BAB 2 | ALGORITMA DAN METODE DATA MINING

Felix Reba, S.Si., M.Sc.

A. Clustering

Clustering adalah sebuah teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok (clusters) berdasarkan kesamaan karakteristik. Tujuan utama dari clustering adalah untuk menemukan struktur tersembunyi dalam data, di mana data dalam satu kelompok memiliki kesamaan yang tinggi antara satu sama lain, sedangkan data antar kelompok memiliki perbedaan yang signifikan. Beberapa metode umum dalam clustering adalah (Granell *et al.*, 2011; Ye, 2013):

1. K-Means
2. Agglomerative Hierarchical Clustering
3. DBSCAN
4. Mean Shift Clustering
5. Fuzzy C-Means (FCM)
6. Gaussian Mixture Models (GMM)
7. OPTICS
8. Self-Organizing Maps (SOM)
9. Affinity Propagation

Ada banyak lagi metode clustering lainnya, dan pilihan metode tergantung pada sifat data dan tujuan analisis. Setiap metode memiliki kelebihan dan kelemahan yang berbeda, dan pemilihan metode yang tepat sangat penting untuk mendapatkan hasil clustering yang baik.

B. Classification

Classification adalah sebuah teknik dalam data mining yang digunakan untuk memprediksi kategori atau label dari suatu data berdasarkan fitur-fitur yang diberikan. Tujuannya adalah untuk membangun model yang dapat mengklasifikasikan data baru ke dalam kelas yang telah ditentukan sebelumnya. Metode dalam data mining yang digunakan untuk melakukan klasifikasi termasuk (Granell *et al.*, 2011; Ye, 2013):

1. **Decision Trees**: Membagi data berdasarkan serangkaian keputusan berbasis fitur untuk memprediksi kelas target.
2. **Random Forest**: Kumpulan dari beberapa pohon keputusan yang digunakan untuk meningkatkan akurasi klasifikasi.
3. **Support Vector Machines (SVM)**: Membangun hyperplane atau beberapa hyperplane di ruang fitur yang memisahkan kelas-kelas dengan margin maksimum.
4. **Logistic Regression**: Model regresi yang digunakan untuk memprediksi probabilitas bahwa suatu pengamatan akan menjadi bagian dari satu kelas tertentu.
5. **k-Nearest Neighbors (k-NN)**: Mengklasifikasikan pengamatan berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur.
6. **Naive Bayes**: Menggunakan teorema Bayes untuk memperkirakan probabilitas kelas target berdasarkan fitur-fitur yang diberikan.
7. **Neural Networks**: Jaringan saraf tiruan yang terdiri dari lapisan-lapisan neuron yang digunakan untuk memodelkan hubungan kompleks antara fitur dan kelas target.
8. **AdaBoost**: Metode boosting yang menggabungkan beberapa model lemah menjadi satu model yang kuat.
9. **Ensemble Learning**: Kombinasi dari beberapa model klasifikasi untuk meningkatkan akurasi dan ketahanan terhadap overfitting.

Setiap metode memiliki kelebihan dan kelemahan, serta berlaku untuk berbagai jenis data dan skenario. Pemilihan metode yang tepat tergantung pada sifat data, kompleksitas masalah klasifikasi, dan tujuan analisis.

C. Association Rule Mining (Pencarian Aturan Asosiasi)

Association Rule Mining, atau Pencarian Aturan Asosiasi, adalah sebuah teknik dalam data mining yang digunakan untuk menemukan pola hubungan antara item-item dalam dataset transaksional atau dataset yang berisi himpunan item. Tujuan utama dari Association Rule Mining adalah untuk menemukan asosiasi atau korelasi yang kuat antara item-item yang sering muncul bersama-sama dalam transaksi atau dataset. Metode dalam Association Rule Mining meliputi (Dunhan, 2013; Ye, 2013):

1. **Apriori Algorithm**: Algoritma yang paling populer untuk pencarian aturan asosiasi. Algoritma ini bekerja dengan menghasilkan kandidat-kandidat aturan asosiasi berdasarkan prinsip apriori, di mana setiap subset dari aturan harus sering terjadi dalam dataset.
2. **FP-Growth Algorithm**: Algoritma yang efisien untuk pencarian aturan asosiasi. FP-Growth bekerja dengan membangun struktur pohon yang disebut FP-Tree, yang kemudian digunakan untuk mengekstraksi aturan asosiasi secara efisien.
3. **Eclat Algorithm**: Algoritma alternatif untuk pencarian aturan asosiasi. Eclat menggunakan struktur tumpukan vertikal (vertical format) untuk menghitung frekuensi itemset secara efisien.
4. **Maximal Frequent Itemsets**: Selain aturan asosiasi, metode ini juga digunakan untuk menemukan himpunan item yang sering muncul bersama-sama, tetapi tidak semua itemset yang muncul secara bersamaan termasuk dalam aturan.

5. **Closed Frequent Itemsets**: Seperti maximal frequent itemsets, namun himpunan item yang dihasilkan adalah himpunan tertutup yang tidak ada subsetnya dengan dukungan yang sama.
6. **Constraint-Based Association Mining**: Pendekatan yang memungkinkan pengguna untuk menyertakan aturan atau batasan tertentu yang harus dipenuhi oleh aturan asosiasi yang dihasilkan.
7. **Sequential Pattern Mining**: Metode yang digunakan untuk menemukan pola urutan dari item-item yang sering muncul bersama-sama dalam transaksi atau urutan waktu tertentu.

Association Rule Mining memiliki banyak aplikasi dalam berbagai bidang, termasuk analisis keranjang belanja (market basket analysis), analisis log web, personalisasi rekomendasi produk, dan analisis interaksi item dalam sistem e-commerce.

D. Regresi

Regresi adalah sebuah teknik dalam data mining yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (fitur) dan variabel dependen (target) dalam dataset. Tujuan utama dari regresi adalah untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen. Beberapa metode dalam regresi yang umum digunakan meliputi (Ye, 2013):

1. **Linear Regression**: Model regresi yang paling sederhana, di mana hubungan antara variabel independen dan dependen diasumsikan linear. Regresi Linier Berganda:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

Dengan:
Y = Variabel Dependen; X = Variabel Independen; β_0 = Konstanta; $\beta_1, \ldots, \beta_n$ = Koefisien Determinasi; ε = Error Term

2. **Logistic Regression**: Digunakan ketika variabel dependen adalah biner (dua kategori). Model ini memprediksi probabilitas kejadian suatu kategori berdasarkan variabel independen. Saat data yang lebih relevan ditambahkan,

algoritma meningkatkan kemampuannya untuk memprediksi klasifikasi dalam kumpulan data:

$$p = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k)}}$$

3. **Polynomial Regression**: Memodelkan hubungan non-linear antara variabel independen dan dependen dengan menggunakan fungsi polinomial.

4. **Ridge Regression**: Versi regularized dari regresi linear yang mengurangi overfitting dengan menambahkan komponen regularisasi ke fungsi biaya. Jadi dalam Regresi Ridge fungsi tujuan kita adalah:

$$Min\left(\sum \varepsilon^2 + \lambda \sum \beta^2\right)$$

$$= Min \sum (y - (\beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k))^2 + \lambda \sum \beta^2$$

Berikut ini λ adalah parameter regularisasi yang merupakan bilangan non negatif.

5. **Lasso Regression**: Regresi Lasso (Least Absolute Shrinkage and Selection Operator) adalah metode regresi yang digunakan untuk melakukan pemilihan variabel dan mengurangi dimensi dalam model regresi. Tujuannya adalah untuk meminimalkan jumlah koefisien yang signifikan secara statistik, dengan memaksa beberapa koefisien menjadi nol. Ini membantu dalam menghasilkan model yang lebih sederhana dan mudah diinterpretasikan, serta mengatasi masalah multicollinearity di antara prediktor. Berikut ini adalah Model Regresi Lasso:

$$Min\left(\sum \varepsilon^2 + \lambda \sum |\beta|\right)$$

$$= Min \sum (y - (\beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k))^2 + \lambda \sum |\beta|$$

6. **ElasticNet Regression**: Regresi ElasticNet adalah teknik regresi yang menggabungkan dua metode regresi yang populer, yaitu regresi Lasso (Least Absolute Shrinkage and

Selection Operator) dan regresi Ridge. Tujuan dari regresi ElasticNet adalah untuk mengatasi beberapa kelemahan dari kedua metode tersebut, serta memperoleh keuntungan dari keduanya. Berikut adalah Model Regresi Jaringan Elastis:

$$Min\left(\sum \varepsilon^2 + \lambda_1 \sum \beta^2 + \lambda_2 \sum |\beta|\right)$$

$$= Min \sum (y - (\beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k))^2 \lambda_1 \sum \beta^2 + \lambda_2 \sum |\beta|$$

7. **Support Vector Regression (SVR)**: Menggunakan pendekatan dari Support Vector Machines (SVM) untuk regresi, dengan tujuan untuk menemukan hyperplane yang memiliki kesalahan prediksi minimum.
8. **Decision Tree Regression**: Memodelkan hubungan antara variabel independen dan dependen dengan membagi ruang fitur ke dalam segmen-segmen berdasarkan serangkaian keputusan.
9. **Random Forest Regression**: Kumpulan dari beberapa pohon keputusan yang digunakan untuk meningkatkan akurasi dan ketahanan terhadap overfitting dalam regresi.

E. Anomaly Detection

Anomaly Detection, atau deteksi anomali, adalah sebuah teknik dalam data mining yang digunakan untuk mengidentifikasi pola atau instansi yang tidak biasa atau langka dalam dataset. Tujuan utama dari Anomaly Detection adalah untuk menemukan observasi yang berbeda secara signifikan dari pola umum dalam data. Anomali seringkali menunjukkan kejadian yang tidak biasa, kegagalan sistem, atau perilaku yang mencurigakan. Beberapa metode dalam Anomaly Detection termasuk (Ville, 2001):

1. **Statistical Methods**: Pendekatan yang menggunakan statistik deskriptif untuk menentukan apakah sebuah observasi adalah anomali berdasarkan ukuran seperti jarak, deviasi standar, atau distribusi.

2. **Density Based Methods**: Menggunakan metrik kerapatan data seperti jarak ke k-tetangga terdekat atau estimasi kerapatan kernel untuk mengidentifikasi area yang jarang atau langka dalam ruang fitur.
3. **Clustering Methods**: Memisahkan data ke dalam kelompok-kelompok dan mengidentifikasi observasi yang berada di luar kelompok utama sebagai anomali.
4. **Classification Methods**: Membangun model klasifikasi untuk membedakan antara data normal dan anomali. Contohnya adalah One-Class SVM dan algoritma klasifikasi lainnya yang dilatih dengan hanya menggunakan data normal.
5. **Proximity-Based Methods**: Mengukur jarak atau kemiripan antara setiap observasi dalam dataset, dan mengidentifikasi observasi yang memiliki jarak atau kemiripan yang rendah sebagai anomali.
6. **Deep Learning Approaches**: Menggunakan arsitektur jaringan saraf yang kompleks untuk menangkap pola yang rumit dalam data dan mengidentifikasi anomali berdasarkan perbedaan yang signifikan dari pola umum.
7. **Ensemble Methods**: Menggabungkan beberapa model deteksi anomali untuk meningkatkan keakuratan dan ketahanan terhadap noise.
8. **Time Series Analysis**: Melakukan analisis pada data deret waktu untuk mendeteksi perubahan atau kejadian yang tidak biasa dalam pola waktu.
9. **Isolation Forest**: Algoritma yang mengisolasi anomali dengan membagi ruang fitur secara acak dan mengukur jumlah partisi yang diperlukan untuk mengisolasi observasi.

F. Dimensionality Reduction

Dimensionality reduction adalah proses mengurangi jumlah fitur (variabel) dalam kumpulan data dengan mempertahankan sebanyak mungkin informasi yang relevan. Tujuan utamanya adalah untuk mengatasi masalah "kutukan dimensi" (curse of dimensionality) di mana semakin banyak fitur

yang ada dalam data, semakin sulit untuk menganalisisnya, memahaminya, dan menghasilkan model yang baik. Dimensi yang tinggi sering kali mengakibatkan kompleksitas komputasi yang tinggi, peningkatan memori yang besar, dan risiko overfitting (Ville, 2001).

1. **Principal Component Analysis (PCA):** Sebuah teknik dalam data mining yang digunakan untuk mereduksi dimensi dari dataset yang kompleks dengan memproyeksikan data ke dalam ruang dimensi yang lebih rendah, sambil mempertahankan sebanyak mungkin informasi yang relevan. PCA berusaha untuk menemukan arah (komponen) yang paling bervariasi dalam data dan mentransformasikan data ke dalam ruang yang terdefinisi oleh komponen-komponen tersebut.

2. **t-Distributed Stochastic Neighbor Embedding (t-SNE):** Sebuah teknik dalam data mining yang digunakan untuk mereduksi dimensi dari dataset yang kompleks dengan memetakan data ke dalam ruang dimensi yang lebih rendah. t-SNE terutama efektif dalam memvisualisasikan data yang kompleks dalam ruang dimensi yang lebih rendah, sehingga memungkinkan untuk menangkap struktur dan pola yang tersembunyi dalam data. Dimana, P_{ij} adalah probabilitas bersyarat antara pasangan pengamatan i dan j dalam ruang dimensi asli, dan Q_{ij} adalah probabilitas bersyarat yang dipetakan ke dalam ruang dimensi yang direduksi, maka fungsi biaya untuk t-SNE dapat dinyatakan sebagai berikut:

$$C = \sum_i KL(P_i|||Q_i) = \sum_i \sum_j P_{ij} log\left(\frac{P_{ij}}{Q_{ij}}\right)$$

Di sini, $KL(P_i||Q_i)$ adalah divergensi Kullback-Leibler antara distribusi probabilitas P_i dan Q_i untuk pengamatan ke-i. Tujuan dari t-SNE adalah untuk meminimalkan nilai dari fungsi biaya ini.

3. **Linear Discriminant Analysis (LDA)**: Sebuah teknik dalam data mining yang digunakan untuk mereduksi dimensi dari dataset dan pada saat yang sama memaksimalkan pemisahan antara kelas-kelas yang berbeda dalam dataset. LDA mencoba untuk menemukan kombinasi linear dari fitur-fitur yang memaksimalkan pemisahan antara kelas-kelas, sehingga membuatnya menjadi metode yang umum digunakan dalam klasifikasi.

G. Ensemble Learning

Ensemble learning adalah teknik dalam pembelajaran mesin di mana beberapa model (biasanya disebut sebagai "anggota ensemble") digabungkan bersama untuk meningkatkan kinerja prediksi secara keseluruhan. Konsep di balik ensemble learning adalah bahwa gabungan dari beberapa model yang secara individu mungkin tidak sempurna dapat menghasilkan prediksi yang lebih akurat dan stabil (Zhou, 2012).

1. **Bagging (Bootstrap Aggregating)**: Sebuah teknik dalam data mining yang digunakan untuk meningkatkan kinerja model dengan cara menggabungkan hasil dari beberapa model yang dilatih secara independen. Random Forest merupakan salah satu contoh populer dari algoritma yang menggunakan teknik bagging.

   Dimana, $f_i(x)$ adalah fungsi prediksi dari model ke-i untuk pengamatan x, dan terdapat M model dalam ensemble, maka prediksi dari ensemble bagging bisa dinyatakan sebagai:

$$\hat{f}_{bagging}(x) = \frac{1}{M}\sum_{i=1}^{M} f_i(x)$$

   $\hat{f}_{bagging}(x)$: prediksi akhir dari ensemble bagging untuk pengamatan x

2. **Boosting**: Sebuah teknik dalam data mining yang digunakan untuk meningkatkan kinerja model dengan cara menggabungkan hasil dari beberapa model yang lemah (weak learners) menjadi model yang kuat (strong learner).

Ada beberapa algoritma boosting yang populer, seperti AdaBoost (Adaptive Boosting) dan Gradient Boosting Machines (GBM).

Rumus umum untuk prediksi dari ensemble boosting dapat dinyatakan sebagai berikut:

$$\hat{f}_{boosting}(x) = \sum_{i=1}^{T} \alpha_t f_t(x)$$

$\hat{f}_{boosting}(x)$: prediksi akhir dari ensemble boosting untuk pengamatan x, T adalah jumlah iterasi, α_t adalah bobot yang diberikan pada model ke-t, dan $f_t(x)$ adalah prediksi dari model ke-t.

3. **Stacking**: Sebuah teknik ensemble learning yang digunakan untuk meningkatkan kinerja model dengan cara menggabungkan hasil dari beberapa model yang dilatih secara independen. Dalam stacking, prediksi dari model-model dasar (base models) digunakan sebagai fitur input untuk model meta (meta model) yang kemudian digunakan untuk membuat prediksi akhir.

   Dimana, x adalah prediksi dari model dasar ke-i pengamatan dan g(x) adalah prediksi dari model meta, maka prediksi dari stacking dapat dinyatakan sebagai:

$$\hat{y}_{stacking}(x) = g(\{f_1(x), f_2(x), \ldots, f_n(x)\})$$

$\hat{y}_{stacking}(x)$: prediksi akhir dari stacking untuk pengamatan x.

H. Deep Learning

Deep Learning adalah sebuah cabang dari machine learning yang menggunakan arsitektur jaringan saraf tiruan (neural networks) yang dalam untuk memodelkan dan mempelajari representasi hierarkis dari data. Deep Learning bertujuan untuk mengekstraksi fitur yang paling relevan atau representasi yang paling abstrak dari data input untuk melakukan tugas-tugas seperti klasifikasi, regresi, pengenalan pola, atau generasi data baru.

Beberapa metode dalam Deep Learning meliputi (Yang *et al*., 2022; Yi, 2022):

1. **Convolutional Neural Networks (CNN)**: Digunakan khusus untuk data berstruktur grid, seperti gambar. CNN memiliki lapisan konvolusi yang memungkinkan model untuk menangkap pola spasial dalam data.
2. **Recurrent Neural Networks (RNN)**: Dirancang untuk bekerja dengan data berurutan, seperti teks atau time series. RNN memiliki loop yang memungkinkan informasi untuk diproses secara berurutan.
3. **Long Short-Term Memory (LSTM)**: Jenis khusus dari RNN yang dirancang untuk mengatasi masalah vanishing gradient, yang memungkinkan model untuk mempelajari ketergantungan jangka panjang dalam data berurutan.
4. **Generative Adversarial Networks (GAN)**: Terdiri dari dua jaringan saraf yang bersaing, yaitu generator dan diskriminator, yang bekerja bersama untuk menghasilkan data baru yang terlihat seperti data yang ada.
5. **Autoencoders**: Jaringan saraf yang digunakan untuk mempelajari representasi tersembunyi dari data input dengan mengurangi dimensinya ke dalam ruang fitur yang lebih kecil.
6. **Transformer Models**: Arsitektur jaringan saraf yang berfokus pada perhatian dan dapat menangani input sekuensial dengan baik, seperti teks atau audio. Transformer terkenal karena kemampuannya dalam pemrosesan bahasa alami dan mampu menghasilkan hasil yang luar biasa dalam tugas-tugas seperti terjemahan mesin dan generasi teks.
7. **Deep Reinforcement Learning**: Menggabungkan Deep Learning dengan reinforcement learning, yang melibatkan agen yang belajar tindakan yang optimal dalam lingkungan yang dinamis.
8. **Deep Belief Networks (DBN)**: Arsitektur yang terdiri dari beberapa lapisan yang berbeda jenis, termasuk lapisan yang bertindak sebagai pembelajaran unsupervised dan lapisan yang bertindak sebagai pembelajaran supervised.

I. Text Mining

Text Mining adalah sebuah teknik dalam data mining yang digunakan untuk mengekstraksi informasi yang bermakna dan terstruktur dari teks yang tidak terstruktur, seperti dokumen, artikel, email, dan media sosial. Tujuan utama dari Text Mining adalah untuk mengidentifikasi pola, tren, hubungan, dan wawasan baru dalam teks yang dapat digunakan untuk pengambilan keputusan. Beberapa metode dalam Text Mining termasuk (Jiawei Han; Kamber Micheline, 2011) dan (Bidin A, 2017):

1. **Tokenization;** Proses memecah teks menjadi unit-unit yang lebih kecil seperti kata-kata, frasa, atau token.
2. **Stopwords Removal:** Menghapus kata-kata yang umum dan tidak memiliki makna (stopwords) dari teks untuk meningkatkan kualitas analisis.
3. **Stemming and Lemmatization:** Mengubah kata-kata ke bentuk dasar (stemming) atau kata dasar (lemmatization) untuk mengurangi variasi kata yang memiliki arti yang sama.
4. **Bag of Words (BoW):** Representasi teks sebagai himpunan kata-kata unik dalam dokumen tanpa memperhatikan urutan atau struktur kalimat.
5. **Term Frequency-Inverse Document Frequency (TF-IDF):** Menghitung bobot kata-kata dalam dokumen berdasarkan seberapa sering kata-kata tersebut muncul dalam dokumen (term frequency) dan seberapa jarang kata-kata tersebut muncul dalam seluruh korpus dokumen (inverse document frequency).
6. **Word Embeddings:** Representasi vektor yang dipelajari dari kata-kata dalam teks yang menangkap makna dan konteks kata-kata.
7. **Topic Modeling:** Mengidentifikasi topik-topik utama yang ada dalam korpus teks menggunakan algoritma seperti Latent Dirichl*et al*location (LDA).
8. **Sentiment Analysis:** Mengidentifikasi dan mengekstraksi sentimen atau opini dari teks, seperti positif, negatif, atau netral.

9. **Text Classification:** Mengklasifikasikan dokumen atau teks ke dalam kategori-kategori yang telah ditentukan sebelumnya berdasarkan konten atau topiknya, menggunakan metode seperti Naive Bayes, Support Vector Machines (SVM), atau Deep Learning.

J. Time Series Analysis

Time series analysis adalah sebuah metode statistik yang digunakan untuk menganalisis data yang dihasilkan dalam interval waktu yang teratur. Data dalam time series biasanya diambil pada titik-titik waktu yang berurutan, seperti harian, bulanan, atau tahunan, dan dianalisis untuk mengidentifikasi pola, tren, dan perilaku dalam data tersebut (Jiawei Han; Kamber Micheline, 2011).

1. *Autoregressive Integrated Moving Average* (ARIMA) adalah model statistik yang digunakan untuk menganalisis dan meramalkan data deret waktu. Model ini merupakan kombinasi dari tiga komponen utama: autoregressive (AR), integrated (I), dan moving average (MA). Rumus umum ARIMA(p, d, q) untuk nilai-nilai deret waktu y_t adalah sebagai berikut:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t$$

dengan:
c : konstanta, $\phi_1, \phi_2, \ldots, \phi_p)$: parameter autoregressive, $(\theta_1, \theta_2, \ldots, \theta_q)$: parameter moving average, e_t : nilai gangguan pada waktu t.

2. Metode Exponential Smoothing adalah teknik yang digunakan dalam analisis deret waktu untuk meramalkan data berdasarkan pola eksponensial. Rumus dasar metode Exponential Smoothing untuk meramalkan nilai $\hat{y}_{t+1}$ berdasarkan nilai observasi sebelumnya $\hat{y}_t$ adalah sebagai berikut:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

Di mana,
α : parameter smoothing yang berada di antara 0 dan 1
y_t : nilai observasi pada waktu t
$\hat{y}_t$: ramalan sebelumnya (biasanya ramalan pada waktu sebelumnya).

3. *Seasonal Decomposition of Time Series* (STL) adalah metode statistik yang digunakan untuk memisahkan sebuah deret waktu menjadi tiga komponen utama: komponen musiman, komponen tren, dan komponen sisa (residual). Secara umum, STL memisahkan deret waktu y_t menjadi tiga komponen sebagai berikut:

$$y_t = S_t + T_t + R_t$$

S_t : komponen musiman, T_t : komponen tren dan R_t : komponen sisa atau residu.

DAFTAR PUSTAKA

Bidin A. (2017). Опыт аудита обеспечения качества и безопасности медицинской деятельности в медицинской организации по разделу «Эпидемиологическая безопасностьNo Title. In *Вестник Росздравнадзора* (Vol. 4, Issue 1).

Dunhan, M. H. (2013). Data Mining: Introductory and Advanced Topics 1st Edition. *Engineering*, 1–89.

Granell, C., Gómez, S., & Arenas, A. (2011). Mesoscopic analysis of networks: Applications to exploratory analysis and data clustering. In *Chaos* (Vol. 21, Issue 1). https://doi.org/10.1063/1.3560932

Jiawei Han; Kamber Micheline. (2011). *Data Mining: Concept & Technique*.

Ville, B. de. (2001). Introduction to Data Mining. In *Microsoft Data Mining*. https://doi.org/10.1016/b978-155558242-5/50003-6

Yang, H., Zheng, Z., & Sun, C. (2022). E-Commerce Marketing Optimization of Agricultural Products Based on Deep Learning and Data Mining. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/6564014

Ye, N. (2013). Data Mining: Theories, Algorithms, and Examples. In *Data Mining: Theories, Algorithms, and Examples*.

Yi, J. (2022). Deep Learning in Data Mining Management of Industrial and Commercial Enterprises. *Mobile Information Systems*. https://doi.org/10.1155/2022/6974993

Zhou, Z. H. (2012). Ensemble methods: Foundations and algorithms. In *Ensemble Methods: Foundations and Algorithms*. https://doi.org/10.1201/b12207