

BAB II

LANDASAN TEORI

2.1 Kualitas Air

Air merupakan sumber daya alam yang sangat vital bagi kehidupan manusia dan makhluk hidup lainnya. Kualitas air yang baik menjadi salah satu indikator penting dalam menjamin kesehatan masyarakat, mendukung kegiatan domestik, industri, maupun pertanian. Menurut *World Health Organization* (WHO, 2017), air minum yang aman harus bebas dari mikroorganisme patogen, zat kimia berbahaya, serta tidak berwarna, tidak berbau, dan tidak memiliki rasa yang menyimpang.

Kualitas air ditentukan oleh sejumlah variabel fisik, kimia, dan biologi. Variabel fisik meliputi suhu, kekeruhan, dan warna, sedangkan variabel kimia mencakup pH, daya hantar listrik (DHL), serta kandungan zat seperti nitrat (NO_3^-), fluorida (F^-), klorida (Cl^-), dan logam berat seperti besi (Fe), mangan (Mn), dan seng (Zn). Variabel-variabel ini memberikan gambaran tentang kondisi air serta potensi pencemaran yang mungkin terjadi (Sutaryo *et al.*, 2021).

Pemerintah Indonesia melalui *Peraturan Menteri Kesehatan No. 492/Menkes/Per/IV/2010* telah menetapkan standar baku mutu air minum yang wajib dipenuhi oleh penyedia layanan air, termasuk PDAM. Ketidaksesuaian terhadap standar ini dapat menimbulkan dampak kesehatan, baik akut maupun kronis, seperti gangguan sistem pencernaan, akumulasi logam berat dalam tubuh, hingga risiko kanker.

Beberapa variabel seperti *Total Dissolved Solids* (TDS) dan Daya Hantar Listrik (DHL) berperan sebagai indikator konsentrasi ion-ion terlarut dalam air, yang berhubungan erat dengan tingkat kesadahan. Air dengan tingkat kesadahan yang terlalu tinggi atau terlalu rendah tidak hanya menimbulkan permasalahan teknis, seperti kerak pada pipa dan ketidakefisienan sabun, tetapi juga dapat berdampak terhadap rasa dan kenyamanan air (Afifah *et al.*, 2022).

Keberadaan zat seperti sisa klor (Cl_2), aluminium (Al), dan zat organik yang diukur melalui reaksi dengan kalium permanganat (KMnO_4) juga memengaruhi kejernihan dan kualitas visual air. Meskipun variabel ini tidak selalu berbahaya secara langsung, perubahan tampilan air dapat memengaruhi persepsi konsumen terhadap kualitas dan keamanan air yang dikonsumsi.

2.2 Variabel Kualitas Air

Penilaian kualitas air minum sangat bergantung pada berbagai variabel fisik, kimia, dan biologis. Dalam penelitian ini, analisis difokuskan pada variabel-variabel kimia dan fisika yang umumnya digunakan dalam evaluasi kualitas air oleh instansi penyedia air bersih seperti PDAM.

1. Padatan Total (*Total Dissolved Solids*/TDS)

Padatan total merupakan jumlah zat padat terlarut dalam air yang mencakup garam anorganik dan senyawa organik dalam jumlah kecil. Menurut *World Health Organization* (WHO, 2017), kadar TDS yang melebihi 500 mg/L dapat menimbulkan rasa tidak enak dan menjadi indikator adanya pencemaran air.

2. Sisa Klor (*Residual Chlorine*)

Sisa klor merupakan klorin yang tersisa setelah proses desinfeksi air. Meskipun penting untuk menjamin tidak adanya mikroorganisme patogen, klorin dalam kadar tinggi dapat berdampak buruk bagi kesehatan, sehingga *Permenkes No. 492/Menkes/Per/IV/2010* menetapkan batas maksimal sebesar 0,3 mg/L (Kementerian Kesehatan RI, 2010).

3. Kekeruhan (*Turbidity*)

Kekeruhan (*turbidity*) merupakan indikator penting yang menunjukkan keberadaan partikel tersuspensi seperti lumpur atau mikroorganisme yang dapat menghambat proses desinfeksi. Nilai ideal kekeruhan menurut *World Health Organization* (WHO) berada di bawah 5 NTU (*Nephelometric Turbidity Units*) (Afifah *et al.*, 2022).

4. pH air

pH air yang menunjukkan tingkat keasaman atau kebasaan, memiliki peran penting dalam menjaga stabilitas kimia air dan mencegah korosi. Nilai pH optimal untuk air minum berkisar antara 6,5 hingga 8,5 (Sutaryo *et al.*, 2021).

5. Daya Hantar Listrik (DHL)

Daya Hantar Listrik (DHL) merupakan variabel penting dalam penilaian kualitas air karena mencerminkan jumlah ion terlarut dalam air, seperti kalsium, magnesium, natrium, dan klorida. Semakin tinggi kandungan ion dalam air, semakin tinggi pula nilai konduktivitas listriknya.

Satuan untuk daya hantar listrik adalah mikrosiemens per sentimeter ($\mu\text{S}/\text{cm}$). Nilai DHL yang terlalu tinggi dapat mengindikasikan potensi kelebihan garam atau senyawa anorganik terlarut, yang mungkin berdampak pada rasa air dan efektivitas desinfeksi.

Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 492/Menkes/Per/IV/2010, batas maksimum yang diperbolehkan untuk daya hantar listrik dalam air minum adalah $1500 \mu\text{S}/\text{cm}$. Batas ini ditetapkan untuk menjaga kualitas dan keamanan air yang dikonsumsi masyarakat.

6. Warna

Warna merupakan salah satu variabel visual penting dalam penilaian kualitas air minum. Warna dalam air biasanya disebabkan oleh keberadaan zat organik alami, logam seperti besi dan mangan, atau hasil dari proses pengolahan seperti sisa koagulan. Warna yang tinggi pada air tidak selalu berbahaya secara langsung, namun dapat mengurangi tingkat penerimaan masyarakat terhadap air tersebut, baik dari segi estetika maupun persepsi kebersihan.

Warna diukur dalam satuan *True Color Unit* (TCU), yang menunjukkan intensitas warna air setelah partikel tersuspensi disaring.

Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 492/Menkes/Per/IV/2010, batas maksimum kadar warna dalam air minum yang diperbolehkan adalah 15 TCU. Nilai ini ditetapkan untuk memastikan kenyamanan dan persepsi kualitas air oleh konsumen.

7. Suhu

Suhu air berperan penting dalam proses pengolahan karena memengaruhi kelarutan oksigen dan laju reaksi kimia, dengan suhu ideal berkisar antara $10 - 25^{\circ}\text{C}$ (Afifah, *et al.*, 2022).

8. Aluminium (Al)

Aluminium merupakan logam yang umumnya berasal dari penggunaan koagulan seperti aluminium sulfat dalam proses pengolahan air. Apabila kadarnya berlebihan dalam air minum, aluminium dapat berdampak negatif terhadap kesehatan, termasuk berpotensi memicu gangguan pada sistem saraf. *World Health Organization* (WHO, 2017) menetapkan batas maksimum aluminium dalam air minum adalah $0,2 \text{ mg/L}$.

9. Besi (Fe)

Besi dalam konsentrasi tinggi dapat menyebabkan perubahan rasa air menjadi logam, serta meninggalkan noda pada pakaian dan peralatan. Selain itu, endapan besi dapat menyumbat pipa dan mengganggu proses distribusi air. *World Health Organization* (WHO, 2017) menetapkan ambang batas besi sebesar 0,3 mg/L.

10. Mangan (Mn)

Mangan dapat menyebabkan perubahan warna dan rasa pada air serta meninggalkan endapan hitam pada permukaan. Dalam konsentrasi tinggi, mangan juga berpotensi menyebabkan gangguan neurologis jika terakumulasi dalam tubuh dalam jangka panjang. *World Health Organization* (WHO, 2017) menetapkan batas maksimal mangan sebesar 0,1 mg/L.

11. Kandungan Nitrat (NO_3)

Kandungan nitrat yang tinggi dalam air minum dapat menimbulkan risiko kesehatan serius, terutama pada bayi, dalam bentuk methemoglobinemia. *World Health Organization* (WHO, 2017) menetapkan batas aman nitrat sebesar 50 mg/L.

12. Fluorida (F)

Fluorida merupakan unsur yang secara alami terdapat di air tanah dan dikenal bermanfaat dalam mencegah kerusakan gigi. Namun, jika dikonsumsi dalam kadar tinggi secara terus-menerus, fluorida dapat menyebabkan fluorosis, yaitu kondisi pengeroposan atau perubahan warna pada gigi dan bahkan gangguan pada tulang. Oleh karena itu, *World Health Organization* (WHO, 2017) merekomendasikan batas maksimum fluorida dalam air minum sebesar 1,5 mg/L agar aman bagi kesehatan. Di Indonesia, batas maksimum ini juga diadopsi dalam *Permenkes No. 492/Menkes/Per/IV/2010* tentang Persyaratan Kualitas Air Minum.

13. Sulfat (SO_4)

Sulfat merupakan senyawa yang secara alami terdapat dalam air akibat pelarutan mineral atau aktivitas industri. Konsentrasi sulfat yang tinggi dalam air minum dapat menimbulkan efek laksatif, terutama pada anak-anak atau orang yang tidak terbiasa. Selain itu, kandungan sulfat yang tinggi juga dapat

mempercepat korosi pada pipa distribusi air. Berdasarkan *World Health Organization* (WHO, 2017), batas maksimum sulfat dalam air minum adalah 250 mg/L, sedangkan menurut *Permenkes No. 492/Menkes/Per/IV/2010*, ambang batas yang ditetapkan adalah 400 mg/L.

14. Klorida (Cl^-)

Klorida umumnya terdapat dalam air akibat pelarutan garam, intrusi air laut, atau limbah domestik. Meskipun tidak berbahaya dalam jumlah kecil, kadar klorida yang tinggi dapat menimbulkan rasa asin pada air serta mempercepat proses korosi pada pipa dan perlengkapan logam. *World Health Organization* (WHO, 2017) menyarankan batas maksimum sebesar 250 mg/L, dan hal ini juga diatur dalam *Permenkes No. 492/Menkes/Per/IV/2010*.

15. Seng (Zn)

Seng (Zn) merupakan logam yang dapat masuk ke dalam air minum terutama dari proses korosi pipa galvanis atau sistem distribusi air yang menggunakan material berbasis seng. Meskipun seng bukan logam berat yang sangat berbahaya, kadar yang terlalu tinggi dapat menimbulkan rasa logam pada air serta menyebabkan gangguan gastrointestinal seperti mual dan diare, terutama jika dikonsumsi dalam jangka panjang. Menurut *World Health Organization* (WHO, 2017), batas maksimum yang disarankan untuk kandungan seng dalam air minum adalah 3 mg/L.

16. Tembaga (Cu)

Tembaga (Cu) biasanya berasal dari pipa tembaga atau fitting pada sistem perpipaan rumah tangga. Jika air bersifat korosif, tembaga dapat larut ke dalam air minum. Dalam jumlah kecil, tembaga sebenarnya dibutuhkan oleh tubuh, namun apabila melebihi ambang batas, dapat menyebabkan rasa pahit pada air serta gejala gastrointestinal seperti mual, muntah, dan kram perut. *World Health Organization* (WHO, 2017) merekomendasikan batas maksimum tembaga dalam air minum adalah 2 mg/L, yang juga sejalan dengan ketentuan dalam *Permenkes No. 492/Menkes/Per/IV/2010*.

17. Kesadahan Total

Kesadahan total menggambarkan jumlah ion-ion mineral, terutama kalsium (Ca^{2+}) dan magnesium (Mg^{2+}), yang terkandung dalam air. Air dengan tingkat

kesadahan yang tinggi dapat menyebabkan terbentuknya kerak pada permukaan pipa, pemanas air, serta peralatan rumah tangga lainnya. Selain itu, kesadahan juga mengurangi efektivitas sabun dan deterjen, karena ion kalsium dan magnesium bereaksi membentuk endapan. Oleh karena itu, pengukuran kesadahan total penting untuk menilai kualitas air secara keseluruhan (Pratama *et al.*, 2021; WHO, 2017).

18. Kesadahan Kalsium

Kesadahan kalsium mengacu pada jumlah ion kalsium yang terdapat dalam air. Ion ini merupakan komponen utama dari kesadahan total dan berperan besar dalam pembentukan kerak pada permukaan alat rumah tangga atau sistem perpipaan. Batas maksimum kadar kalsium dalam air minum umumnya tidak ditetapkan secara spesifik, namun pengukuran tetap diperlukan untuk memantau kontributor utama kesadahan *World Health Organization* (WHO, 2017).

19. Kesadahan Magnesium

Kesadahan magnesium menunjukkan jumlah ion magnesium dalam air. Meskipun secara umum magnesium aman dikonsumsi, jika kadarnya tinggi dapat menyebabkan rasa pahit pada air dan berkontribusi terhadap kesadahan total. Seperti halnya kalsium, magnesium juga dapat membentuk endapan yang mengganggu sistem distribusi air dan efisiensi penggunaan sabun *World Health Organization* (WHO, 2017).

20. Zat Organik (KMnO₄)

Zat organik dalam air biasanya diukur dengan menggunakan kalium permanganat (KMnO₄) sebagai indikator. Senyawa organik terlarut ini dapat berasal dari sisa tumbuhan, limbah rumah tangga, atau aktivitas industri. Keberadaan zat organik dapat mengganggu proses desinfeksi karena senyawa tersebut bereaksi dengan desinfektan seperti klorin dan membentuk produk sampingan yang berbahaya. Selain itu, zat organik juga dapat menjadi sumber nutrisi bagi pertumbuhan mikroorganisme. Oleh karena itu, pengukuran variabel ini penting dalam evaluasi kualitas air yang akan diolah *World Health Organization* (WHO, 2017) serta merekomendasikan batas maksimum zat

organik (KMnO_4) dalam air minum adalah 10 mg/L, yang juga sejalan dengan ketentuan dalam *Permenkes No. 492/Menkes/Per/IV/2010*.

Seluruh variabel kualitas air yang telah dijelaskan di atas memberikan gambaran menyeluruh mengenai aspek fisik, kimia, dan biologis dari air minum yang diproduksi PDAM. Mengingat banyaknya variabel yang saling berkaitan dan kompleksitas data yang dihasilkan, diperlukan pendekatan statistik multivariat untuk menganalisisnya secara simultan. Oleh karena itu, pada subbab berikutnya akan dibahas teori dan landasan dari metode analisis multivariat yang digunakan dalam penelitian ini, yaitu analisis komponen utama (*Principal Component Analysis/PCA*) dan Hotelling's T^2 Control Chart.

2.3 Analisis Multivariat

Analisis multivariat merupakan metode statistik yang digunakan untuk menganalisis data dengan lebih dari satu variabel secara simultan. Pendekatan ini sangat berguna dalam memahami struktur data yang kompleks, mengidentifikasi pola-pola hubungan antar variabel, serta mereduksi dimensi data tanpa kehilangan informasi penting (Hair *et al.*, 2010). Dalam konteks kualitas air, pendekatan multivariat sangat relevan karena data yang diperoleh biasanya terdiri dari berbagai variabel fisika dan kimia yang saling berkaitan dan berinteraksi satu sama lain.

Salah satu teknik utama dalam analisis multivariat adalah Analisis Komponen Utama (*Principal Component Analysis/PCA*). PCA bertujuan mereduksi jumlah variabel menjadi sejumlah kecil komponen utama yang mampu menjelaskan sebagian besar variabilitas dalam data. Teknik ini bekerja dengan mengubah sekumpulan variabel yang saling berkorelasi menjadi variabel baru yang saling bebas (tidak berkorelasi) yang disebut komponen utama (Jolliffe *et al.*, 2016).

1. Teorema 1 : Variansi Maksimum dalam PCA

Pernyataan:

Bila data sudah di-center, maka komponen utama ke- k (PC_k) adalah arah (vektor satuan) yang memaksimumkan variansi proyeksi data, dengan syarat orthogonal terhadap komponen sebelumnya. Nilai Maksimum ini sama dengan nilai *eigen* ke- k dari matriks kovarians Σ (Jolliffe, 2002).

Secara matematis, jika Σ adalah matriks kovarians dari data, maka komponen utama $\mathbf{w}_k$ adalah vektor satuan yang memenuhi:

$$\mathbf{w}_k = \arg \max_{\mathbf{w}: \|\mathbf{w}\|=1, \mathbf{w} \perp \mathbf{w}_1, \dots, \mathbf{w}_{k-1}} \mathbf{w}^T \Sigma \mathbf{w}$$

dan nilai maksimum itu sama dengan nilai *eigen* ke- k , λ_k .

- Komponen utama pertama, $\mathbf{w}_1$, adalah arah di mana proyeksi data memiliki variansi terbesar dari semua arah satuan.
- Komponen utama kedua, $\mathbf{w}_2$, adalah arah yang memaksimalkan variansi di antara arah-arahan yang ortogonal terhadap $\mathbf{w}_1$, dan seterusnya.
- Karena Σ adalah simetri positif semidefinit, dekomposisi *eigen* memberikan basis ortonormal yang menyusun komponen utama ini.
- Dengan demikian, PCA bisa juga dilihat sebagai memecah variansi total ke dalam komponen-komponen yang tidak saling berkorelasi (“*variance decomposition*”).

2. Teorema 2 : Model Faktor Linear

Pernyataan:

Setiap variabel observasi $\mathbf{X}_i$ dapat direpresentasikan sebagai:

$$\mathbf{X}_i = \boldsymbol{\mu}_i + \lambda_{i1}\mathbf{F}_1 + \lambda_{i2}\mathbf{F}_2 + \dots + \lambda_{im}\mathbf{F}_m + \boldsymbol{\varepsilon}_i$$

dengan $\mathbf{F}_j$ adalah faktor laten yang saling tidak berkorelasi dan $\boldsymbol{\varepsilon}_i$ adalah *error* unik yang tidak berkorelasi dengan faktor maupun *error* lainnya. Model ini menjelaskan kovarians antar variabel observasi sebagai $\Sigma_x = \Lambda\Lambda^T + \Psi$ (Anderson, 2003).

3. Teorema 3 : Distribusi Hotelling’s T^2

Pernyataan:

Misalkan $\mathbf{X}_1, \dots, \mathbf{X}_n \sim N_p(\boldsymbol{\mu}, \Sigma)$. Jika $\bar{\mathbf{X}}$ adalah rata-rata sampel dan S adalah kovarians sampel, maka

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T S^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$$

memenuhi distribusi

$$\frac{n-p}{p(n-1)} T^2 \sim F_{p, n-p}$$

dengan $n > p$. Hasil ini menjadikan Hotelling’s T^2 sebagai generalisasi multivariat dari uji t (Jhonson & Wichern, 2007).

2.3.1 Uji Prasyarat Sebelum Analisis Multivariat

Sebelum melakukan analisis multivariat seperti *Principal Component Analysis* (PCA) dan analisis faktor, penting untuk memastikan bahwa data memenuhi syarat awal agar hasil analisis dapat diinterpretasikan dengan lebih akurat dan bermakna. Dua uji prasyarat yang dilakukan dalam penelitian ini adalah uji normalitas dan uji korelasi antar variabel. Meskipun kedua uji ini tidak bersifat wajib untuk semua jenis analisis multivariat, keduanya sangat membantu dalam mengevaluasi kelayakan data serta memilih pendekatan statistik yang tepat.

1. Uji Normalitas (*Shapiro-Wilk Test*)

Uji normalitas bertujuan untuk mengetahui apakah distribusi data pada masing-masing variabel menyerupai distribusi normal. Distribusi normal merupakan salah satu asumsi umum dalam banyak metode statistik klasik, termasuk analisis komponen utama (PCA) berbasis kovarians. Salah satu metode uji yang digunakan adalah *Shapiro-Wilk Test*, yang dikenal efektif terutama untuk ukuran sampel kecil hingga menengah ($n < 2000$).

Dalam uji ini, jika nilai $p\text{-value} < 0,05$, maka data dianggap tidak berdistribusi normal secara statistik. Jika $p\text{-value} \geq 0,05$, maka data dianggap berdistribusi normal. Walaupun PCA dan analisis faktor tidak secara mutlak mempersyaratkan normalitas, hasil uji ini tetap penting sebagai dasar untuk memilih metode korelasi yang tepat serta menafsirkan hasil dengan lebih cermat.

Rumus *Shapiro-Wilk* (Shapiro *et al.*, 1965):

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3.1)$$

keterangan:

- $x_{(i)}$: nilai ke-i yang sudah diurutkan
- $\bar{x}$: rata-rata nilai
- a_i : konstanta yang dihitung berdasarkan distribusi normal.

2. Korelasi *Rank Spearman*

PCA dan analisis faktor bekerja dengan prinsip utama hubungan (korelasi) antar variabel. Oleh karena itu, sebelum melakukan analisis lebih lanjut, dilakukan pengujian korelasi untuk melihat apakah terdapat hubungan yang cukup kuat antar variabel yang akan dianalisis.

Jika data tidak berdistribusi normal, maka digunakan korelasi *Rank Spearman*, yaitu teknik korelasi nonparametrik yang tidak mengasumsikan distribusi normal. *Spearman* digunakan untuk mengukur kekuatan dan arah hubungan monotonik antar dua variabel, berdasarkan peringkat nilai (*ranking*), bukan nilai mentah.

Korelasi *Spearman* bernilai antara -1 dan 1:

- $r_s > 0$: hubungan positif (searah)
- $r_s < 0$: hubungan negatif (berlawanan arah)
- $r_s = 0$: tidak ada hubungan monotonik

Rumus koefisien *spearman* (Spearman, 1904):

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad (2.3.2)$$

keterangan:

- d_i : selisih antara peringkat dua variabel
- n : jumlah pasangan data

Uji ini dilakukan terhadap seluruh pasangan variabel yang diamati untuk mengetahui apakah terdapat korelasi yang cukup untuk melanjutkan analisis multivariat. Korelasi yang terlalu rendah (< 0.3) menunjukkan bahwa variabel tersebut tidak cukup berhubungan dengan variabel lainnya untuk dianalisis dalam satu struktur komponen atau faktor.

Kesimpulannya, setelah data melalui uji normalitas dan uji korelasi *Spearman*, peneliti dapat lebih yakin bahwa data memenuhi syarat dasar yang dibutuhkan dalam analisis multivariat. Uji ini juga berfungsi sebagai dasar untuk menentukan pendekatan yang paling tepat dalam tahapan analisis selanjutnya.

2.4 Analisis Komponen Utama (*Principal Component Analysis/PCA*)

Dalam penelitian ini, salah satu teknik yang digunakan untuk menganalisis variabel kualitas air secara menyeluruh adalah analisis komponen utama (*Principal Component Analysis/PCA*). PCA merupakan metode statistik yang digunakan untuk menyederhanakan data yang memiliki banyak variabel tanpa menghilangkan informasi penting di dalamnya. Teknik ini sangat bermanfaat dalam pengolahan data multivariat, terutama ketika terdapat banyak variabel yang saling berkaitan, seperti pada kualitas air. Dengan menggunakan PCA, peneliti dapat

mengidentifikasi pola, mengurangi kompleksitas data, dan menyoroti variabel-variabel yang paling mempengaruhi kualitas air.

PCA tidak hanya memudahkan interpretasi data, tetapi juga dapat digunakan sebagai dasar untuk analisis lanjutan, seperti klasifikasi, pemetaan, atau pengendalian kualitas multivariat. Oleh karena itu, dalam studi ini PCA menjadi langkah penting untuk memperoleh pemahaman yang lebih mendalam terhadap kondisi kualitas air produksi PDAM Kota Padang.

2.4.1 Pengertian PCA

Analisis komponen utama (*Principal Component Analysis/PCA*) adalah teknik reduksi dimensi yang digunakan dalam analisis multivariat untuk merangkum informasi dari sejumlah besar variabel ke dalam sejumlah kecil komponen utama (*principal components*) yang saling bebas (tidak berkorelasi). Komponen-komponen ini merupakan kombinasi linier dari variabel-variabel awal dan disusun sedemikian rupa sehingga komponen pertama (PC1) memuat variasi data terbesar, diikuti oleh komponen kedua (PC2), dan seterusnya (Jolliffe *et al.*, 2016).

PCA juga memungkinkan kita untuk melihat struktur tersembunyi dalam data dan mengetahui variabel mana yang memberikan kontribusi terbesar terhadap perbedaan antar observasi. Teknik ini sering digunakan dalam bidang lingkungan, termasuk dalam analisis kualitas air, karena dapat mengurangi jumlah variabel yang dianalisis tanpa kehilangan informasi penting.

2.4.2 Tujuan dan Fungsi PCA

Tujuan utama dari analisis komponen utama (*Principal Component Analysis/PCA*) adalah untuk menyederhanakan struktur data yang kompleks dengan banyak variabel, menjadi bentuk yang lebih ringkas tanpa kehilangan informasi penting di dalamnya. PCA berfungsi untuk mengurangi jumlah variabel yang dianalisis, namun tetap mempertahankan sebagian besar variasi yang ada di dalam data. Hal ini sangat membantu ketika kita ingin memahami pola umum dari data yang besar dan saling berkaitan.

Dalam konteks analisis kualitas air, PCA digunakan untuk mengidentifikasi variabel-variabel yang paling memengaruhi kondisi air. Dengan teknik ini, peneliti dapat mengetahui kombinasi variabel yang memiliki kontribusi paling besar

terhadap variasi kualitas air, sehingga lebih mudah untuk dilakukan interpretasi dan pengambilan keputusan.

PCA juga berfungsi untuk mendeteksi adanya korelasi atau hubungan antarvariabel, sehingga dapat digunakan sebagai dasar dalam proses pemodelan, klasifikasi, ataupun pengendalian mutu. Dengan kata lain, PCA tidak hanya membantu dalam merangkum data, tetapi juga membuka wawasan tentang bagaimana struktur data tersebut bekerja secara keseluruhan (Abdi *et al.*, 2010).

2.4.3 Asumsi dan Syarat PCA (KMO, Bartlett, Korelasi Antar Variabel)

Sebelum menerapkan *Principal Component Analysis* (PCA), terdapat beberapa asumsi dan prasyarat yang perlu dipenuhi agar hasil analisis valid dan dapat diinterpretasikan secara bermakna. PCA merupakan teknik statistik yang sensitif terhadap struktur hubungan antar variabel, sehingga kualitas data sangat memengaruhi hasil akhir analisis.

1. Hubungan antar Variabel

PCA bertujuan menyederhanakan data berdasarkan pola korelasi antar variabel. Oleh karena itu, diperlukan adanya hubungan yang cukup kuat antar variabel. Jika sebagian besar korelasi antar variabel sangat rendah (misalnya di bawah 0,3), maka PCA tidak akan menghasilkan komponen yang bermakna.

2. Standardisasi Data

Langkah penting sebelum melakukan PCA adalah standardisasi data, terutama ketika data memiliki satuan dan skala yang berbeda antar variabel. Hal ini dilakukan agar tidak ada variabel yang mendominasi hasil analisis hanya karena nilai angkanya besar. Dalam penelitian ini, seluruh 20 variabel kualitas air PDAM Kota Padang telah distandardisasi menggunakan *Python* agar memiliki nilai rata-rata = 0 dan standar deviasi = 1. Proses standardisasi dilakukan dengan rumus:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2.4.1)$$

keterangan:

- X_{ij} : nilai observasi ke-I pada variabel ke-j
- $\bar{x}_j$: rata-rata variabel ke-j
- s_j : standar deviasi variabel ke-j
- Z_{ij} : data hasil standardisasi.

Data yang telah distandardisasi, PCA kemudian dilanjutkan dengan membentuk matriks korelasi antar variabel.

3. *Kaiser-Meyer-Olkin (KMO) Test*

Uji *Kaiser-Meyer-Olkin* (KMO) digunakan untuk mengevaluasi kecukupan sampel dalam penerapan analisis komponen utama (PCA). KMO mengukur sejauh mana variabel-variabel dalam dataset dapat dikelompokkan berdasarkan struktur korelasi parsial. Nilai KMO berada dalam rentang 0 hingga 1, dan mengindikasikan apakah korelasi antar variabel cukup kuat untuk dianalisis secara multivariat. Nilai KMO yang tinggi menunjukkan bahwa pola hubungan antar variabel memadai dan data layak dianalisis lebih lanjut menggunakan PCA.

Rumus KMO dapat dituliskan sebagai berikut:

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum q_{ij}^2} \quad (2.4.2)$$

keterangan:

- r_{ij} : Korelasi sederhana antar variabel ke-i dan ke-j
- q_{ij} : Korelasi parsial antara variabel ke-i dan ke-j

Artinya, nilai KMO membandingkan jumlah kuadrat korelasi sederhana dengan jumlah kuadrat korelasi sederhana ditambah korelasi parsial. Jika nilai korelasi parsial relatif kecil terhadap korelasi sederhana, maka nilai KMO akan tinggi, menandakan struktur korelasi yang kuat.

Interpretasi umum nilai KMO menurut Kaiser (1974) adalah sebagai berikut:

- 0.90 – 1.00: Sangat Baik
- 0.80 – 0.89: Baik
- 0.70 – 0.79: Cukup
- 0.60 – 0.69: Sedang
- 0.50 – 0.59: Buruk (masih bisa diterima)
- < 0.50: Tidak layak untuk PCA

Nilai KMO yang berada di atas 0,5 menunjukkan bahwa pola korelasi antar variabel cukup baik untuk dilakukan PCA. Sebaliknya, jika nilainya kurang dari 0,5, maka PCA sebaiknya tidak dilakukan karena struktur data tidak mendukung untuk reduksi dimensi.

4. *Bartlett's Test of Sphericity*

Uji *Bartlett's Test of Sphericity* digunakan untuk menilai apakah matriks korelasi antar variabel secara statistik berbeda signifikan dari matriks identitas. Jika matriks korelasi menyerupai matriks identitas (di mana nilai diagonal = 1 dan elemen lainnya = 0), maka hal ini mengindikasikan bahwa tidak terdapat korelasi antar variabel dan analisis komponen utama (PCA) tidak layak dilakukan.

Uji *Bartlett* menguji hipotesis nol:

- H_0 : Matriks korelasi adalah matriks identitas (tidak ada korelasi antar variabel).
- H_1 : Matriks korelasi bukan matriks identitas (ada korelasi yang cukup antar variabel).

Rumus statistik uji *Bartlett* dapat dituliskan sebagai:

$$x^2 = \left(n - 1 - \frac{2p+5}{6} \right) \ln|\mathbf{R}| \quad (2.4.3)$$

keterangan:

- x^2 : nilai statistik *chi-square*
- n : jumlah sampel
- p : jumlah variabel
- $|\mathbf{R}|$: determinan dari matriks korelasi $\mathbf{R}$
- $\ln$: logaritma natural.

Jika hasil perhitungan menghasilkan $p\text{-value} < 0,05$ (taraf signifikansi 5%), maka hipotesis nol (H_0) ditolak, yang berarti terdapat korelasi signifikan antar variabel, sehingga data memenuhi syarat untuk dilakukan PCA (Bartlett, 1954).

Dengan terpenuhinya tiga syarat utama berikut:

1. Terdapat korelasi yang cukup antar variabel (hasil korelasi $> 0,3$),
2. Nilai KMO $\geq 0,5$,
3. $p\text{-value}$ uji *Bartlett* $< 0,05$

maka dapat disimpulkan bahwa data layak dianalisis menggunakan PCA. Selain itu, proses standardisasi data yang telah dilakukan juga memastikan bahwa semua variabel memiliki skala kontribusi yang seimbang dalam proses analisis, sehingga hasil PCA menjadi lebih valid dan interpretatif.

2.4.4 Langkah-langkah Analisis Komponen Utama (PCA)

Analisis Komponen Utama (*Principal Component Analysis/PCA*) dilakukan melalui beberapa tahapan yang bertujuan untuk menyederhanakan kumpulan variabel yang saling berkorelasi menjadi sejumlah komponen utama (*principal components*) yang lebih sedikit, namun tetap mampu menjelaskan sebagian besar informasi dari data asli. Langkah-langkah utama dalam penerapan PCA dijelaskan sebagai berikut:

1. Standardisasi Data

Langkah awal adalah memastikan data sudah dalam bentuk skor standar (*z-score*), sehingga setiap variabel memiliki rata-rata nol dan standar deviasi satu. Proses dan rumus standardisasi telah dijelaskan pada Subbab 2.4.3 mengenai Asumsi PCA. Dengan demikian, semua variabel memiliki skala kontribusi yang seimbang dalam analisis (Jolliffe et al., 2016).

2. Membentuk Matriks Korelasi atau Kovarians (**R**)

Setelah data distandardisasi, langkah selanjutnya adalah membentuk matriks korelasi antar variabel atau matriks kovarians. Matriks ini digunakan untuk menggambarkan kekuatan dan arah hubungan linear antar semua pasangan variabel dalam data (Johnson & Wichern, 2007; Hair *et al.*, 2010).

Jika data belum distandardisasi, maka yang digunakan adalah **matriks kovarians**, yang setiap elemennya dihitung dengan rumus:

$$S_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{n-1} \quad (2.4.4)$$

keterangan:

- S_{ij} = kovarians antara variabel ke- i dan ke- j
- x_{ki} = nilai observasi ke- k pada variabel ke- i
- $\bar{x}_i$ = rata-rata variabel ke- i
- n = jumlah observasi

Namun, jika data sudah distandardisasi, maka yang digunakan adalah matriks korelasi, yang elemennya dihitung dari kovarians yang telah dinormalisasi:

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} \cdot S_{jj}}} \quad (2.4.5)$$

keterangan:

- r_{ij} = koefisien korelasi *Pearson* antara variabel ke- i dan ke- j
- S_{ij} = kovarians antara variabel ke- i ke- j
- S_{ii} dan S_{jj} = varians masing-masing variabel

Tahapan pada penelitian ini seluruh variabel telah melalui proses standardisasi, maka digunakan matriks korelasi untuk analisis lebih lanjut. Bentuk umum (notasi simbolik) dari matriks korelasi atau kovarians adalah:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (2.4.6)$$

dengan p menyatakan jumlah variabel yang dianalisis.

3. Menentukan Nilai *Eigen* (*Eigenvalues*) dan Vektor *Eigen* (*Eigenvectors*)

Nilai *eigen* (λ) menunjukkan seberapa besar variasi dalam data yang dapat dijelaskan oleh masing-masing komponen. Komponen utama dipilih berdasarkan besar nilai *eigen*-nya. Sedangkan vektor *eigen* menunjukkan arah dari masing-masing komponen utama dalam ruang multidimensi data.

Persamaan dasar:

$$(R - \lambda I)\vec{v} = 0 \quad (2.4.7)$$

keterangan:

- R : matriks korelasi
- λ : nilai *eigen*
- I : matriks identitas
- $\vec{v}$: *eigenvector*

Untuk menentukan nilai *eigen*:

$$\det(R - \lambda I) = 0 \quad (2.4.8)$$

(lalu dihitung akar-akarnya)

4. Menentukan Jumlah Komponen Utama

Tidak semua komponen utama digunakan dalam analisis. Komponen yang dipilih biasanya adalah yang memiliki nilai *eigen* lebih dari 1 (dikenal sebagai *Kaiser's Criterion*). Selain itu, *scree plot* juga dapat digunakan untuk menentukan

titik "*elbow*", yaitu titik di mana tambahan komponen tidak lagi menambah banyak informasi. Misalnya kita ambil dua *eigenvector* $\vec{v}_1$ dan $\vec{v}_2$, maka:

$$PC_1 = v_{11}Z_1 + v_{12}Z_2 + \dots + v_{1p}Z_p \quad (2.4.9)$$

$$PC_2 = v_{21}Z_1 + v_{22}Z_2 + \dots + v_{2p}Z_p$$

- PC_1, PC_2 : komponen utama
- Z_i : variabel hasil standardisasi
- v_{ij} : koefisien dari *eigenvector* ke-i dan variabel ke-j.

5. Membentuk Komponen Utama

Komponen utama dihitung sebagai kombinasi linier dari variabel-variabel asli dengan menggunakan bobot dari vektor *eigen*. Hasil ini disebut juga dengan skor komponen.

Gunakan:

- *Kaiser Criterion* → pilih komponen dengan $\lambda > 1$
- *Scree Plot* → titik “tekuk” menunjukkan batas optimal jumlah komponen.

6. Interpretasi Komponen

Komponen utama yang terbentuk kemudian diinterpretasikan berdasarkan *loading factor*, yaitu kontribusi masing-masing variabel terhadap komponen. Komponen yang memiliki *loading* tinggi dari variabel-variabel tertentu dapat diberi nama atau interpretasi sesuai dengan karakteristik variabel tersebut.

Dalam analisis komponen utama, transformasi data dilakukan dengan membentuk kombinasi linier dari variabel-variabel asli. Komponen utama ke- i (Z_i) dapat dinyatakan sebagai:

$$Z_i = a'_i X = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (2.4.10)$$

keterangan:

- Z_i = komponen utama ke- i
- $a'_i X$ = vektor *eigen* transpose ke- i
- X_j = variabel asli ke- j
- a_{ij} = koefisien atau bobot komponen utama (*eigenvector*)
- p = jumlah variabel asli

Koefisien a_{ij} diperoleh dari *eigenvector matriks kovarians* atau matriks korelasi data, dan komponen utama (Z_i) membentuk sumbu baru dengan variansi

maksimum secara berurutan. Variansi dari masing-masing Z_i disebut sebagai nilai *eigen*, yang mencerminkan kontribusi relatif dari komponen tersebut terhadap total variansi data. Dengan demikian, komponen utama pertama (Z_1) merupakan kombinasi linier yang menangkap variansi terbesar, diikuti oleh komponen utama kedua (Z_2), dan seterusnya, dengan asumsi antar komponen saling tidak berkorelasi.

Melalui tahapan-tahapan di atas, PCA mampu menyederhanakan kompleksitas data kualitas air yang terdiri dari banyak variabel menjadi beberapa komponen utama yang saling bebas (tidak berkorelasi). Hasil PCA tidak hanya memberikan gambaran struktur data secara menyeluruh, tetapi juga menjadi landasan kuat dalam proses analisis faktor dan evaluasi pengendalian kualitas menggunakan pendekatan statistik multivariat.

2.4.5 Interpretasi Hasil Analisis Komponen Utama (PCA)

Setelah mendapatkan komponen utama dari proses PCA, langkah selanjutnya adalah melakukan interpretasi terhadap hasil tersebut. Interpretasi bertujuan untuk memahami makna dari setiap komponen dan bagaimana kontribusi masing-masing variabel dalam membentuk komponen tersebut.

Setiap komponen utama merupakan kombinasi linier dari variabel-variabel asli. Besarnya kontribusi tiap variabel terhadap suatu komponen ditunjukkan oleh faktor muat (*loading factor*). Semakin besar nilai absolut dari *loading*, maka semakin besar pula peran variabel tersebut dalam membentuk komponen tersebut. Biasanya, *loading* yang bernilai lebih dari 0,5 (baik positif maupun negatif) dianggap signifikan dan layak dipertimbangkan dalam interpretasi (Hair *et al.*, 2010).

Interpretasi juga melibatkan persentase keragaman (variansi) yang dijelaskan oleh masing-masing komponen. Komponen pertama (PC1) biasanya menjelaskan keragaman data terbesar, diikuti oleh PC2, dan seterusnya. Komponen-komponen yang memiliki kontribusi variansi besar dan *loading* yang tinggi pada variabel-variabel tertentu dapat dijadikan dasar dalam menarik kesimpulan atau mengelompokkan data.

Selain melihat *loading*, jumlah variansi yang dijelaskan oleh masing-masing komponen juga menjadi dasar penting dalam interpretasi PCA. Setiap nilai

eigen (λ_i) yang dihasilkan dari proses ekstraksi komponen utama merepresentasikan jumlah variansi yang dapat dijelaskan oleh komponen tersebut. Oleh karena itu, untuk mengetahui kontribusi relatif dari tiap komponen terhadap total variansi, digunakan rumus:

$$\text{Persentase Keragaman} = \frac{\lambda_i}{\sum \lambda_i} \times 100\% \quad (2.4.5)$$

keterangan:

- λ_i : nilai *eigen* dari komponen ke-i
- $\sum \lambda_i$: jumlah total semua nilai *eigen* dari seluruh komponen utama

Persentase ini menunjukkan seberapa besar bagian dari keseluruhan keragaman data yang dijelaskan oleh masing-masing komponen. Komponen utama pertama (PC1) umumnya memiliki nilai persentase keragaman tertinggi, yang berarti komponen tersebut paling dominan dalam menjelaskan variasi yang terjadi pada data. Semakin besar persentase keragaman suatu komponen, semakin penting peran komponen tersebut dalam representasi data awal.

Contohnya, jika pada PC1 terdapat *loading* tinggi dari variabel seperti TDS, daya hantar listrik, dan kesadahan, maka PC1 dapat diinterpretasikan sebagai "komponen kesadahan dan kandungan zat terlarut". Interpretasi seperti ini bersifat subjektif namun penting untuk menjelaskan informasi utama dari data secara ringkas.

Secara umum, hasil PCA dapat digunakan untuk:

1. Mengidentifikasi kelompok variabel yang saling berkorelasi,
2. Menyederhanakan jumlah variabel dalam analisis lanjutan,
3. Memvisualisasikan data dalam dua atau tiga dimensi untuk pengamatan pola.

Setelah melalui proses analisis komponen utama (PCA), peneliti dapat memahami pola keterkaitan antar variabel dan menentukan kelompok variabel utama yang paling berpengaruh terhadap kualitas air. Hasil dari PCA juga memberikan gambaran mengenai struktur data dan dimensi penting yang dapat digunakan untuk analisis lanjutan. Namun, untuk memperoleh pemahaman yang lebih dalam terhadap faktor-faktor yang mendasari hubungan antar variabel, maka dilakukan analisis faktor sebagai pendekatan lanjutan dalam penelitian ini.

2.5 Analisis Faktor

Setelah dilakukan analisis komponen utama (*Principal Component Analysis* /PCA) untuk mereduksi jumlah variabel dan mengidentifikasi komponen utama dalam data kualitas air, langkah selanjutnya adalah menggunakan analisis faktor untuk menggali lebih dalam mengenai struktur laten di balik data tersebut. Analisis ini digunakan untuk mengelompokkan variabel-variabel yang saling berkorelasi tinggi ke dalam satu kelompok faktor yang sama. Dengan pendekatan ini, peneliti dapat mengetahui faktor-faktor tersembunyi yang memengaruhi kualitas air, serta menyederhanakan kompleksitas data tanpa kehilangan informasi penting.

2.5.1 Pengertian Analisis Faktor

Analisis faktor adalah suatu metode statistik multivariat yang digunakan untuk mengidentifikasi sejumlah kecil faktor yang mendasari variabel-variabel yang saling berkorelasi dalam satu set data. Tujuannya adalah untuk menemukan struktur laten (faktor) yang secara konseptual dapat menjelaskan pola hubungan antar variabel yang diamati. Metode ini sangat berguna dalam penelitian yang melibatkan banyak variabel, karena dapat menyederhanakan data menjadi kelompok-kelompok variabel yang memiliki kesamaan karakteristik (Johnson *et al.*, 2007).

Analisis faktor bertujuan untuk menyederhanakan struktur korelasi antar variabel dengan mengidentifikasi sejumlah kecil faktor laten yang dapat menjelaskan hubungan tersebut. Dalam pendekatannya, setiap variabel yang diamati dianggap sebagai kombinasi dari beberapa faktor bersama dan faktor unik.

Secara matematis, hubungan antara variabel yang diamati dan faktor-faktor laten dapat dituliskan dalam model berikut:

$$X - \mu = LF + \varepsilon \quad (2.5.1)$$

keterangan:

- X : vektor variabel pengamatan
- μ : rata-rata dari masing-masing variabel
- L : matriks *loading* faktor
- F : faktor laten
- ε : *error* atau faktor spesifik (unik)

Model ini menjelaskan bahwa nilai dari masing-masing variabel bukan hanya dipengaruhi oleh satu faktor saja, melainkan oleh kombinasi beberapa faktor laten dan komponen unik masing-masing variabel. Inilah dasar dari proses ekstraksi dan rotasi dalam analisis faktor.

Secara umum, analisis faktor bertujuan untuk:

1. Mengidentifikasi hubungan laten antar variabel
2. Mengelompokkan variabel berdasarkan kemiripan atau korelasi yang tinggi
3. Menyusun indikator atau faktor baru yang dapat mewakili sekelompok variabel secara bersamaan.

Menggunakan metode ini, peneliti dapat memahami dimensi tersembunyi dari data yang kompleks, sehingga analisis menjadi lebih terarah dan interpretasi hasil menjadi lebih informatif.

2.5.2 Tujuan dan Manfaat serta Proses Perhitungan Analisis Faktor

Analisis faktor memiliki peran penting dalam penelitian yang melibatkan banyak variabel, khususnya ketika variabel-variabel tersebut saling berkorelasi. Tujuan utama dari analisis faktor adalah mengidentifikasi struktur laten yang mendasari hubungan antar variabel, sehingga sejumlah besar variabel dapat diringkas menjadi beberapa faktor yang lebih sedikit namun tetap informatif (Hair *et al.*, 2010).

Beberapa tujuan utama dari analisis faktor adalah sebagai berikut:

1. Mengelompokkan variabel yang saling berkaitan ke dalam satu faktor yang sama, untuk menyederhanakan data.
2. Mengungkap dimensi laten, yaitu konstruk atau karakteristik tersembunyi yang tidak dapat diukur secara langsung tetapi memengaruhi sejumlah variabel yang diamati.
3. Mengurangi jumlah variabel, sehingga memudahkan interpretasi data dalam analisis lanjutan.

Adapun manfaat dari penggunaan analisis faktor dalam konteks penelitian, antara lain:

1. Menyederhanakan kompleksitas data tanpa kehilangan informasi penting.
2. Membantu dalam pemodelan, seperti untuk pengembangan instrumen pengukuran atau identifikasi faktor utama yang memengaruhi suatu fenomena.

3. Mendukung pengambilan keputusan, dengan cara mengidentifikasi kelompok variabel yang relevan terhadap suatu isu atau kondisi.

Dengan manfaat tersebut, analisis faktor menjadi metode yang sangat berguna dalam mengolah data yang kompleks dan tinggi dimensinya, seperti dalam penelitian kualitas air yang melibatkan berbagai variabel kimia dan fisik.

Langkah-langkah umum analisis faktor:

1. Menyusun Model Faktor

Analisis faktor memodelkan variabel-variabel teramati sebagai kombinasi dari faktor-faktor laten:

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + e_j \quad (2.5.2)$$

keterangan:

- X_j : variabel ke-j
 - $F_1, F_2, \dots, F_m$: faktor-faktor umum (*common factors*)
 - a_{j2} : *loading* faktor dari variabel ke-j pada faktor ke-k
 - e_j : *error* spesifik variabel ke-j
2. Matriks Korelasi dan Uji Kelayakan
Sebelum lanjut, pastikan kelayakan data dengan:
 - $KMO \geq 0.5 \rightarrow$ menunjukkan cukup korelasi parsial antar variabel
 - *Bartlett's Test* signifikansi ($p < 0.05$) $\rightarrow$ menunjukkan korelasi antar variabel tidak nol.
 3. Ekstraksi Faktor

Metode umum yang digunakan adalah metode *Principal Axis Factoring* (PAF) yang bertujuan untuk menghitung nilai *eigen* dan *loading* awal dari matriks korelasi (Fabrigar *et al.*, 1999; Hair *et al.*, 2010).

. Rumus dasar ekstraksi (matriks *loading* awal):

$$Loading = \sqrt{\lambda_i} \cdot \vec{v}_i \quad (2.5.3)$$

- λ_i : nilai *eigen* dari faktor ke-i
- v_i : *eigenvector* dari matriks korelasi

Setelah proses ekstraksi faktor dilakukan dan nilai awal *loading* diperoleh, langkah selanjutnya adalah menghitung komunalitas. Komunalitas menunjukkan proporsi variansi dari masing-masing variabel yang dapat dijelaskan oleh faktor-

faktor umum yang terbentuk. Dengan kata lain, komunalitas memberikan gambaran seberapa besar kontribusi keseluruhan faktor terhadap masing-masing variabel.

Langkah ini penting untuk mengetahui apakah variabel-variabel yang diamati memiliki kecocokan yang baik dengan struktur faktor yang terbentuk. Variabel dengan nilai komunalitas tinggi berarti dapat dijelaskan dengan baik oleh faktor-faktor yang ada, sementara nilai komunalitas rendah menunjukkan bahwa variabel tersebut kurang sesuai dimasukkan dalam model faktor.

2.5.3 Komunalitas, Rotasi, dan Penamaan Faktor

Dalam analisis faktor, komunalitas menunjukkan seberapa besar proporsi variansi suatu variabel yang dapat dijelaskan oleh faktor-faktor yang diekstrak. Nilai komunalitas berada dalam rentang 0 hingga 1. Jika nilai komunalitas suatu variabel mendekati 1, artinya sebagian besar informasi dari variabel tersebut dapat dijelaskan oleh faktor-faktor yang terbentuk, sebaliknya jika mendekati 0, maka variabel tersebut kurang mampu dijelaskan oleh faktor.

Secara matematis, komunalitas (h_i^2) untuk variabel ke- i dihitung sebagai jumlah kuadrat dari *loading* faktor variabel tersebut terhadap seluruh faktor yang diambil:

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 \quad (2.5.3)$$

keterangan:

- h_i^2 = komunikasi untuk variabel ke- i .
- l_{ij} = *loading* faktor dari variabel ke- i pada faktor ke- j .
- m = jumlah faktor yang digunakan.

Setelah faktor diekstrak, dilakukan rotasi faktor untuk membuat hasil lebih mudah diinterpretasikan. Rotasi membantu memaksimalkan *loading* faktor terhadap satu faktor tertentu dan meminimalkan terhadap faktor lainnya, sehingga struktur faktor menjadi lebih jelas. Setelah rotasi, nilai *loading* akan lebih tegas yang mana nilai *loading* tersebut (≥ 0.7) menunjukkan variabel dominan pada faktor itu. Dua jenis rotasi umum adalah:

1. Rotasi *Orthogonal* (seperti *varimax*) faktor tetap saling bebas (tidak berkorelasi).
2. Rotasi *Oblique* faktor dapat berkorelasi satu sama lain.

Langkah berikutnya adalah penanaman (*naming*) faktor, yaitu memberikan nama pada setiap faktor berdasarkan variabel-variabel dengan *loading* tertinggi

dalam faktor tersebut. Penamaan ini bersifat subjektif namun harus mencerminkan makna atau tema dari kumpulan variabel yang berkaitan dalam faktor.

Menurut Supranto (2004), proses rotasi dan penamaan faktor penting untuk memperoleh makna dari hasil analisis dan menjadikan hasil tersebut dapat diinterpretasikan dalam konteks penelitian.

2.5.4 Uji Reliabilitas Faktor

Setelah menentukan struktur faktor dalam analisis faktor, langkah penting berikutnya adalah menguji reliabilitas atau konsistensi internal dari masing-masing faktor. Tujuan dari uji reliabilitas ini adalah untuk mengetahui sejauh mana variabel-variabel yang membentuk suatu faktor memiliki konsistensi dalam mengukur konsep yang sama.

Salah satu metode yang paling umum digunakan untuk mengukur reliabilitas adalah *Cronbach's Alpha*. Nilai *Cronbach's Alpha* berkisar antara 0 hingga 1. Semakin tinggi nilainya, maka semakin baik reliabilitas faktor tersebut. Secara umum, kriteria interpretasi nilai *Cronbach's Alpha* adalah sebagai berikut (Ghozali, 2016):

- Nilai $> 0,90$ = sangat baik (*excellent*)
- Nilai $0,70-0,90$ = baik (*good*)
- Nilai $0,60-0,70$ = dapat diterima (*acceptable*)
- Nilai $< 0,60$ = kurang baik (*poor*)

Persamaan *Cronbach's Alpha* dapat dituliskan sebagai berikut:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_t^2} \right) \quad (2.5.4)$$

keterangan:

- α = nilai *Cronbach's Alpha*
- k = jumlah item/variabel dalam satu faktor
- σ_i^2 = varians item ke- i
- σ_t^2 = varians total skor keseluruhan item

Jika nilai *alpha* tinggi, artinya variabel-variabel dalam faktor tersebut saling berkorelasi cukup kuat dan memiliki konsistensi dalam mengukur konsep yang sama. Sebaliknya, nilai *alpha* yang rendah menunjukkan bahwa variabel-variabel tersebut kurang dapat diandalkan sebagai pengukur satu faktor tertentu.

Reliabilitas sangat penting untuk memastikan bahwa hasil analisis faktor tidak hanya menggambarkan pola data yang kebetulan, tetapi juga menggambarkan struktur yang stabil dan dapat dipercaya untuk digunakan dalam interpretasi lebih lanjut.

Berdasarkan uraian sebelumnya, dapat disimpulkan bahwa analisis komponen utama (PCA) dan analisis faktor merupakan dua pendekatan multivariat yang saling melengkapi dalam mereduksi dimensi data serta memahami struktur keterkaitan antar variabel. PCA bertujuan menyederhanakan data dengan tetap mempertahankan informasi sebanyak mungkin, sedangkan analisis faktor lebih fokus pada pengelompokan variabel ke dalam faktor-faktor laten yang mendasarinya.

Kedua metode ini sangat bermanfaat dalam mengolah data kompleks, terutama ketika jumlah variabel cukup banyak dan saling berkorelasi. Dengan melakukan reduksi dimensi dan identifikasi struktur faktor, peneliti dapat memperoleh gambaran yang lebih jelas dan terfokus terhadap pola yang muncul dalam data.

Setelah struktur data teridentifikasi dengan baik, langkah selanjutnya adalah memastikan bahwa sistem yang diamati berada dalam kondisi terkendali. Oleh karena itu, pembahasan selanjutnya akan difokuskan pada pengendalian kualitas multivariat, khususnya menggunakan pendekatan statistik seperti Hotelling's T^2 , untuk memantau kestabilan proses berdasarkan variabel-variabel yang telah dianalisis sebelumnya.

2.6 Pengendalian Kualitas Multivariat

Pengendalian kualitas merupakan aspek penting dalam proses produksi dan distribusi, termasuk dalam pengelolaan kualitas air minum. Tujuan utama dari pengendalian kualitas adalah memastikan bahwa proses berjalan dalam batas yang telah ditentukan, sehingga produk akhir yang dihasilkan memenuhi standar mutu yang telah ditetapkan. Dalam konteks kualitas air, hal ini berarti bahwa berbagai variabel fisik dan kimia air harus tetap berada dalam rentang aman dan sesuai dengan standar kesehatan.

Ketika proses melibatkan banyak variabel yang saling berkaitan, seperti dalam pemantauan kualitas air, pendekatan pengendalian kualitas yang hanya

mempertimbangkan satu variabel (*univariate*) menjadi kurang efektif. Oleh karena itu, pendekatan pengendalian kualitas multivariat diperlukan untuk memantau semua variabel secara simultan dan mendeteksi perubahan yang mungkin terjadi akibat variasi proses.

Salah satu metode yang umum digunakan dalam pengendalian kualitas multivariat adalah diagram kendali Hotelling's T^2 . Teknik ini dikembangkan dari konsep analisis kovarians dan memungkinkan pemantauan beberapa variabel sekaligus dalam satu kerangka statistik. Dengan metode ini, penyimpangan kecil sekalipun dalam kombinasi variabel dapat terdeteksi lebih cepat dibandingkan diagram kendali univariat (Montgomery, 2009).

2.6.1 Konsep Dasar Pengendalian Kualitas

Pengendalian kualitas adalah proses sistematis yang dilakukan untuk memastikan bahwa suatu produk atau jasa sesuai dengan standar atau spesifikasi yang telah ditetapkan. Dalam praktiknya, pengendalian kualitas tidak hanya bertujuan untuk mendeteksi produk cacat, tetapi juga untuk mencegah terjadinya penyimpangan selama proses berlangsung. Salah satu cara yang umum digunakan untuk mencapai hal ini adalah dengan memanfaatkan diagram kendali (*control chart*) yang dikembangkan oleh Walter A. Shewhart pada awal abad ke-20.

Diagram kendali bekerja dengan cara memantau data secara berkala dan memberikan sinyal apabila terjadi penyimpangan dari pola yang diharapkan. Dengan demikian, pengendalian kualitas menjadi alat penting untuk menjaga stabilitas proses dan memperbaiki masalah sejak dini (Montgomery, 2009).

Dalam proses pengolahan air, pengendalian kualitas dilakukan terhadap berbagai variabel seperti pH, kekeruhan, dan kadar logam berat. Jika salah satu atau beberapa dari variabel ini menunjukkan nilai yang menyimpang secara signifikan, maka proses harus segera diperiksa dan diperbaiki. Hal ini penting agar kualitas air yang didistribusikan ke masyarakat tetap aman dan sesuai dengan ketentuan standar kesehatan.

Menurut Tjiptono (2008), konsep dasar pengendalian kualitas meliputi: identifikasi kebutuhan mutu, penetapan standar mutu, pengukuran terhadap hasil aktual, dan tindakan korektif jika terjadi ketidaksesuaian. Prinsip ini juga berlaku dalam pengendalian kualitas air yang bersifat berkelanjutan dan berbasis data.

2.6.2 Hotelling T^2

Dalam pengendalian kualitas yang melibatkan banyak variabel secara bersamaan, metode *univariat* seperti diagram kendali $\bar{X}$ dan R tidak lagi cukup efektif. Hal ini karena setiap variabel dalam proses sering kali saling berhubungan (berkorelasi), sehingga memantau masing-masing variabel secara terpisah dapat menyebabkan informasi penting terabaikan. Untuk mengatasi hal tersebut, digunakan metode pengendalian kualitas multivariat, salah satunya adalah Hotelling's T^2 .

Hotelling's T^2 adalah perpanjangan dari statistik *t-Student* yang dirancang untuk menganalisis lebih dari satu variabel sekaligus. Metode ini sangat berguna untuk mendeteksi penyimpangan dalam proses yang melibatkan hubungan antar variabel. Dalam konteks kualitas air, Hotelling's T^2 dapat membantu memantau berbagai variabel kimia dan fisika air secara bersamaan, sehingga lebih cepat mendeteksi jika ada perubahan yang tidak wajar pada kombinasi variabel tersebut (Montgomery, 2009).

Statistik T^2 dihitung dari data hasil pengamatan beberapa variabel dengan mempertimbangkan korelasi antar variabel tersebut. Nilai T^2 yang lebih besar dari batas kendali atas (*Upper Control Limit/UCL*) menunjukkan bahwa observasi tersebut berada di luar kendali dan perlu ditindaklanjuti. Prosedur ini memungkinkan identifikasi yang lebih akurat terhadap potensi penyebab masalah dalam sistem yang kompleks.

Penggunaan Hotelling's T^2 dalam pengendalian kualitas telah banyak diterapkan dalam industri makanan, farmasi, dan juga pengolahan air, karena kemampuannya dalam menangkap variasi multivariat secara lebih komprehensif dibanding metode univariat (Kourti *et al.*, 1996).

2.6.3 Diagram Kendali T^2 : Pembuatan dan Interpretasi

Diagram kendali T^2 adalah alat bantu visual dalam metode Hotelling's T^2 yang digunakan untuk memantau kestabilan suatu proses multivariat. Diagram ini menampilkan nilai statistik T^2 dari setiap observasi terhadap batas kendali atas (*Upper Control Limit/UCL*). Jika suatu titik (observasi) berada di atas garis UCL, maka observasi tersebut dianggap berada di luar kendali, dan perlu dilakukan penyelidikan lebih lanjut (Montgomery, 2009).

Langkah awal dalam pembuatan diagram kendali T^2 adalah menghitung nilai rata-rata dan matriks kovarian dari seluruh variabel yang dianalisis. Menghitung rata-rata dan matriks kovarian, untuk data multivariat standar X :

- Hitung rata-rata tiap variabel: $\bar{x}$
- Hitung matriks kovarian S

Selanjutnya, menghitung invers matriks kovarian, cari S^{-1} dari matriks kovarian menggunakan metode invers matriks (dalam *Python* sudah otomatis, tapi saat seminar hasil kamu bisa sebut pakai metode determinan dan *adjoint* jika manual).

Selanjutnya, untuk setiap observasi, dihitung nilai statistik T^2 menggunakan rumus:

$$T^2 = (\bar{x} - \mu)' S^{-1} (\bar{x} - \mu) \quad (2.6.1)$$

keterangan:

- n : jumlah observasi
- $\bar{x}$: vektor rata-rata dari seluruh data
- μ : vektor rata-rata populasi (target kualitas)
- S^{-1} : *invers* dari matriks kovarian antar variabel
- T^2 : nilai statistik Hotelling's

Rumus ini digunakan untuk menghitung nilai statistik T^2 yang mengukur seberapa jauh rata-rata observasi dari mean populasi secara multivariat. Semakin besar nilai T^2 , semakin mungkin data berada di luar kendali (Wahyuningsih *et al.*, 2005).

Setelah nilai T^2 dari semua observasi diperoleh, nilai batas kendali atas (UCL) dihitung berdasarkan distribusi F. Untuk sampel sebanyak n dengan p variabel, dan N total data, rumus:

$$UCL = \frac{p(n-1)}{n-p} F_{p, N-p, \alpha} \quad (2.6.2)$$

keterangan:

- $F_{p, N-p, \alpha}$ adalah nilai kritis dan distribusi F dengan derajat bebas p dan $N - p$ pada tingkat signifikansi α , umumnya 0,05.

Interpretasi diagram kendali T^2 cukup sederhana:

- Jika $T_i^2 < UCL$, maka proses dianggap berada dalam kendali (*in-control*).

- Jika $T_i^2 > UCL$, maka proses dianggap berada di luar kendali (*out-of-control*), dan analisis lebih lanjut perlu dilakukan untuk mengetahui penyebabnya.

Batas kendali atas (*Upper Control Limit*) ditentukan berdasarkan distribusi F, dengan p variabel dan n jumlah sampel. Titik ini menjadi batas untuk menilai apakah suatu observasi masih dalam kendali atau tidak (Wahyuningsih *et al.*, 2005).

Dengan diagram ini, peneliti dapat memantau banyak variabel kualitas secara bersamaan dan mendeteksi penyimpangan secara lebih cepat dan efisien dibandingkan metode univariat. Hal ini membuat diagram kendali T^2 sangat cocok digunakan dalam pemantauan kualitas air yang melibatkan banyak variabel secara simultan (Alt, 1985).

2.7 Analisis Data Menggunakan Python

Python merupakan bahasa pemrograman tingkat tinggi yang saat ini banyak digunakan dalam analisis data dan statistik. Kemampuannya dalam memproses data, analisis statistik, hingga visualisasi membuatnya menjadi alat yang sangat populer di kalangan peneliti dan akademisi (VanderPlas, 2016). Dalam penelitian ini, *Python* digunakan untuk melakukan berbagai tahap analisis, mulai dari eksplorasi data, uji asumsi, hingga analisis multivariat.

Beberapa pustaka (*library*) yang digunakan dalam penelitian ini antara lain:

- *Pandas*: Untuk mengelola data dalam bentuk dataframe, mempermudah pembacaan dan manipulasi data (McKinney, 2010).
- *NumPy*: Untuk perhitungan numerik dan operasi matriks yang efisien (Oliphant, 2006).
- *Matplotlib* dan *Seaborn*: Digunakan untuk membuat visualisasi grafik seperti histogram, *scatter plot*, *heatmap*, dan *scree plot* (Hunter, 2007; Waskom, 2021).
- *Scikit-learn*: *Library* yang menyediakan fungsi-fungsi untuk analisis statistik lanjutan seperti standardisasi data (*StandardScaler*), PCA, dan pemodelan multivariat lainnya (Pedregosa *et al.*, 2011).
- *SciPy* dan *Statsmodels*: Untuk melakukan uji statistik seperti uji normalitas (*Shapiro-Wilk*) dan korelasi *rank Spearman* (Jones *et al.*, 2001; Seabold *et al.*, 2010).

Penggunaan *Python* dalam penelitian ini memungkinkan setiap langkah analisis dilakukan secara sistematis dan replikatif, sekaligus menghasilkan visualisasi yang mendukung interpretasi data secara lebih komprehensif. *Python* juga membantu meningkatkan efisiensi dan akurasi dalam pengolahan data, terutama saat menangani dataset yang kompleks dan multidimensi.

2.8 Penelitian Terdahulu

Penelitian mengenai analisis kualitas air minum menggunakan pendekatan multivariat telah dilakukan oleh berbagai peneliti sebelumnya. Salah satu penelitian yang menjadi acuan utama dalam skripsi ini adalah studi oleh Wahyuningsih dan Pusdikarta (2005) yang berjudul Analisis Pengendalian Kualitas Multivariate Air Minum (Studi Kasus di PDAM Gresik).

Penelitian tersebut bertujuan untuk mengidentifikasi variabel-variabel yang paling berpengaruh terhadap kualitas air serta mengevaluasi apakah variabel-variabel tersebut telah berada dalam kondisi terkendali secara statistik. Dalam analisisnya, digunakan dua pendekatan utama yaitu *Principal Component Analysis* (PCA) dan analisis faktor, yang kemudian dilanjutkan dengan pengendalian kualitas multivariat menggunakan diagram kendali Hotelling's T^2 .

Dari hasil penelitian tersebut, ditemukan enam variabel penting yang memengaruhi kualitas air di PDAM Gresik, yaitu padatan total, kesadahan, kekeruhan, timbal (Pb), besi (Fe), dan suhu. Hasil pengujian lebih lanjut menunjukkan bahwa dari keenam variabel tersebut, tiga di antaranya padatan total, suhu belum terkendali secara statistik. Oleh karena itu, digunakan diagram kendali T^2 Hotelling untuk memantau kestabilan proses berdasarkan kombinasi dari beberapa variabel sekaligus.

Selain itu, dilakukan juga evaluasi terhadap kemampuan proses dengan menghitung indeks kapabilitas proses multivariat. Hasil analisis menunjukkan bahwa proses pengolahan air untuk kedua kelompok faktor (faktor 1 dan faktor 2) dinilai kapabel karena memiliki nilai $C_p > 1$, yang berarti proses berada dalam kondisi baik dan sesuai spesifikasi.

Penelitian ini relevan dan menjadi dasar bagi studi yang dilakukan oleh penulis, terutama dalam hal struktur metode yang digunakan, yaitu penggabungan

PCA, analisis faktor, dan Hotelling's T^2 untuk mengevaluasi data kualitas air PDAM secara menyeluruh.

2.9 Kerangka Pemikiran

Kerangka pemikiran berfungsi sebagai panduan logis yang menjelaskan alur berpikir peneliti dalam menyusun dan menjalankan penelitian. Dalam penelitian ini, kerangka pemikiran disusun berdasarkan konsep kualitas air, pendekatan multivariat, serta teknik pengendalian kualitas statistik.

Air minum yang diproduksi oleh PDAM harus memenuhi standar kualitas yang ditetapkan agar aman dikonsumsi masyarakat. Kualitas air ini diukur berdasarkan berbagai variabel fisik dan kimia seperti TDS, pH, kekeruhan, suhu, logam berat, dan zat organik. Karena jumlah variabel cukup banyak dan saling berkorelasi, maka diperlukan pendekatan statistik yang mampu menangani data multivariat secara efektif (WHO, 2017; Permenkes No. 492/Menkes/Per/IV/2010).

Untuk memahami karakteristik utama dari data yang kompleks tersebut, digunakan metode *Principal Component Analysis* (PCA). Teknik ini membantu mereduksi jumlah variabel dengan cara menggabungkannya ke dalam beberapa komponen utama yang dapat menjelaskan sebagian besar variansi data (Jolliffe *et al.*, 2016). Selanjutnya, digunakan analisis faktor untuk mengelompokkan variabel-variabel yang berkorelasi ke dalam satu faktor laten, sehingga struktur data menjadi lebih jelas dan bermakna (Hair *et al.*, 2010).

Setelah faktor terbentuk, dilakukan uji reliabilitas untuk memastikan bahwa kelompok variabel dalam suatu faktor memiliki konsistensi yang cukup. Kemudian, untuk mengevaluasi apakah proses kualitas air berada dalam kondisi yang stabil, digunakan pendekatan pengendalian kualitas multivariat dengan metode Hotelling's T^2 *control chart*. Metode ini dapat memantau kestabilan proses berdasarkan kombinasi dari beberapa variabel secara simultan (Montgomery, 2009).

Melalui alur pemikiran ini, penelitian diharapkan mampu memberikan gambaran yang utuh mengenai struktur variabel kualitas air dan mendeteksi potensi penyimpangan proses dalam produksi air minum PDAM Kota Padang.

Bab ini telah menguraikan berbagai teori dan konsep penting yang menjadi dasar dalam penelitian, dimulai dari pengertian kualitas air serta variabel-variabel

fisik dan kimia yang digunakan dalam penilaian kualitas air produksi. Kemudian dijelaskan pula teori-teori statistik multivariat yang relevan, termasuk uji prasyarat seperti uji normalitas dan korelasi *Spearman*, yang penting dilakukan sebelum menerapkan metode analisis multivariat.

Selanjutnya, disajikan pembahasan mendalam mengenai PCA sebagai metode reduksi dimensi untuk menyederhanakan kompleksitas data, diikuti dengan analisis faktor sebagai tahap lanjutan dalam memahami struktur laten dari variabel-variabel yang telah direduksi. Penjelasan tentang Hotelling's T^2 *Control Chart* juga diberikan sebagai pendekatan statistik untuk memantau kestabilan proses secara simultan dengan mempertimbangkan seluruh variabel sekaligus.

Sebagai tambahan, dijelaskan pula mengenai penggunaan *Python* sebagai alat bantu analisis yang modern dan efisien dalam pengolahan data multivariat. Keberadaan *Python* dan pustaka-pustaka pendukungnya memungkinkan implementasi metode statistik dilakukan secara akurat, transparan, dan mudah direplikasi.

Landasan teori yang komprehensif dan didukung oleh sitasi dari berbagai sumber terpercaya, Bab ini diharapkan dapat memberikan dasar pemahaman yang kuat terhadap pendekatan metodologis yang digunakan dalam penelitian. Pemahaman tersebut akan menjadi pijakan penting dalam tahap analisis data pada Bab selanjutnya.