis data, pola data dan

juga trend data. Pada fase ini juga ditentukan atribut apa saja yang akan digunakan dalam fase selanjutnya. Selanjutnya dilakukan pengumpulan data dengan cara ekspor penuh data mentah. Tujuan dari fase ini adalah didapatkan pemahaman terhadap data serta didapatkan data yang sesuai untuk pemecahan masalah.

Kesimpulan dari pemahaman data yang dilakukan secara langsung adalah:

1. Dari dataset yang ada, terlihat bahwa data tersebut merupakan dataset time series yang menunjukkan tren peningkatan utilisasi storage.
2. Ukuran data akan selalu mengalami perubahan. Operasi insert atau update dapat menambah ukuran data, sementara operasi delete dapat mengurangi ukuran data yang tersimpan.
3. Terdapat pola musiman dalam pertumbuhan data, terutama selama musim penilaian perilaku pada awal dan tengah tahun.

Setelah mengetahui karakteristik dari data, dilakukan penarikan data set. Data tersebut masih berupa data mentah yang berisi atribut asli antara lain Date & Time, Value, Low DT, High DT, Smooth dengan interval waktu 10 (sepuluh) menit.

Output dari tahapan data understanding didapatkan data primer berupa hasil observasi dan hasil wawancara serta data sekunder berupa 11 file csv dataset hasil ekspor utilisasi storage server database sistem persuratan Kemenkeu atau Nadine (Sistem Naskah Dinas Elektronik Kemenkeu).

3.3. Data Preparation

Tahap pertama data awal yang menjadi rujukan utama adalah data utilisasi storage server database per disk. Pertama dilakukan kegiatan reduksi data. Kegiatan reduksi data perlu dilakukan untuk membersihkan data dari unsur-unsur yang dianggap tidak berpengaruh pada saat proses perhitungan data. Dalam data utilisasi storage server database per disk, output dari tahap data understanding, perlu dilakukan proses penghilangan atribut dan data yang tidak diperlukan. Atribut awal pada data ini adalah Date & Time, Value, Low DT, High DT, Smooth. Tidak semua atribut diperlukan untuk penelitian ini. Untuk kebutuhan penelitian ini, atribut yang dibutuhkan adalah Date & Time dan Value. Atribut Date & Time adalah atribut yang berisi data dengan tipe *date*. Isi data dari atribut Date & Time adalah waktu pencatatan kapan utilisasi storage server tersebut

diambil / direkam. Sedangkan atribut Value adalah atribut yang berisi data dengan tipe Number. Isi data dari atribut Value adalah besaran utilisasi storage server database dalam satuan GB (GigaByte).

Setelah itu dilanjutkan dengan melakukan transformasi data. Dikarenakan data awal yang didapatkan untuk atribut Date & Time memiliki interval 10 menit, pada tahap ini dilakukan transformasi data dengan melakukan generalisasi data menjadi data dengan interval waktu per bulan.

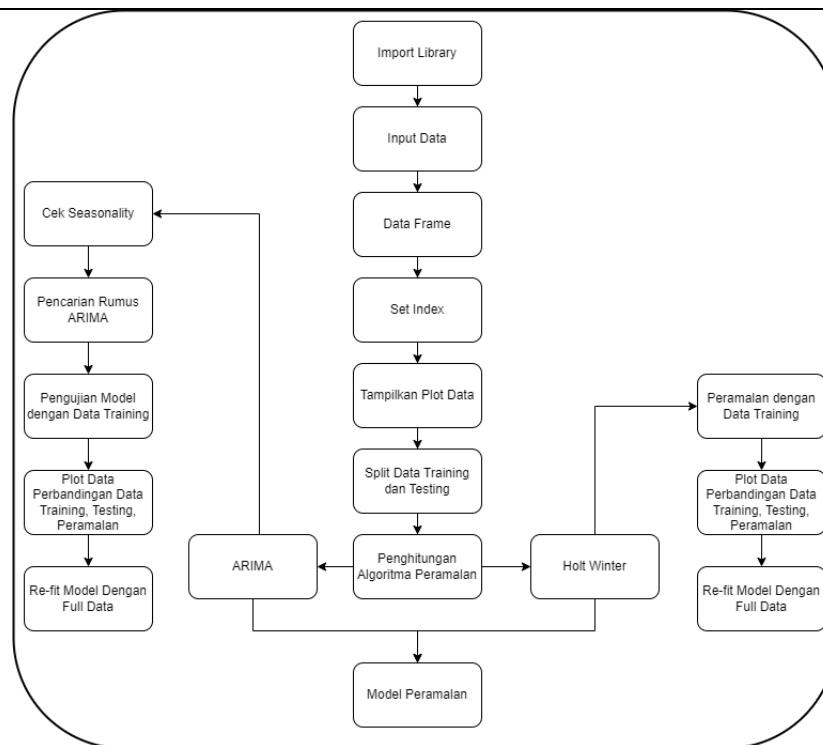
Selain itu akan dilakukan pemeriksaan terhadap konsistensi data. Misalnya tidak ada data *null* atau kosong pada atribut Date & Time dan Value. Selanjutnya digabungkan dengan data utilisasi storage server database untuk disk yang lain (C, D, E, F, G, H, I, J, K, L, M) sehingga didapatkan nilai total dari atribut Value untuk utilisasi storage server database dengan satuan GB. Output akhir dari tahapan ini berupa satu file data set keseluruhan utilisasi disk server database sistem persuratan Kemenkeu atau Nadine (Sistem Naskah Dinas Elektronik Kemenkeu) sesuai pada Tabel 1.

Tabel 1 Data Set Utilisasi Storage

Bulan	2021	2022
	Total Utilisasi Storage (GB)	Total Utilisasi Storage (GB)
Januari	42214.29	59474.38
Februari	42411.42	61028.27
Maret	53900.33	67935.03
April	54908.93	80210.63
Mei	56416.96	81040.48
Juni	56408.51	84841.33
Juli	58106.64	77029.70
Agustus	59814.66	76774.70
September	58523.08	80610.72
Oktober	60639.49	86533.10
November	59105.22	90123.80
Desember	60010.90	95647.11

3.4. Data Modelling

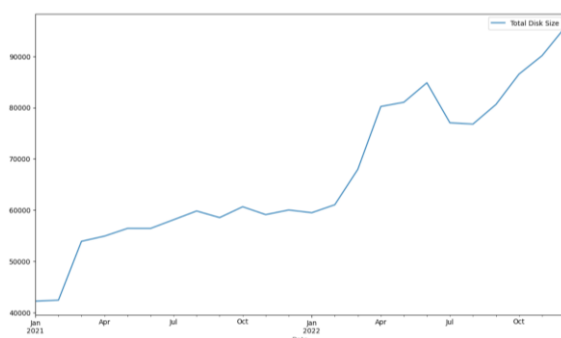
Pada tahap ini dilakukan pemilihan metode data mining yang sesuai dengan karakter data set dan tujuan penelitian. Tahap ini bertujuan mengolah dan memanfaatkan data untuk menghasilkan informasi tersembunyi dari balik data tersebut. Output dari tahap ini adalah hasil pemodelan peramalan utilisasi storage server database dengan menggunakan metode Holt-Winters dan ARIMA. Pada tahap ini didapatkan juga nilai hasil peramalan utilisasi storage server database dengan menggunakan metode Holt-Winters dan ARIMA.



Gambar 4 Tahapan Data Modeling

Tahapan yang dilakukan pada tahap ini digambarkan pada Gambar 4. Tahapan pertama dimulai dengan import library yang dibutuhkan pada lingkungan penelitian. Dalam penelitian kali ini lingkungan yang digunakan adalah menggunakan bahasa pemrograman python.

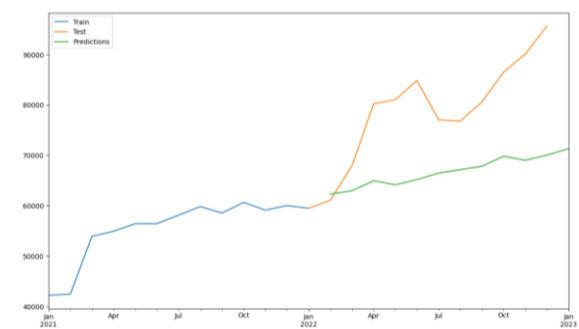
Tahapan selanjutnya adalah dengan melakukan input data hasil dari output data preparation sebelumnya (Tabel 1). Setelah itu dibuatlah data frame yang selanjutnya dilakukan set index pada kolom Date dengan frekuensi bulanan, sehingga didapatlah total 24 index baris data. Tampilan dari plot data awal ditunjukkan pada Gambar 5. Parameter trend yang digunakan adalah dengan nilai seasonality enam bulan. Nilai enam bulan didapatkan karena periode penilaian kinerja di lingkungan Kemenkeu adalah enam bulan. Dengan adanya periode ini, diduga mempengaruhi utilisasi storage server database dikarenakan banyaknya transaksi persuratan yang dilakukan.



Gambar 5 Plot Data Awal

Pemisahan data training dan testing dipisahkan secara manual dengan cara linear sampling. Linear sampling adalah membagi dataset menjadi beberapa bagian tanpa merubah urutan data sample sehingga menghasilkan bagian data yang berurutan. Dari total 24 data dilakukan split data training dan testing, dengan komposisi data training sejumlah 12 data awal, dan data testing 12 data akhir. Setelah data siap diolah, dilanjutkan pembuatan model peramalan dengan menggunakan metode ARIMA dan Holt-Winters.

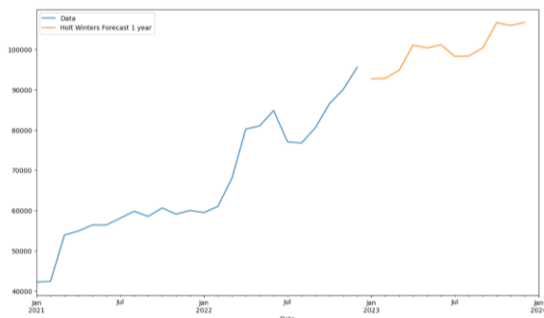
Pembuatan model peramalan dengan metode Holt-Winters diawali dengan melakukan peramalan dengan menggunakan data training. Setelah itu dibandingkan hasil peramalan awal yang didapatkan dengan data testing. Grafik hasil perbandingan data training, data testing dan hasil peramalan awal ditunjukkan pada Gambar 6.



Gambar 6 Grafik Perbandingan Data Metode Holt-Winters

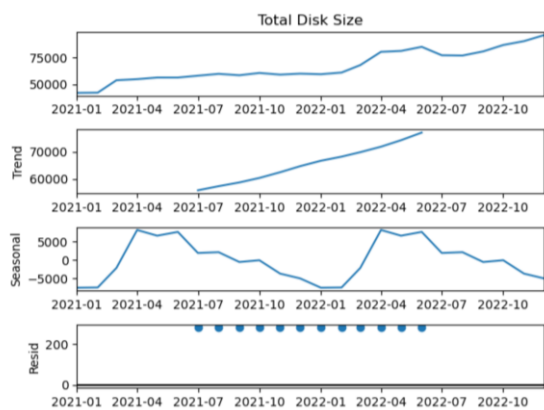
Proses selanjutnya adalah dengan melakukan peramalan atas seluruh data dengan menggunakan model yang dihasilkan dengan metode Holt-Winters.

Grafik hasil peramalan pada tahun 2023 ditunjukkan pada Gambar 7.



Gambar 7 Grafik Hasil Peramalan Dengan Metode Holt-Winters

Metode peramalan pembanding adalah dengan melakukan peramalan dengan metode ARIMA. Langkah pertama yang dilakukan adalah dengan melakukan pengecekan seasonality pada data awal. Hasil dari pengecekan seasonality ditampilkan pada Gambar 8.



Gambar 8 Pengecekan Seasonality ARIMA

Setelah itu dilakukan pencarian model arima paling tepat untuk data utilisasi storage database. Didapatkan model ARIMA(2,0,2)(1,1,1)[6] (Gambar 9).

```
auto_arima(df['Total Disk Size'],seasonal=True, m=6, trace=True).summary()
```

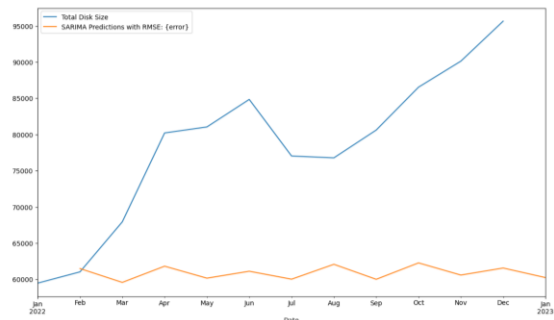
Performing stepwise search to minimize aic

Model	Intercept	AIC	Time
ARIMA(2,0,2)(1,1,1)[6]	intercept	360.645	0.71 sec
ARIMA(0,0,0)(0,1,0)[6]	intercept	374.693	0.02 sec
ARIMA(1,0,0)(1,1,0)[6]	intercept	376.968	0.06 sec
ARIMA(0,0,1)(0,1,1)[6]	intercept	371.163	0.06 sec
ARIMA(0,0,0)(0,1,0)[6]	intercept	394.812	0.02 sec
ARIMA(2,0,2)(0,1,1)[6]	intercept	375.474	0.16 sec
ARIMA(2,0,2)(1,1,0)[6]	intercept	374.560	0.17 sec
ARIMA(2,0,2)(2,1,1)[6]	intercept	370.910	0.67 sec
ARIMA(2,0,2)(1,1,2)[6]	intercept	374.267	0.50 sec
ARIMA(2,0,2)(0,1,0)[6]	intercept	371.464	0.22 sec
ARIMA(2,0,2)(0,1,2)[6]	intercept	375.570	0.19 sec
ARIMA(2,0,2)(2,1,0)[6]	intercept	376.186	0.25 sec
ARIMA(2,0,2)(2,1,2)[6]	intercept	inf	0.71 sec
ARIMA(1,0,2)(1,1,1)[6]	intercept	374.925	0.22 sec
ARIMA(2,0,1)(1,1,1)[6]	intercept	374.763	0.23 sec
ARIMA(3,0,2)(1,1,1)[6]	intercept	inf	0.57 sec
ARIMA(2,0,3)(1,1,1)[6]	intercept	378.649	0.48 sec
ARIMA(1,0,1)(1,1,1)[6]	intercept	374.413	0.22 sec
ARIMA(1,0,3)(1,1,1)[6]	intercept	374.852	0.36 sec
ARIMA(3,0,1)(1,1,1)[6]	intercept	inf	0.49 sec
ARIMA(3,0,3)(1,1,1)[6]	intercept	inf	0.58 sec
ARIMA(2,0,2)(1,1,1)[6]	intercept	370.998	0.45 sec

Best model: ARIMA(2,0,2)(1,1,1)[6] intercept
Total fit time: 7.454 seconds

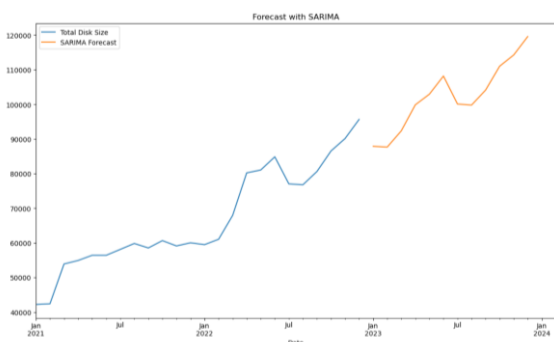
Gambar 9 Pencarian Model ARIMA

Dari model yang didapatkan dilakukan pengujian peramalan dengan menggunakan data training. Setelah itu dibandingkan hasil peramalan awal yang didapatkan dengan data testing. Grafik hasil perbandingan data training, data testing dan hasil peramalan awal ditunjukkan pada Gambar 10.



Gambar 10 Grafik Perbandingan Data Metode ARIMA

Proses selanjutnya adalah dengan melakukan peramalan atas seluruh data dengan menggunakan model yang dihasilkan dengan metode ARIMA. Grafik hasil peramalan pada tahun 2023 ditunjukkan pada Gambar 11.



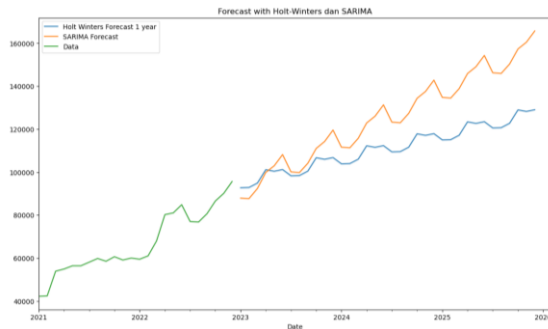
Gambar 11 Grafik Hasil Peramalan Dengan Metode ARIMA.

3.5. Evaluation

Berdasarkan hasil penelitian, dilakukan evaluasi dengan menghitung Root Mean Square Error (RMSE) untuk membandingkan nilai masing-masing algoritma peramalan.denga menggunakan metode Holt-Winters dan ARIMA. Dengan hasil MEAN 78437,4375, didapatkan hasil RMSE dengan metode Holt-Winters adalah 14332,661717740748 dan RMSE dengan metode ARIMA adalah 20498,977982137938..

3.6. Deployment

Pada tahap akhir dilakukan deployment dengan melakukan implementasi metode Holt-Winters dan ARIMA untuk meramalkan utilisasi storage server database untuk tiga (3) tahun kedepan yaitu tahun 2024, 2025 dan 2026. Grafik hasil peramalan digambarkan pada Gambar 12.



Gambar 12 Grafik Hasil Peramalan.

4. DISKUSI

Dari hasil penelitian didapatkan hasil kebutuhan storage server database Nadine pada Bulan Desember 2023 adalah sebesar 106.755,69 GB untuk peramalan dengan metode Holt-Winter dan 119.474,80 GB untuk peramalan dengan menggunakan metode ARIMA. Sedangkan pada tahun akhir tahun 2024 diramalkan dibutuhkan 117.888,01 GB dengan perhitungan metode Holt-Winter dan 142.564,17 dengan metode ARIMA. Dan pada Bulan Desember 2025 didapatkan hasil peramalan dengan nilai 129.020,33 GB untuk utilisasi storage server database dengan peramalan metode Holt-Winter dan 165.275,12 GB untuk peramalan menggunakan metode ARIMA.

Kedua hasil peramalan memiliki trend yang sama naik, namun hasil peramalan menggunakan metode Holt-Winter menghasilkan trend kenaikan yang lebih landai dibandingkan trend yang dihasilkan oleh peramalan menggunakan metode ARIMA.

Sedangkan dari hasil evaluasi yang dilakukan didapatkan hasil yaitu peramalan menggunakan metode Holt-Winter memiliki nilai RMSE yang lebih rendah dibandingkan menggunakan metode ARIMA.

Perbedaan penelitian terdahulu yang telah dilakukan tinjauan studi pada awal fase dengan penelitian yang penulis lakukan adalah objek penelitian pada utilisasi storage server database serta subjek penelitian yang dilakukan pada instansi pemerintah. Beberapa penelitian terdahulu melakukan peramalan beban kerja database untuk mendapatkan perkiraan performa database di masa depan. Sedangkan pada penelitian ini dilakukan peramalan utilisasi storage server database untuk mengetahui kebutuhan *resource* atau ketersediaan storage server database.

Perbedaan juga terdapat pada dataset yang digunakan pada penelitian ini berupa dataset time series dengan rentang waktu data selama 24 bulan yang memiliki pola musiman dan trend. Dan ditambahkan dengan analisis perbandingan peramalan utilisasi storage server database dengan metode Holt-Winters dan ARIMA sehingga didapatkan metode yang sesuai dengan kebutuhan penyusunan analisis kapasitas TIK di Pusintek Kemenkeu.

5. KESIMPULAN

Pada penelitian ini, dilakukan peramalan utilisasi storage server database dengan metode Holt-Winters dan ARIMA. Kedua algoritma digunakan untuk membandingkan akurasi mana yang lebih baik dalam memprediksi kebutuhan storage dengan mengukur RMSE. Dari hasil evaluasi dapat disimpulkan bahwa peramalan utilisasi storage server database dengan menggunakan metode Holt-Winters lebih baik daripada metode ARIMA.

Hasil dari peramalan ini dikomunikasikan ke Pusintek selaku user, sehingga dapat digunakan sebagai salah satu acuan untuk memprediksi kebutuhan storage server pada periode berikutnya yang bermanfaat dalam melakukan penyusunan analisis kapasitas TIK khususnya pada perencanaan kebutuhan storage server database.

Dalam penelitian ini masih didapatkan banyak hal yang dapat dikembangkan lagi. Saran yang dapat kami sampaikan kepada peneliti berikutnya yang ingin melanjutkan penelitian ini adalah melakukan penelitian terkait penggunaan algoritma-algoritma lain untuk meramalkan utilisasi storage server dan melakukan penelitian pada utilisasi objek lain yang sejenis dengan kapasitas.

DAFTAR PUSTAKA

- [1] M. Taufiq, D. Sugiarto, and A. Rochman, "Peramalan Utilisasi Perangkat Jaringan dan Bandwidth Dengan Metode Holt-Winters dan Multi Layer Perceptron," *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 42–51, 2020, doi: 10.31937/ti.v12i1.1575.
- [2] E. P. Mariz *et al.*, "Sistem Informasi Prediksi Stok Barang Menggunakan Metode Holt Winters Pada Cv Budi Jaya Mandiri," *J. Ilm. Fak. Ilk. Univ. Bina Darma*, no. April 2018, pp. 0–11, 2017.
- [3] L. Nababan *et al.*, "PENGUNAAN METODE WINTER EXPONENTIAL SMOOTHING," *J. Sist. Inf. Kaputama*, vol. 6, no. 2, pp. 373–381, 2022.
- [4] A. Asrirawan, S. U. Permata, and M. I. Fauzan, "Pendekatan Univariate Time Series Modelling untuk Prediksi Kuartalan Pertumbuhan Ekonomi Indonesia Pasca Vaksinasi COVID-19," *Jambura J. Math.*, vol. 4, no. 1, pp. 86–103, 2022, doi: 10.34312/jjom.v4i1.11717.
- [5] R. Rumini and N. Norhikmah, "Perbandingan Metode Arima Dan Exponential Smoothing Holt-Winters Untuk Peramalan Data Kunjungan," *Sistemasi*, vol. 9, no. 3, p. 622, 2020, doi: 10.32520/stmsi.v9i3.975.
- [6] I. R. Akolo, "Perbandingan Exponential Smoothing Holt-Winters Dan Arima Pada Peramalan Produksi Padi Di Provinsi Gorontalo," *J. Technopreneur*, vol. 7, no. 1,

- pp. 20–26, 2019, doi: 10.30869/jtech.v7i1.314.
- [7] N. R. Wiyanti, P. S. Manajemen, K. Bekasi, and L. Belakang, “Analisis Perbandingan Model Arima Dan Exponential Smoothing Dalam Memprediksi,” *JBME J. Bisnis Manaj. dan Ekon.*, vol. 19, no. 2, pp. 1–10, 2019.
- [8] T. Safitri, N. Dwidayati, and K. Kunci, “Perbandingan Peramalan Menggunakan Metode Exponential Smoothing Holt-Winters dan Arima,” *Unnes J. Math.*, vol. 6, no. 1, pp. 48–58, 2017, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [9] E. Staviana, Kusriani, and E. T. Luthfi, “Sistem Peramalan Kebutuhan Stok Obat Menggunakan Metode Holt-Winters,” *J. Informa, J. Penelit. Dan Pengadain Masyarakat, Politek. Indonusa, Surakarta*, vol. 5, 2019, doi: <https://doi.org/10.46808/informa.v5i4.154>.
- [10] A. T. Namel, M. A. Sahib, and S. M. Hasan, “Bandwidth Utilization Prediction in LAN Network Using Time Series Modeling,” *Iraqi J. Comput. Commun. Control Syst. Eng.*, pp. 78–89, 2019, doi: 10.33103/uot.ijccce.19.2.9.
- [11] Ema Utami; Hanif Al Fatta; Muhammad Surahmanto, “Prediksi Kapasitas Kargo Pada Bandara Deo Sorong,” *Insect (Informatics Secur. J. Tek. Inform.*, vol. 8, no. 1, pp. 58–66, 2022, [Online]. Available: <http://www.ejournal.um-sorong.ac.id/index.php/insect/article/view/1912>
- [12] D. Kurniawati, R. C. Buwono, and A. M. M. D. . Satriya, “EKSPLOKASI SINGLE MOVING AVERAGE UNTUK PERAMALAN PERSEDIAAN BARANG,” *JIRK. J. Innov. Res. Knowl.*, vol. 2, no. 5, pp. 2239–2246, 2022.
- [13] Harsiti, Z. Muttaqin, and E. Srihartini, “Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet,” *JSil (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 12–16, 2022, doi: 10.30656/jsii.v9i1.4426.
- [14] Mustika *et al.*, *Data Mining dan Aplikasinya*. Penerbit Widina Bhakti Persada Bandung, 2021.
- [15] M. A. Witten, I.H., Eibe, F.and Hall, *Data mining Practical Machine Learning Tools and Techniques 3rd Edition*. Elsevier, 2011.
- [16] B. Santosa, *Data mining: Teknik pemanfaatan data untuk keperluan bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [17] J. P. Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques*. Elsevier, 2012.
- [18] M. N. F. Rahmi Roza, *Tutorial Sistem Informasi Prediksi Jumlah Pelanggan Menggunakan Metode Regr... - Google Books*. Kreatif Industri Nusantara, 2020. [Online]. Available: https://www.google.co.id/books/edition/Tutorial_Sistem_Informasi_Prediksi_Jumla/ixH9DwAAQBAJ?hl=id&gbpv=1&dq=XAMPP+adalah+perangkat+lunak+bebas,+yang+mendukung+banyak+sistem+operasi,+merupakan+kompilasi+dari+beberapa+program.+XAMPP+merupakan+tool+yang+men
- [19] W. Y. Rusyida, *Teknik Peramalan: Metode ARIMA dan Holt Winter*. Penerbit NEM, 2022. [Online]. Available: <https://books.google.co.id/books?id=z9-mEAAAQBAJ>
- [20] D. Nugraha, I. P. Wardana, and M. O. Adnyana, “Pendugaan Produksi dan Tantangan Usahatani Kedelai di Indonesia Menggunakan Metode ARIMA,” *J. Penelit. Pertan. Tanam. Pangan*, vol. 2, no. 3, p. 155, 2019, doi: 10.21082/jpftp.v2n3.2018.p155-163.
- [21] D. T. Larose, *Data Mining Methods & Models*. Wiley-Interscience, 2006. [Online]. Available: <https://books.google.co.id/books?id=xveeIUOdkfEC>
- [22] R. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS inc, 2000.