



INNOVATIVE: Journal Of Social Science Research

Volume 3 Nomor 6 Tahun 2023 Page 6405-6421

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Peramalan Konsumsi Gas Alam Amerika Serikat dengan *Double Seasonality* menggunakan *Singular Spectrum Analysis (SSA)*

Qurnia Amanah Dwiadi^{1✉}, Ayu Indriani², Theresia Samaria Nauli³, Hani Nurhapilah⁴, Gumgum Darmawan⁵

Universtas Padjadjaran

Email: qurnia23001@mail.unpad.ac.id[✉]

Abstrak

Gas alam memainkan peran penting dalam memenuhi kebutuhan energi Amerika Serikat, yang merupakan salah satu konsumen terbesar di dunia. Konsumsi gas alam terus meningkat sejak implementasi proyek "*An America First Energy Plan*" pada tahun 2017. Peramalan yang akurat tentang jumlah konsumsi gas alam sangat dibutuhkan. Dalam penelitian ini, kita akan menggunakan model *Singular Spectrum Analysis (SSA)*. Metode ini tidak memerlukan pemenuhan asumsi parametrik dan diterapkan dengan baik pada data musiman. SSA dapat menggambarkan pola tren dan komponen lainnya dengan struktur sederhana. Konsep utamanya adalah 'pemisahan' yang mengkarakterisasi seberapa baik komponen berbeda dapat dipisahkan satu sama lain. SSA terdiri dari dua tahap yang saling melengkapi, yaitu tahap dekomposisi dan tahap rekonstruksi. Dari hasil pengujian, data konsumsi gas alam yang digunakan dalam penelitian ini memiliki pola musiman. Berdasarkan analisis model tersebut, yang memberikan nilai MAPE sebesar 1,62 % , dengan hasil peramalan yang relatif konstan setiap tahunnya.

Kata Kunci: *Peramalan, Gas Alam, Singular Spectrum Analysis (SSA)*

Abstract

Natural gas plays an important role in meeting the energy needs of the United States, which is one of the world's largest consumers. Natural gas consumption has continued to increase since the implementation of the "An America First Energy Plan" project in 2017. Accurate forecasting of the amount of natural gas consumption is urgently needed. In this research, we will use the Singular Spectrum Analysis (SSA) model. This method does not require fulfilling parametric assumptions and is well applied to seasonal data. SSA can describe trend patterns and other components with a simple structure. The key concept is 'separability' which characterizes how well different components can be separated from each other. SSA consists of two complementary stages, namely the decomposition stage and the reconstruction stage. From the test results, the natural gas consumption data used in this research has a seasonal pattern. Based on the model analysis, which gives a MAPE value of 1.62%, with forecasting results that are relatively constant every year.

Keywords: *Forecasting, Natural Gas, Singular Spectrum Analysis (SSA)*

PENDAHULUAN

Gas alam memiliki peran yang sangat penting dalam menyokong kebutuhan energi di Amerika Serikat. Hal ini dikarenakan gas alam merupakan salah satu energi utama yang digunakan untuk keperluan pembangkit listrik, pemanas, dan sektor-sektor krusial lain di Amerika Serikat. Konsumsi gas alam di negara ini terus meningkat dan menjadikannya sebagai negara dengan konsumsi gas alam terbesar kedua di dunia (Lu et al., 2020). Peningkatan konsumsi gas alam yang signifikan terjadi sejak dilakukan project "An America First Energy Plan" pada tahun 2017, dimana jumlah konsumsi gas alam yang semula 27,13 *Trillion Cubic Feet* meningkat menjadi 30,13 *Trillion Cubic Feet* dan terus meningkat hingga mencapai 32,28 *Trillion Cubic Feet* pada tahun 2022. Oleh karena itu, peramalan konsumsi gas alam sangat krusial sebagai acuan perencanaan strategis dan pengelolaan sumber daya energi yang efisien dalam upaya menjaga stabilitas pasokan energi yang keberlanjutan. Berbagai pendekatan metode dan model peramalan telah diaplikasikan untuk memperoleh hasil peramalan konsumsi gas alam yang akurat, termasuk pendekatan klasik, modern, maupun *hybrid*.

Sejumlah pendekatan klasik untuk meramalkan konsumsi gas alam telah dilakukan, diantaranya adalah model ARIMA untuk meramalkan konsumsi gas alam di Pakistan (Hussain et al., 2022), model SARIMA untuk meramalkan konsumsi gas alam musiman di China (Xu et al., 2023), model Holt-Winters dan ARIMA/SARIMA untuk meramalkan konsumsi gas alam jangka menengah di Turki (Akpınar & Yumuşak, 2020). Pendekatan modern menggunakan *machine learning* juga banyak diaplikasikan, di antaranya *multiple linear regression* (MLR), *artificial neural network* (ANN) dan *support vector regression* (SVR)

untuk meramalkan konsumsi gas alam di provinsi Istanbul (Turki) (Beyca et al., 2019), serta pendekatan ANN dengan *multilayer perceptron model* (MLP) untuk meramalkan konsumsi gas alam di Szczecin (Polandia) (Szoplik, 2015). Sejumlah pendekatan hybrid juga dilakukan untuk memperoleh akurasi peramalan yang lebih tinggi, seperti hybrid model ARIMA - LSTM untuk meramalkan konsumsi gas alam di Polandia (Manowska et al., 2021), CNN-LSTM untuk meramalkan konsumsi gas alam harian di China (Du et al., 2022), dan lainnya.

Peramalan konsumsi gas alam di AS juga telah dilakukan sebelumnya melalui berbagai pendekatan, seperti pendekatan model SARIMA dan SARIMAX (Manigandan et al., 2021), *Deep Neural Network*, *Long Short-Term Memory Neural Network*, dan *Gated Recurrent Unit Neural Network* (Zu et al., 2023), ANN dan SVM (Singh et al., 2023), *gray model* dengan variabel *dummy* (Z.-X. Wang et al., 2021). Meskipun berbagai pendekatan telah dilakukan, namun pola yang rumit dalam data konsumsi gas alam AS masih sulit untuk diatasi karena data tersebut mengandung pola tren dan multi-musiman yang terkait dengan faktor musim dan pola konsumsi. Oleh karena itu, penelitian ini mengintegrasikan pendekatan *Singular Spectrum Analysis* (SSA) pada peramalan konsumsi gas alam AS untuk meningkatkan keakuratannya.

Singular Spectrum Analysis (SSA) adalah teknik yang berakar dalam teori dekomposisi Karhunen-Loève dan menjadi populer karena kemampuannya untuk mengidentifikasi dan menggambarkan karakteristik sinyal dalam domain waktu-frekuensi (J. Wang et al., 2013). SSA menjadi salah satu teknik non-parametrik yang banyak digunakan dalam analisis data runtun waktu pendek maupun panjang dan telah diterapkan di berbagai disiplin ilmu, termasuk ekonomi, data iklim, dan keuangan (Abrego-Perez Adriana & Penagos-Londoño, 2022). SSA telah dikembangkan menjadi sebuah metode yang dapat digunakan untuk menganalisis variasi, periodisitas, dan tren dalam deret waktu (Broomhead & King, 1986)(Fraedrich, 1986), hingga saat ini SSA juga telah dikembangkan menjadi model prediksi (Lisi et al., 1995). Keunggulan SSA terletak pada kemampuannya untuk mendekomposisi pola-pola deret waktu kompleks yang melibatkan tren, beberapa komponen musiman, dan siklis maupun komponen noise secara rinci yang disebabkan oleh pengaruh variasi kalender (Sulandari et al., 2020). Dengan demikian, penggunaan SSA dalam konteks peramalan konsumsi gas alam Amerika Serikat menjadi relevan dan berpotensi memberikan hasil yang akurat.

METODE PENELITIAN

Data Penelitian

Pada penelitian ini, model yang digunakan adalah model persamaan *Singular Spectrum Analysis* (SSA). Data yang digunakan adalah data sekunder yang diperoleh dari Energy Information Administration (EIA) Amerika Serikat. Objek penelitian ini adalah jumlah konsumsi gas alam di Amerika Serikat (*Billion Cubic Feet*). Data yang digunakan dalam peramalan ini berdasarkan data deret waktu dengan periode bulanan konsumsi gas alam Januari 2010 - Desember 2022.

Singular Spectrum Analysis (SSA)

1. *Embedding* dalam Tahap Dekomposisi

Embedding adalah proses transformasi data deret waktu menjadi matriks lintasan dengan cara mentransformasikan data dimensi satu (vektor) menjadi data multidimensional (matriks). $X = \{x_1, x_2, \dots, x_N\}$ merupakan data deret waktu sepanjang N tanpa data hilang. Data tersebut ditransformasi menjadi matriks berdimensi $L \times K$, dimana nilai L (*window length*) menjadi baris matriks yang memenuhi syarat $2 < L < \frac{N}{2}$ dan nilai K (*lagged vectors*) menjadi kolom matriks dengan vektor lag $X_i = (x_i, \dots, x_{i+L-1}) \in R^L, i = 1, \dots, K$. Nilai L diperoleh dengan proses *trial and error*, sedangkan nilai K dapat dihitung dengan rumus berikut:

$$K = N - L + 1 \quad (1)$$

Matriks lintasan data diperoleh berdasarkan vektor lag, sehingga bentuk matriks lintasan partisi kolomnya adalah sebagai berikut:

$$X_1 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{1+L-1} \end{pmatrix}; X_2 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{2+L-1} \end{pmatrix}; \dots; X_K = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{K+L-1} \end{pmatrix} \quad (2)$$

Maka, bentuk matriks lintasan X adalah sebagai berikut:

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix} \quad (3)$$

2. *Singular Value Decompositioin* (SVD) dalam Tahap Dekomposisi

Singular Value Decompositioin (SVD) atau Dekomposisi Nilai Singular bertujuan untuk memperoleh pemisahan komponen dalam dekomposisi dari data deret waktu. SVD dimulai dengan menentukan nilai *eigenvalue* ($\lambda_1, \lambda_2, \dots, \lambda_L$) dari matriks simetris $S = \mathbf{X}\mathbf{X}^T$ di mana $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L > 0$, serta *eigenvector* (u_1, u_2, \dots, u_L) dari matriks S tersebut. Langkah-langkah SVD yaitu:

- 1) Definisikan matriks $\mathbf{S} = \mathbf{X}\mathbf{X}^T$

Dari perkalian matriks $\mathbf{X}\mathbf{X}^T$ menghasilkan matriks simetris \mathbf{S} berukuran $L \times L$ yang juga merupakan matriks kovarians.

- 2) Menentukan *eigenvalue* dan *eigenvector* pada matriks \mathbf{S}

Perhitungan *eigenvalue* dan *eigenvector* dari matriks persegi \mathbf{S} dapat dilakukan dengan menggunakan persamaan:

$$\det(\mathbf{S} - \lambda\mathbf{I}) = 0 \quad (4)$$

selanjutnya akan diperoleh *eigenvalue* λ_i untuk $i = 1, 2, \dots, L$ dan $\mathbf{U}_i = (u_1, u_2, \dots, u_L)$ adalah *eigenvector* dari *eigenvalue* yang bersesuaian.

- 3) Membuat matriks nilai singular

Matriks *singular value* adalah matriks diagonal dengan diagonal utama berisi akar positif dari *eigenvalue* ($\sqrt{\lambda_i}$) dengan urutan menurun $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L > 0$

Maka,

$$\sqrt{\lambda_i} = \begin{bmatrix} \sqrt{\lambda_i} & \dots & 0_L \\ \vdots & \ddots & \vdots \\ 0_L & \vdots & \sqrt{\lambda_L} \end{bmatrix} \quad (5)$$

- 4) Membuat matriks *principal component*

Misalkan \mathbf{S} matriks real berukuran $L \times L$ bilangan real positif σ disebut *singular value* dari matriks \mathbf{S} jika ada vektor tak nol $\mathbf{U} \in R^m$ dan $\mathbf{V} \in R^m$ sehingga $\mathbf{X}\mathbf{V} = \sigma_i\mathbf{U}$ dan $\mathbf{X}^T\mathbf{U} = \sigma_i\mathbf{V}$. Vektor \mathbf{U} disebut vektor singular kiri dan \mathbf{V} disebut vektor singular kanan. Selanjutnya (σ, \mathbf{U}) disebut pasangan singular kiri dari \mathbf{S} dan (σ, \mathbf{V}) disebut pasangan singular kanan dari \mathbf{S} (Ariyanti, 2010).

Dari definisi di atas diperoleh persamaan berikut:

$$\mathbf{X}\mathbf{V}_i = \sigma_i\mathbf{U}_i \text{ dan } \mathbf{X}^T\mathbf{U}_i = \sigma_i\mathbf{V}_i \quad (6)$$

Untuk Vektor \mathbf{V}_i dengan $i = (1, 2, \dots, L)$ mengikuti persamaan berikut:

$$\begin{aligned} \mathbf{X}^T\mathbf{U}_i &= \sigma_i\mathbf{V}_i \\ \mathbf{X}^T\mathbf{U}_i &= \sqrt{\lambda_i}\mathbf{V}_i \\ \mathbf{V}_i &= \frac{\mathbf{X}^T\mathbf{U}_i}{\sqrt{\lambda_i}} \end{aligned} \quad (7)$$

\mathbf{V}_i^T merupakan transpose dari matriks \mathbf{V}_i

$$\mathbf{V}_i^T = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1K} \\ v_{21} & v_{22} & \dots & v_{2K} \\ \vdots & \dots & \ddots & \vdots \\ v_{K1} & v_{K2} & \dots & v_{KK} \end{bmatrix} \quad (8)$$

Setelah diperoleh *singular value* (σ_i), *vector eigen* (\mathbf{U}_i) dan *principal component* (\mathbf{V}_i^T), maka SVD dari matriks lintasan \mathbf{X} adalah

$$\mathbf{X} = \mathbf{U}_i \sqrt{\lambda_i} \mathbf{V}_i^T \quad (9)$$

Dengan

\mathbf{U}_i : matriks orthornormal berukuran $L \times L$

$\sqrt{\lambda_i}$: matrik diagonal berukuran $L \times K$

\mathbf{V}_i^T : matriks orthornormal berukuran $K \times K$

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1L} \\ u_{21} & u_{22} & \dots & u_{2L} \\ \vdots & \dots & \ddots & \vdots \\ u_{L1} & u_{L2} & \dots & u_{LL} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_L \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1K} \\ v_{21} & v_{22} & \dots & v_{2K} \\ \vdots & \dots & \ddots & \vdots \\ v_{K1} & v_{K2} & \dots & v_{KK} \end{bmatrix} \\ &= \mathbf{U}_1 \sigma_1 \mathbf{V}_1^T + \mathbf{U}_2 \sigma_2 \mathbf{V}_2^T + \dots + \mathbf{U}_d \sigma_d \mathbf{V}_d^T \\ &= \sum_{i=1}^d \mathbf{U}_i \sigma_i \mathbf{V}_i^T \text{ untuk } i = 1, 2, \dots, d \text{ dan } d = \max\{i\} \text{ serta } \sigma_i > 0. \end{aligned} \quad (10)$$

Ketiga komponen dalam matrix \mathbf{X}_i yaitu *singular value* (σ_i), *eigenvector* (\mathbf{U}_i) dan *principal component* (\mathbf{U}_i^T) disebut *eigen triple* ke- i dari SVD. Sehingga SVD untuk matriks \mathbf{X}_i dapat dituliskan sebagai berikut (Darmawan, 2016):

$$\mathbf{X}_i = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d. \quad (11)$$

3. *Grouping* dalam Tahap Rekonstruksi

Dalam tahap ini, matriks lintasan dengan dimensi $L \times K$ dibagi menjadi beberapa sub-grup, seperti pola tren, musiman, periodik, dan noise yang merupakan pola pembentuk data deret waktu. Proses pengelompokan ini terkait erat dengan memisahkan matriks \mathbf{X}_i menjadi beberapa kelompok dan melakukan penjumlahan matriks di setiap kelompok yang dapat ditulis $\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$. Dimana matriks \mathbf{X}_i akan dibagi menjadi m sub-set yang saling tidak berhubungan, yang kita sebut sebagai $I = \{I_1, \dots, I_m\}$ (Darmawan, 2016).

Ada dua cara yang dapat digunakan untuk mengelompokkan komponen, yaitu:

- (1) Memeriksa grafik satu dimensi dari serangkaian data yang telah direkonstruksi untuk mengenali komponen tren dan musiman. Komponen-komponen yang menunjukkan perubahan yang lambat dalam grafik akan dikelompokkan sebagai komponen tren, sementara jika terdapat pola dan periode musiman yang serupa, komponen-komponen tersebut akan digabungkan dalam kelompok musiman.
- (2) Mengamati grafik nilai singular. Jika nilai-nilai tersebut menurun perlahan, ini umumnya menandakan keberadaan komponen noise dalam data.

Metode SSA menekankan konsep "separabilitas" yang menunjukkan sejauh mana komponen-komponen dengan karakteristik yang berbeda dapat dipisahkan satu sama lain. Keberhasilan dalam proses dekomposisi data deret waktu dapat diukur dengan melihat plot korelasi W , yang menggambarkan sejauh mana korelasi antara kelompok-kelompok. Nilai

korelasi yang semakin tinggi ditunjukkan pada warna pada plot korelasi W akan semakin gelap, sedangkan nilai korelasi yang rendah akan menampilkan warna yang lebih terang.

Nilai korelasi yang mendekati nol menandakan semakin kecil nilai korelasi antara dua komponen atau kelompok, maka dua komponen tersebut dapat dipisahkan (Irwan et al., 2022). Nilai korelasi W dapat dihitung menggunakan persamaan berikut:

$$r_{12}^{(u)} = \frac{(A_N^{(i)} \cdot A_N^{(j)})_u}{\sqrt{(A_N^{(i)} \cdot A_N^{(j)})_u (A_N^{(i)} \cdot A_N^{(j)})_u}} \quad (12)$$

dengan

$$(A_N^{(i)} \cdot A_N^{(j)})_u = \sum_{p=1}^N u_p^{L,N} \cdot y_p^{(i)} \cdot y_p^{(j)} \quad (13)$$

dan

$$u_p^{L,N} = \frac{n+1}{2} - \frac{\left| \frac{N+1}{2} - L \right|}{2} - \frac{\left| \frac{N+1}{2} - P \right|}{2} - \frac{\left| \left| \frac{N+1}{2} - L \right| - \left| \frac{N+1}{2} - P \right| \right|}{2} \quad (14)$$

4. *Diagonal Averaging* dalam Tahap Rekonstruksi

Proses *Diagonal Averaging* yaitu dengan merekonstruksi masing-masing matriks yang terdapat pada matriks X saat *grouping* menjadi data deret waktu yang baru dengan panjang N. Tujuan dari tahap ini mendapatkan *singular value* dari komponen-komponen yang telah dipisahkan, kemudian akan digunakan dalam peramalan. Hasil pada tahap ini merupakan matriks F sebagai berikut:

$$F = \begin{pmatrix} f_{11} & f_{21} & \dots & f_k \\ f_{21} & f_{22} & \dots & f_{k+1} \\ \vdots & \dots & \ddots & \vdots \\ f_L & f_{L+1} & f_{11} & f_N \end{pmatrix} \quad (15)$$

Kemudian mencari rata-rata diagonal matriks dapat dipakai persamaan berikut

$$g_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k f_{m,k-m+1}^*; \text{ untuk } 1 \leq k \leq L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*-1} f_{m,k-m+1}^*; \text{ untuk } L^* \leq k \leq K^* \\ \frac{1}{N-k+1} \sum_{m=K-K^*+1}^{N-K^*+1} f_{m,k-m+1}^*; \text{ untuk } K^* \leq k \leq N \end{cases} \quad (16)$$

Dimana $L^* = \min(L, K)$ dan $K^* = \max(L, K)$ pada persamaan diatas jika di aplikasikan kedalam matriks X_{im} akan membentuk deret $\tilde{Y}^{(k)} = (\tilde{Y}_1^{(k)}, \dots, \tilde{Y}_N^{(k)})$. Maka deret akan dikomposisi menjadi jumlah dari m deret:

$$y_n = \sum_{k=1}^m \tilde{Y}_N^{(k)} \quad (17)$$

Periodegram

Tahapan-tahapan dalam menguji pola musiman dengan periodogram adalah sebagai berikut:

- 1) Menentukan data yang akan diuji.
- 2) Persamaan (2), pada dasarnya, secara prinsip, koefisien Fourier serupa dengan

koefisien dalam regresi standar. Oleh karena itu, ketika diterapkan pada data, kita dapat menyusun persamaan Fourier sebagai berikut:

$$Y_t = \sum_{i=0}^{\lfloor n/2 \rfloor} a_k \cos \omega_i t + b_i \sin \omega_i t \quad (18)$$

dimana $i = 0, 1, \dots, \lfloor n/2 \rfloor$ dan ω_i adalah frekuensi fourier yang diperoleh dari perhitungan $\omega_i = 2\pi \cdot i/n$

- 3) Lakukan perhitungan untuk a_i dan b_i dengan perhitungan sebagai berikut:

$$a_i = \begin{cases} \frac{1}{n} \sum_{t=1}^n y_t \cos \omega_i t, & i = 0 \text{ dan } i = \frac{n}{2} \text{ jika } n \text{ genap} \\ \frac{2}{n} \sum_{t=1}^n y_t \cos \omega_i t, & i = 1, 2, \dots, \frac{(n-1)}{2} \end{cases} \quad (19)$$

$$b_i = \frac{2}{n} \sum_{t=1}^n y_t \sin \omega_i t, \quad i = 1, 2, \dots, \frac{(n-1)}{2} \quad (20)$$

- 4) Menghitung nilai ordinat $I(\omega_i)$ dengan perhitungan sebagai berikut:

$$I(\omega_i) = \begin{cases} na_0^2 & k = 0 \\ \frac{n}{2} (a_i^2 + b_i^2) & k = 1, 2, \dots, \frac{(n-1)}{2} \\ na_i^2 & \text{dan } k = \frac{n}{2} \text{ ketika } n \text{ genap} \end{cases} \quad (21)$$

- 5) Melakukan uji hipotesis untuk melihat letak pola musiman mengujikan nilai uji T.

Hipotesis Statistik:

$$H_0: I^{(1)}(\omega_1) = 0 \quad (\text{periodisitas tidak berarti})$$

$$H_1: I^{(1)}(\omega_1) \neq 0 \quad (\text{periodisitas berarti})$$

Statistik Uji:

$$T = \frac{I^{(1)}(\omega_1)}{\sum_{i=1}^{\lfloor n/2 \rfloor} I(\omega_i)} \quad (22)$$

dimana, $I^{(1)}(\omega_1)$ merupakan ordinat maksimum untuk periodogram pada frekuensi fouries dan $I(\omega_1)$ merupakan nilai ordinat periodogram pada frekuensi fouries ke-i.

Kriteria uji:

Tolak H_0 jika $T_{hitung} > g_\alpha$, dan terima dalam hal lainnya.

- 6) Tentukan frekuensi dari hasil nilai periodogram yang terbesar. Frekuensi komponenn periodic didapat dari persamaan sebagai berikut.

$$\omega_i = \frac{2\pi i}{n}, i = 0, 1, 2, \dots, \frac{n}{2} \quad (23)$$

Peramalan (*Forecasting*)

Peramalan yang digunakan dalam penelitian ini adalah SSA *recurent*. Dalam membangun modelnya menggunakan *Linear Recurrent Formula* (LRF) (Darmawan, 2016). Dengan model persamaan sebagai berikut:

$$X_{i+d} = \sum_{k=1}^d r_k X_{i+d-k} \quad (24)$$

Di mana $1 \leq i \leq N - d$. Untuk menaksir koefisien LRF, yaitu (r_1, r_2, \dots, r_d) digunakan *eigenvector* yang diperoleh dari langkah SVD. Dengan $P = (p_1, p_2, \dots, p_{L-1}, p_L)^T$, $P^{\bar{v}} =$

$(p_1, p_2, \dots, p_{L-1})^T$, π_i komponen terakhir dari vektor $(p_1, p_2, \dots, p_{L-1}, p_L)$, dan $v^2 = \sum_{i=1}^{L-1} \pi_i^2$ maka koefisien LRF (vektor R) dapat dihitung dengan persamaan:

$$(r_{L-1}, \dots, r_1) = \frac{1}{1-v^2} \sum_{i=1}^{L-1} \pi_i P_i^p \quad (25)$$

Dalam peramalan SSA *forecasting* ini, deret waktu yang digunakan adalah deret hasil rekonstruksi yang berupa kombinasi linear komponen pertama dan vektor R. Kemudian akan ditentukan M buah titik data baru yang akan diramalkan.

$$g_i \begin{cases} \hat{X}_i, & 0 \leq i \leq N \\ \sum_{j=1}^{L-1} r_j g_{i-j}, & N+1 \leq i \leq N+M \end{cases} \quad (26)$$

Maka terbentuk deret hasil peramalan, yaitu $G_{N+M}(g_1, \dots, g_{N+M})$, di mana g_{N+1}, \dots, g_{N+M} adalah hasil ramalan dari SSA *forecasting*.

Evaluasi Peramalan

Untuk melihat kecocokan metode terhadap data. Dengan menggunakan ukuran ketepatan dan pengujian keandalan peramalan. ukuran ketepatan peramalan dipandang sebagai kriteria penolakan untuk memilih suatu metode peramalan sehingga dapat digunakan untuk menentukan kemungkinan yang lebih baik. Metode yang digunakan adalah *Mean Absolute Percentage Error* (MAPE) sebagai berikut :

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_T - \hat{Y}_t}{Y_t} \right| \times 100\% \quad (27)$$

Tabel 1. Tingkat Standar Evaluasi Model MAPE (%)

< 10%	Hasil peramalan sangat akurat
10-20%	Hasil peramalan baik
20-50%	Hasil peramalan wajar
> 50%	Hasil peramalan lemah dan tidak akurat

HASIL DAN PEMBAHASAN

Embedding

Berikut ini merupakan tabel nilai MAPE yang diperoleh pada beberapa nilai *windows length* (L) yang dicoba untuk menentukan nilai L optimum:

Tabel 2. *Trial and Error* Untuk Menemukan Nilai L Yang Optimal

L	7	12	22	27	32	37	42
MAPE	29,71	3,07	5,98	6,94	6,35	3,58	5,04
L	47	52	57	62	67	72	77
MAPE	3,35	3,002	4,12	6,03	5,43	4,88	6,65
L	82	87	92	97	102	107	112
MAPE	5,92	2,81	1,61	2,04	2,13	2,54	3,49
L	117	122	127	132	137	142	
MAPE	3,74	3,38	6,19	3,09	8,84	14,1	

Berdasarkan di atas, dapat dilihat bahwa nilai *window length* terbaik yang dapat digunakan dalam penelitian ini adalah 92 dengan MAPE terkecil sebesar 1,617%, dan nilai *lagged vectors* sebesar 53. Maka diperoleh matriks lintasan berukuran 92×53 sebagai berikut:

$$X = (x_{ij})_{i,j=1}^{92,53} = \begin{pmatrix} 2809,8 & 2481,0 & \dots & 1810,2 \\ 1481,0 & 2142,9 & \dots & 1745,4 \\ 2142,9 & 1691,8 & \dots & 1881,0 \\ \vdots & \vdots & \ddots & \vdots \\ 2093,8 & 1920,5 & \dots & 3007,3 \end{pmatrix}$$

Singular Value Decompositioin (SVD)

Pada tahap ini dilakukan perhitungan untuk mencari nilai *eigen triple* berdasarkan matriks lintasan $X_{(92 \times 53)}$. Langkah awal yang dilakukan yaitu membentuk matriks simetris $S = XX^T$ sebagai berikut:

$$S = X_{(92 \times 50)} X_{(50 \times 92)}^T = \begin{bmatrix} 248093238 & 243371826 & 23771826 & \dots & 262585799 \\ 243371826 & 243244683 & 239683810 & \dots & 261870776 \\ 23771826 & 239683810 & 240627483 & \dots & 258114358 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 280386037 & 277893810 & 276439173 & \dots & 300538676 \end{bmatrix}$$

Setelah diperoleh matriks simetris $S_{(34 \times 34)}$ selanjutnya dihitung nilai-nilai *eigen triple*:

1) *Singular Value* ($\sqrt{\lambda_i}$)

Dari matriks simetris $S_{(34 \times 34)}$ tersebut dihitung nilai *eigenvalue* dan selanjutnya dari nilai *eigenvalue* dihitung nilai *singular value* yang hasilnya ditampilkan pada Tabel 2 sebagai berikut:

Tabel 3. *Eigenvalues and Singular Values*

No	Eigenvalues	Singular value
1	2.577279e+10	2.577279e+10
2	2.762413e+08	2.762413e+08
3	2.698909e+08	2.698909e+08
⋮	⋮	⋮
91	-1.044479e-06	1.044479e-06
92	-2.535461e-06	2.535461e-06

2) *Eigenvector* (U_i)

Setelah mendapatkan nilai *singular value* selanjutnya menghitung nilai *eigenvector* yang hasilnya disajikan pada Tabel 3 sebagai berikut:

Tabel 4. *Eigenvectors*

No	U_1	U_2	...	U_{50}
1	-0.09634913	-0.1518297847	...	0.104307269
2	-0.09549247	-0.1236305377	...	-0.146491136

3	-0.09500770	-0.0633475477	...	0.077988876
⋮	⋮	⋮	⋮	⋮
92	-0.11457751	0.1219819983	...	0.073797847

3) Principal Component (V_i)

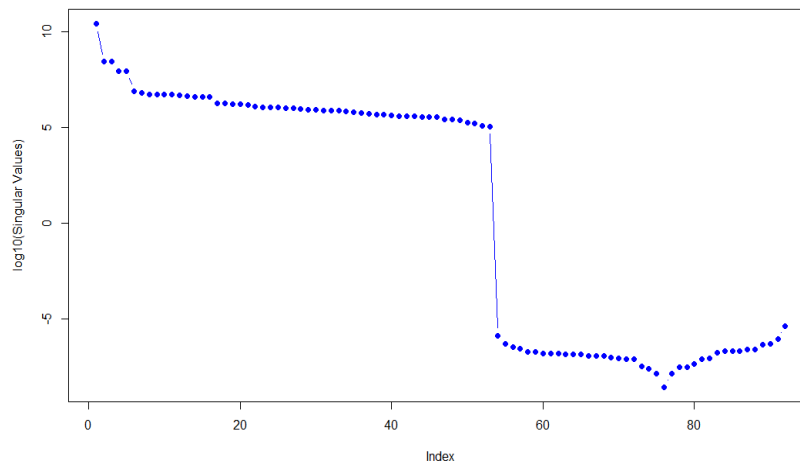
Setelah *eigenvalue* dan *eigenvector* diperoleh, maka selanjutnya dihitung nilai *principal component* yang hasilnya disajikan pada Tabel 4 sebagai berikut:

Tabel 5. *Principal Component*

No	V_1	V_2	...	V_{50}
1	-8.069001e-07	-1.118530e-05	...	-1.412437e-04
2	-8.039011e-07	-9.970008e-06	...	-1.012974e-04
3	-8.023196e-07	-5.999397e-06	...	3.594072e-04
⋮	⋮	⋮	⋮	⋮
53	-8.935982e-07	5,781290e-06	...	-3.68226e-04

Grouping dalam tahap rekonstruksi

Pada awal proses pengelompokan, dilakukan penyusunan *eigen triple* yang sudah dihasilkan dari langkah SVD. *Eigen triple* yang menunjukkan kesamaan karakteristik akan ditempatkan dalam satu grup atau komponen. Proses pengelompokan ini memanfaatkan analisis visual, dengan cara menampilkan grafik nilai singular dan grafik vektor eigen.

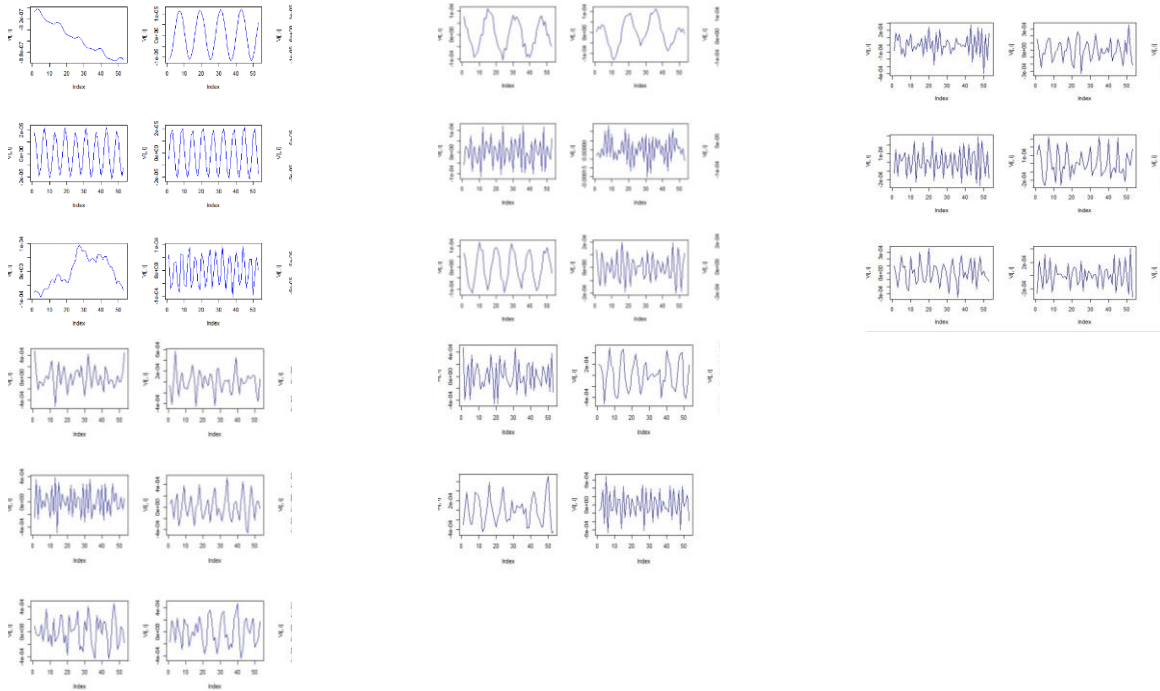


Gambar 1. Nilai Singular

Berdasarkan visualisasi pada gambar di atas, terlihat bahwa nilai singular menunjukkan penurunan yang lambat pada singular ke-53. Penurunan yang lambat pada nilai singular ini dapat dianggap sebagai indikasi adanya komponen noise. Berdasarkan hasil perhitungan menggunakan *package ssa* pada *software R*, *eigenvector* yang digunakan adalah 50. Hasil tersebut menunjukkan bahwa hasil secara *automatic* dan visualisasi tidak berbeda jauh.

Untuk melakukan pengelompokan dengan tingkat spesifik yang lebih tinggi, akan dilakukan identifikasi dengan cara menggambar *eigenvector* satu per satu. Setiap kelompok

akan berisi *eigenvektor* yang memiliki karakteristik serupa. Berikut adalah tampilan plot dari *eigenvektor* berdasarkan hasil dari *package ssa*, dengan $i = 1, 2, \dots, 50$:



Gambar 2. *Eigenvektor Plot*

Hasil pengelompokan ini menghasilkan dua kategori, yakni trend dan musiman. Berdasarkan ilustrasi dalam Gambar 2, dapat diperhatikan bahwa semua vector eigen membentuk pola data yang memiliki kemiripan grafik yang bersebelahan dan dikelompokkan sebagai pola data musiman.

Tabel 6. Hasil *Grouping*

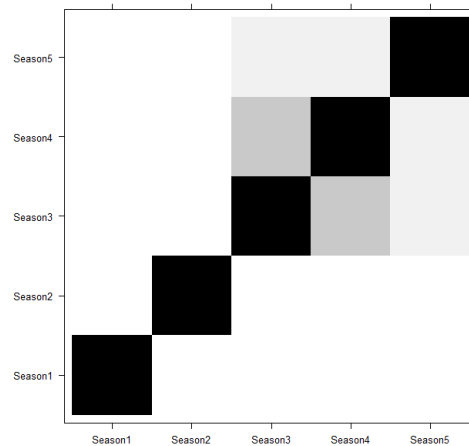
Group	Component	Pattern
1	1,2,3,6,7,10,11,15,16,21,24	Musiman
2	4,5,29,30,32,43,44,45,47,49	Musiman
3	8,9,19,20,33,34,36,37,38,39,41	Musiman
4	17,18,22,23,25,26,27,31,35,48,50	Musiman
5	12,13,14,28,40,42,46	Musiman

Hasil dari pengelompokan komponen berdasarkan kesamaan karakteristiknya, dapat dilihat dari plot *W-correlation*. Tampilan plot *W-correlation* adalah sebagai berikut:

Tabel 7. Korelasi Antar-Kelompok Dari Pola Yang Terbentuk

	Musiman (1)	Musiman (2)	Musiman (3)	Musiman (4)	Musiman (5)
Musiman (1)	1.0000	0.0007	0.0025	0.0021	0.0031
Musiman (2)	0.0007	1.0000	0.0262	0.0149	0.0082
Musiman (3)	0.0025	0.0262	1.0000	0.2146	0.0766
Musiman (4)	0.0021	0.0149	0.2146	1.0000	0.0540
Musiman (5)	0.0031	0.0082	0.0766	0.0540	1.0000

Berdasarkan Tabel 7, terlihat bahwa nilai korelasi antar kelompok yang tertinggi adalah 0.2146 yaitu korelasi antara musiman periode 3 dan musiman periode 4. Nilai korelasi yang tinggi berkisar antara 0.78 hingga 0.933 (Guilford, 1956). Maka, berdasarkan hasil pada tabel 2 tidak terdapat nilai korelasi yang tinggi. Hal ini menunjukkan keterpisahan antar setiap kelompok sudah baik dengan nilai korelasi antar kelompok yang cukup lemah. Secara grafis korelasi antar kelompok adalah sebagai berikut.



Gambar 3. Matriks Kolerasi W

Berdasarkan gambar 3 terlihat bahwa setiap komponen memiliki korelasi yang lemah dan menunjukkan hasil pengelompokkan yang sudah cukup baik. Hasil pengelompokkan akan dilanjutkan dalam proses *diagonal averaging*.

Diagonal Averaging

Setelah mengidentifikasi sejumlah pengelompokkan, pada tahap ini empat kelompok yang telah terbentuk sebelumnya disusun ulang menjadi deret tunggal yang baru dalam bentuk seri unidimensi. Diagonal averaging dapat dihitung dengan menggabungkan hasil rekonstruksi untuk setiap komponen. Selanjutnya, residual diperoleh melalui perbedaan antara data aktual dan hasil rekonstruksi.

Tabel 8. Hasil *Diagonal Averaging*

No	Data	Diagonal Averaging					Data Rekontruksi	Residual
		Musim 1	Musim 2	Musim 3	Musim 4	Musim 5		
1	2809.8	2581.09	277.27	26.706	-83.14	6.143	2808.0683	1.7314
2	2481.0	2463.60	176.05	87.251	-54.17	-12.58	2485.6561	-4.6561
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
144	3007.3	2995.75	81.896	9.4321	-64.48	-16.52	3006.0763	1.22361

Akurasi Peramalan

Akurasi peramalan metode SSA pada penelitian ini menggunakan metode *Mean Absolute Percentage error* (MAPE). Diperoleh nilai MAPE untuk data *out sample* yang disajikan dalam Tabel 9 berikut ini:

Tabel 9. Nilai MAPE pada Hasil Peramalan Data *Out Sample*

Tahun	Bulan	Nilai Error
2023	January	0.056655304
	February	-0.033292694
	March	-0.008784467
	April	0.005843769
	May	-0.0060092629
	June	0.005816415
	July	0.010238845
	August	0.005870780
	September	-0.010033621
	October	-0.007736991
	November	-0.036165310
	December	0.007698582
<i>Mean Absolute Percentage Error (MAPE)</i>		1.617887

Dapat dilihat pada Tabel 8 bahwa nilai MAPE yang diperoleh sebesar 1.617887 atau 1.62 %. Hasil ini menunjukkan bahwa nilai MAPE sangat akurat untuk meramalkan. Maka hal tersebut menunjukkan bahwa hasil peramalan dengan metode SSA dengan windows length 92 sangat akurat. Oleh karena itu, dengan windows length 92 dilakukan peramalan dari bulan Januari Tahun 2023 sampai dengan bulan Desember Tahun 2023. Hasil peramalannya adalah sebagai berikut:

Tabel 10. Hasil Peramalan

Tahun	Bulan	Nilai Ramalan
2023	January	3388.211
	February	3163.012
	March	2805.934
	April	2353.665
	May	2255.876
	June	2304.518
	July	2556.553
	August	2544.474
	September	2311.462
	October	2384.306
	November	23873.286
	December	3355.864

SIMPULAN

Data yang dianalisis pada penelitian ini adalah data konsumsi gas alam Amerika Serikat bulan Januari Tahun 2010 sampai bulan Desember Tahun 2022 dengan menggunakan metode *Singular Spectrum Analysis* (SSA). Berdasarkan hasil analisis diperoleh nilai *Mean Absolute Percentage error* (MAPE) sebesar 1.617887 atau 1.62%. Berdasarkan nilai MAPE tersebut maka didapatkan hasil ramalan konsumsi gas alam pada bulan Januari Tahun 2023 adalah sebesar 3388.211 *Billion Cubic Feet*, bulan Februari sebesar 3163.012 *Billion Cubic Feet*, bulan Maret sebesar 2805.934 *Billion Cubic Feet*, bulan April sebesar 2353.665 *Billion Cubic Feet*, bulan Mei sebesar 2255.876 *Billion Cubic Feet*, dan sampai dengan bulan Desember sebesar 3355.864 *Billion Cubic Feet*.

DAFTAR PUSTAKA

- Abrego-Perez Adriana, L., & Penagos-Londoño, G. I. (2022). Mixture modeling segmentation and singular spectrum analysis to model and forecast an asymmetric condor-like option index insurance for Colombian coffee crops. *Climate Risk Management, 35*, 100421. <https://doi.org/10.1016/J.CRM.2022.100421>
- Akpınar, M., & Yumuşak, N. (2020). Daily basis mid-term demand forecast of city natural gas using univariate statistical techniques. *Journal of the Faculty of Engineering and Architecture of Gazi University, 35*(2), 725–741. <https://doi.org/10.17341/gazimmfd.494094>
- Ariyanti, G. (2010). Dekomposisi nilai singular dan aplikasinya. *Seminar Nasional Matematika Dan Pendidikan Matematika, November 2010*, 33–39.
- Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., & Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics, 80*, 937–949. <https://doi.org/10.1016/j.eneco.2019.03.006>
- Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena, 20*(2–3), 217–236. [https://doi.org/10.1016/0167-2789\(86\)90031-X](https://doi.org/10.1016/0167-2789(86)90031-X)
- Darmawan, G. (2016). Identifikasi Pola Data Curah Hujan pada Proses Grouping dalam Metode Singular Spectrum Analysis. *Seminar Nasional Pendidikan Matematika*, 1–7.
- Du, J., Zheng, J., Liang, Y., Lu, X., Klemeš, J. J., Varbanov, P. S., Shahzad, K., Rashid, M. I., Ali, A. M., Liao, Q., Liao, Q., & Wang, B. (2022). A hybrid deep learning framework for predicting daily natural gas consumption. *Energy, 257*. <https://doi.org/10.1016/j.energy.2022.124689>
- Fraedrich, K. (1986). Estimating the Dimensions of Weather and Climate Attractors. *Journal*

of the Atmospheric Sciences, 43(5), 419–432. [https://doi.org/10.1175/1520-0469\(1986\)043<0419:etdowa>2.0.co;2](https://doi.org/10.1175/1520-0469(1986)043<0419:etdowa>2.0.co;2)

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53(4), 267–293. <https://doi.org/10.1037/h0040755>

Hussain, A., Memon, J. A., Murshed, M., Alam, M. S., Mehmood, U., Alam, M. N., Rahman, M., & Hayat, U. (2022). A time series forecasting analysis of overall and sector-based natural gas demand: a developing South Asian economy case. *Environmental Science and Pollution Research*, 29(47), 71676–71687. <https://doi.org/10.1007/s11356-022-20861-3>

Irwan, Adnan Sauddin, & Anita Kaimuddin. (2022). Proyeksi Produksi Padi Kabupaten Pinrang Dengan Metode Singular Spectrum Analysis. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 10(1), 100–109. <https://doi.org/10.24252/msa.v10i1.29869>

Lisi, F., Nicolis, O., & Sandri, M. (1995). Combining Singular-Spectrum Analysis and neural networks for time series forecasting. *Neural Processing Letters*, 2, 6–10. <https://doi.org/10.1007/BF02279931>

Lu, H., Ma, X., & Azimi, M. (2020). US natural gas consumption prediction using an improved kernel-based nonlinear extension of the Arps decline model. *Energy*, 194. <https://doi.org/10.1016/j.energy.2020.116905>

Manigandan, P., Alam, M. D. S., Alharthi, M., Khan, U., Alagirisamy, K., Pachiyappan, D., & Rehman, A. (2021). Forecasting natural gas production and consumption in united states-evidence from sarima and sarimax models. *Energies*, 14(19). <https://doi.org/10.3390/en14196021>

Manowska, A., Rybak, A., Dylong, A., & Pielot, J. (2021). Forecasting of natural gas consumption in poland based on ARIMA - LSTM hybrid model. *Energies*, 14(24). <https://doi.org/10.3390/en14248597>

Singh, S., Bansal, P., Hosen, M., & Bansal, S. K. (2023). Forecasting annual natural gas consumption in USA: Application of machine learning techniques- ANN and SVM. *Resources Policy*, 80. <https://doi.org/10.1016/j.resourpol.2022.103159>

Sulandari, W., Subanar, Lee, M. H., & Rodrigues, P. C. (2020). Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy*, 190, 116408. <https://doi.org/https://doi.org/10.1016/j.energy.2019.116408>

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85, 208–220. <https://doi.org/https://doi.org/10.1016/j.energy.2015.03.084>

Wang, J., Lian, L., & Shen, Y. (2013). Application of singular spectral analysis to GPS station

- coordinate monitoring series. *Tongji Daxue Xuebao/Journal of Tongji University*, 41(2), 282–288. <https://doi.org/10.3969/j.issn.0253-374x.2013.02.022>
- Wang, Z.-X., He, L.-Y., & Zhao, Y.-F. (2021). Forecasting the seasonal natural gas consumption in the US using a gray model with dummy variables. *Applied Soft Computing*, 113. <https://doi.org/10.1016/j.asoc.2021.108002>
- Xu, G., Chen, Y., Yang, M., Li, S., & Marma, K. J. S. (2023). An outlook analysis on China's natural gas consumption forecast by 2035: Applying a seasonal forecasting method. *Energy*, 284, 128602. <https://doi.org/https://doi.org/10.1016/j.energy.2023.128602>
- Zu, X., Wang, X., & Cui, Y. (2023). Forecasting natural gas consumption in residential and commercial sectors in the US. *Journal of Business Analytics*, 6(1), 77–94. <https://doi.org/10.1080/2573234X.2022.2064777>