

BAB I PENDAHULUAN

1.1. Latar Belakang

Perubahan kualitas udara merupakan masalah yang sangat serius bagi masyarakat dan pemerintah di seluruh dunia karena udara merupakan sumber kehidupan utama bagi manusia dan makhluk hidup lainnya. Oksigen dibutuhkan untuk bertahan hidup, sehingga pencemaran udara akibat berbagai aktivitas manusia, seperti asap rokok, kegiatan industri, transportasi, serta pembakaran hutan dapat berpotensi menimbulkan sejumlah penyakit pada manusia, termasuk gangguan pernapasan, kanker paru-paru, penyakit jantung, hingga kematian. Perubahan kualitas udara selain dapat memengaruhi udara secara langsung juga dapat memengaruhi tanah, air, dan infrastruktur bangunan. Semua ini memberikan efek negatif yang luas pada kehidupan di bumi (Fath, 2021).

Kota-kota besar sering menghadapi masalah polusi udara, salah satunya yaitu kota Makassar dengan tingkat kemajuan yang cukup pesat dalam berbagai bidang, seperti bidang transportasi. Dampak dari majunya industri transportasi sendiri adalah jumlah kendaraan yang terus bertambah setiap tahunnya, yang menyebabkan padatnya kendaraan dan bertambahnya polusi udara. Berdasarkan data dari Samsat Kota Makassar, pada Oktober 2024 jumlah kendaraan bermotor tercatat 2.063.618 unit atau bertambah 207.315 unit dibandingkan 2023. Kondisi ini tentu sangat mengkhawatirkan mengingat polutan yang dikeluarkan oleh kendaraan bermotor. Hasil penelitian Hasna dkk. (2024) terlihat bahwa tiga jenis polutan yang paling besar menjadi pemicu terjadinya pencemaran udara adalah Partikulat PM_{2,5}, Partikulat PM₁₀, dan Karbon Monoksida (CO). Jika senyawa-senyawa tersebut melampaui ambang batas yang dapat diterima, maka berpotensi membahayakan sistem pernafasan, bahkan menyebabkan kematian. Untuk memastikan kualitas udara tetap dalam batas aman, diperlukan metode pemantauan yang efektif dan akurat.

Pengendalian kualitas statistik atau *Statistical Process Control* (SPC) merupakan salah satu bagian yang penting dalam pemantauan suatu proses agar kualitas yang dihasilkan tetap terjaga. Pada teknik SPC terdapat alat untuk mengukur proses dalam keadaan terkendali atau tidak secara statistika yaitu bagan kendali (Gaspersz, 2004). Keadaan terkendali apabila titik-titik berada diantara batas pengendali sedangkan keadaan tidak terkendali apabila terdapat satu atau beberapa titik yang terletak di luar batas pengendali atau apabila titik-titik dalam grafik menunjukkan pola yang tidak random. Bagan kendali terbagi dua, yaitu bagan kendali univariat dan multivariat. Bagan kendali univariat digunakan untuk mengontrol rata-rata dari satu variabel. Namun, dalam banyak kasus seperti pada penelitian ini yang terdiri dari tiga variabel yang saling berkaitan yang lebih relevan adalah menggunakan bagan kendali multivariat mampu mengontrol proses secara bersamaan dengan hubungan antar variabel.



Multivariate Exponentially Weighted Moving Average (MEWMA) adalah salah satu bagan kendali multivariat untuk mendeteksi pergeseran *mean vector* (Montgomery, 2020). Statistik bagan kendali MEWMA mengakumulasikan informasi dari seluruh pengamatan sehingga lebih sensitif untuk mendeteksi pergeseran yang kecil. Bagan kendali MEWMA merupakan pengembangan dari bagan kendali *Univariate Exponentially Weighted Moving Average* (EWMA) dan salah satu bagan kendali multivariat yang dapat mendeteksi pergeseran rata-rata proses dari suatu data berdistribusi normal multivariat (Lowry dkk., 1992). Keunggulan dari bagan kendali MEWMA adalah memiliki kemampuan yang cukup tinggi untuk mendeteksi pergeseran rata-rata proses yang relatif kecil dibanding dengan bagan kendali yang lain. Dalam penerapannya, terdapat beberapa kendala, salah satunya adalah data yang tidak berdistribusi normal multivariat. Jika data yang dianalisis tidak memenuhi asumsi ini, ketepatan dan keandalan bagan kendali bisa menurun sehingga kemampuan untuk mendeteksi perubahan dalam proses menjadi kurang akurat. Oleh karena itu, diperlukan pendekatan yang lebih fleksibel untuk mengatasi ketidaksesuaian distribusi data. Salah satu pendekatan yang sesuai adalah penggunaan Copula.

Copula adalah fungsi yang menggabungkan distribusi multivariat dengan fungsi distribusi marginal satu dimensi, dimana margin satu dimensi tersebut seragam pada interval $[0,1]$ yang memungkinkan pemodelan ketergantungan non-linear secara lebih akurat (Nelsen, 2006). Copula memiliki beberapa tipe yang dapat digunakan, tergantung pada parameter dan banyak variabel yang digunakan. Pemodelan menggunakan copula memiliki beberapa kelemahan, terutama pada kasus multivariat, di mana penentuan fungsi copula menjadi sulit dan pilihan keluarga copula yang tersedia terbatas. Oleh karena itu, jika struktur ketergantungan dalam data multivariat lebih kompleks dapat digunakan pendekatan Vine Copula. Konsep dasar Vine Copula adalah mendekomposisi fungsi copula multivariat menjadi kumpulan fungsi copula berpasangan (copula bivariat). Fungsi copula berpasangan ini diperoleh dari keluarga copula bivariat (Joe, 1997). Di antara model Vine-copula, C-vine dan D-vine telah banyak diterapkan di berbagai bidang. C-vine lebih baik daripada D-vine jika kita mengetahui variabel utama yang mengatur interaksi di antara variabel yang dipilih (Zhang dan Singh, 2019).

Beberapa penelitian sebelumnya yang berkaitan dengan bagan kendali MEWMA dengan Copula diantaranya, dilakukan oleh Busababodhin dan Amphanthong (2016) mengkaji penggunaan Copula dalam pengendalian kualitas statistik multivariat. Hasil penelitian mereka menunjukkan bahwa Copula Gumbel lebih unggul dalam



ran besar pada bagan kendali MEWMA. Selain itu, bagan copula secara umum lebih efektif dibandingkan bagan kendali menangani data dengan ketergantungan positif maupun negatif. ia dkk (2018) membandingkan performa bagan kendali EWMA dengan pendekatan Copula Bivariat untuk mendeteksi enelitiannya menunjukkan bagan kendali MEWMA berbasis

Copula memiliki sensitivitas yang lebih tinggi dalam mendeteksi pergeseran rata-rata proses, dengan performa yang diukur menggunakan nilai *Average Run Length* (ARL). Selanjutnya, Shofwah (2019) juga menerapkan bagan kendali MEWMA dengan Copula Clayton menggunakan data simulasi. Penelitiannya menunjukkan bahwa proses pengendalian kualitas berada dalam kondisi terkendali, ditunjukkan dengan nilai plot yang seluruhnya berada di bawah batas kendali atas.

Dalam penerapannya pada data kualitas udara yang bersifat multivariat, hubungan antara variabel sering kali bersifat nonlinier dan tidak sepenuhnya dapat dijelaskan dengan asumsi normalitas sehingga bagan kendali multivariat perlu dirancang agar dapat menangani karakteristik tersebut. Salah satu pendekatan yang efektif adalah Copula, yang mampu menangkap ketergantungan antara variabel kualitas udara pada kasus multivariat, tanpa terikat pada distribusi marginal tertentu. Oleh karena itu, penulis akan mengkaji mengenai bagan kendali *Multivariate Exponentially Weighted Moving Average* (MEWMA) dengan pendekatan Copula dalam mendeteksi pergeseran rata-rata proses yang diaplikasikan pada data indeks kualitas udara.

1.2. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data sekunder berupa indeks kualitas udara yaitu PM2,5, PM10 dan CO di Kota Makassar periode 1 Januari 2024 – 31 Juli 2024.
2. Copula yang digunakan adalah C-Vine Copula dengan keluarga Copula Archimedean yang meliputi Copula Clayton, Copula Gumbel, dan Copula Frank.

1.3. Tujuan dan Manfaat Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membentuk bagan kendali *Multivariate Exponentially Weighted Moving Average* dengan pendekatan Copula pada data indeks kualitas udara di Kota Makassar.
2. Memperoleh kinerja bagan kendali *Multivariate Exponentially Weighted Moving Average* dengan pendekatan Copula pada data indeks kualitas udara di Kota Makassar.

Manfaat yang diharapkan dari penelitian ini dapat menambah pengetahuan tentang pengendalian kualitas statistik menggunakan bagan kendali MEWMA dengan pendekatan Copula. Serta dapat memberikan rekomendasi yang lebih baik dalam perumusan kebijakan dan strategi pengendalian pencemaran udara.

1.4. Kajian Teori



riat

merupakan data hasil pengamatan terhadap $p \geq 1$ variabel i hasil pengamatan dari setiap variabel disebut item atau unit menjelaskan struktur data multivariat, digunakan symbol x_{ij} yang i menunjukkan item ke- i dari variabel ke- j . Menurut Irwan dan Saudidin (2021), p variabel dapat disusun sebagai berikut:

Tabel 1. Struktur Data Multivariat

Observasi	Variabel					
	1	2	...	j	...	p
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{np}

Susunan tersebut dapat disederhanakan dalam bentuk array, yang disebut matriks X berukuran $n \times p$ sebagai berikut.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}$$

Keterangan:

x_{ij} : Pengamatan ke- i dari variabel ke- j

n : Jumlah pengamatan

p : Jumlah variabel

1.4.2. Uji Dependensi Variabel

Uji dependensi pada kasus multivariat digunakan untuk mengetahui apakah antar variabel saling berhubungan atau tidak. Untuk mengukur ketergantungan antara dua variabel atau lebih dapat digunakan koefisien korelasi Pearson, Tau Kendall, dan Spearman. Namun dalam penelitian ini peneliti menggunakan koefisien korelasi Tau Kendall. Korelasi Tau Kendall adalah korelasi yang berbasis rank dan tidak memerlukan asumsi kenormalan data. Tau Kendall mengukur bentuk bentuk dari dependensi yang disebut sebagai konkordan. Misalkan dalam uji dependensi (x_i, y_i) dan (x_j, y_j) merupakan dua vector (X, Y) dari variabel kontinu. (x_i, y_i) dan (x_j, y_j) dikatakan konkordan adalah jika $(x_i - x_j)(y_i - y_j) > 0$ dan dikatakan diskordan jika $x_i < x_j$ dan $y_i > y_j$ atau jika $x_i > x_j$ dan $y_i < y_j$. Rumus alternatif dari dua variabel bersifat konkordan adalah jika $(x_i - x_j)(y_i - y_j) > 0$ dan diskordan jika $(x_i - x_j)(y_i - y_j) < 0$ (Nelsen, 2006). Menurut Genest dan Favre (2007) pengukuran koefisien Tau Kendall secara empiris ditulis sebagai berikut.

$$\tau = \frac{P_n - Q_n}{\binom{n}{2}} \quad (1)$$



1 konkordan
1 diskordan

Optimized using
trial version
www.balesio.com

Nilai korelasi Tau Kendall yang diperoleh perlu diketahui signifikansinya untuk mengetahui dependensi antara dua variabel dalam pemodelan distribusi Bersama. Adapun hipotesis yang digunakan untuk pengujian tersebut adalah:

$H_0: \tau = 0$ (Tidak ada korelasi antar variabel)

$H_1: \tau \neq 0$ (Ada korelasi antar variabel)

Statistik uji:

$$Z_{hitung} = |\tau| \sqrt{\frac{9n(n-1)}{2(2n+5)}} \sim N(0, 1) \quad (2)$$

Kriteria uji: Dengan taraf signifikansi α , menolak H_0 jika $|Z_{hitung}| > Z_{\frac{\alpha}{2}}$ atau $p\text{-value} < \alpha$.

1.4.3. Uji Normalitas Multivariat

Pemeriksaan distribusi normalitas multivariat secara grafis dapat dilakukan menggunakan QQ plot yang didasarkan pada *square distance*. Data dapat diduga berdistribusi normal multivariat apabila hasil plot membentuk garis lurus (Johnson & Wichern, 2007). Pengujian distribusi normal multivariat dapat juga dilakukan dengan menggunakan Uji Mardia. Uji Mardia merupakan uji normalitas multivariat yang didasarkan pada *Skewness* dan *Kurtosis*. Definisikan multivariate skewness sebagai $b_{1,p}$ dan kurtosis sebagai $b_{2,p}$. Hipotesis yang akan digunakan adalah sebagai berikut.

H_0 : Data berdistribusi normal multivariat

H_1 : Data tidak berdistribusi normal multivariat

Statistik uji:

Statistik uji menggunakan Uji Mardia diperoleh menggunakan persamaan sebagai berikut (Ramdhania dan Rifai, 2022).

Untuk *skewness*:

$$z_1 = \frac{(p+1)(n+1)(n+3)}{6[(n+1)(p+1)-6]} b_{1,p} \sim \chi^2_{p(p+1)(p+2)/6} \quad (3)$$

dimana p adalah jumlah variabel dan $b_{1,p}$:

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [(x_i - \bar{x}_j)^t S^{-1} (x_i - \bar{x}_j)]^3 \quad (4)$$

an invers dari matriks kovarians



Untuk *kurtosis*:

$$z_2 = \frac{b_{2,p} - p(p+2)}{\sqrt{\frac{8p(p+2)}{n}}} \sim Z_\alpha \quad (5)$$

dimana $b_{2,p}$:

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}_j)^t S^{-1} (x_i - \bar{x}_j)]^2 \quad (6)$$

Kriteria keputusan:

Jika nilai statistik *skewness* dan *kurtosis* memenuhi $|Z_{\alpha/2}| > |z_{1 \text{ atau } 2}|$ atau $p\text{-value} > \alpha$ maka H_0 diterima atau data berdistribusi normal multivariat.

1.4.4. Pengendalian Kualitas Statistika

Pengendalian kualitas merupakan salah satu teknik yang perlu dilakukan mulai dan sebelum proses produksi berjalan, pada saat proses produksi, hingga proses produksi berakhir dengan menghasilkan produk akhir. Pengendalian kualitas dilakukan agar dapat menghasilkan produk berupa barang atau jasa yang sesuai dengan standar yang diinginkan dan direncanakan. Serta memperbaiki kualitas produk yang belum sesuai dengan standar yang telah ditetapkan dan mungkin mempertahankan kualitas yang sesuai. Menurut Vincent Gasperz (2005), pengendalian kualitas adalah teknik dan aktivitas operasional yang digunakan untuk memenuhi standar kualitas yang diharapkan. Berdasarkan pengertian tersebut, maka dapat ditarik kesimpulan bahwa pengendalian kualitas adalah suatu teknik dan aktivitas atau tindakan yang terencana yang dilakukan untuk mencapai, mempertahankan, dan meningkatkan kualitas suatu produk dan jasa agar sesuai dengan standar yang telah ditetapkan dan dapat memenuhi kepuasan konsumen.

1.4.5. Bagan Kendali

Ide pembuatan bagan kendali pertama kali dikembangkan oleh Dr. Walter Andrew Shewhart tahun 1924 yang bertujuan untuk memonitor karakteristik kualitas dari suatu proses, apakah proses tersebut sudah berjalan sesuai prosedur yang diharapkan. Hal utama yang dilakukan untuk membuat sebuah bagan kendali adalah melakukan pengambilan data sampel berukuran n . Pada masing-masing data sampel diukur atau diambil karakteristik kualitas yang menjadi perhatian. Data sampel kemudian yang menjadi *input* untuk membuat suatu bagan kendali. Bagan kendali merupakan alat untuk menganalisis apakah proses dalam menghasilkan sesuatu sudah terkendali atau tidak secara statistik. Bagan kendali dapat



k mendeteksi adanya penyimpangan dalam proses yang (Sury, 2009). Bagan kendali disertai dengan tiga garis batas yang ada atau *Center Line* (CL) merupakan garis yang menandakan penyimpangan dari karakteristik sampel. Sepasang garis lainnya garis pusat, yang disebut sebagai batas kendali atas atau *Upper Control Limit* (UCL) dan di bawah garis pusat yang disebut sebagai batas kendali *Lower Control Limit* (LCL).

1.4.6. Copula

Copula merupakan suatu fungsi yang pertama kali diperkenalkan dalam Teorema Sklar yang digunakan untuk memodelkan sebaran multivariat (Nelsen, 2006). Menurut teorema sklar, copula merupakan suatu fungsi yang menghubungkan fungsi distribusi multivariat dengan distribusi marginalnya. Fungsi copula multivariat digunakan untuk menangkap ketergantungan antara dua atau lebih variabel acak. Misalkan sebuah vektor acak $X_1, \dots, X_d$ memiliki fungsi distribusi bersama $F(x_1, \dots, x_d)$ dengan fungsi distribusi marginal kontinu $F_i(x_i) = u_i$, dimana U_i berdistribusi uniform pada interval $[0,1]$ untuk $i = 1, \dots, d$ maka terdapat sebuah Copula berdimensi- d yang unik, yaitu C .

Pola ketergantungan dari lebih dari dua variabel acak tidak dapat ditangkap secara spesifik hanya dengan menggunakan satu Copula. Oleh karena itu, Vine Copula dapat digunakan sebagai alternatif untuk memodelkan ketergantungan tersebut. Konsep dari Vine Copula adalah mendekomposisi fungsi copula multivariat menjadi fungsi copula berpasangan (copula bivariat). Fungsi copula berpasangan ini diperoleh dari keluarga copula bivariat (Joe, 1997).

1.4.6.1. Teorema Sklar

Copula berdimensi d adalah suatu fungsi distribusi yang terdefinisi pada interval $[0,1]^d$ dimana distribusi marginalnya adalah distribusi seragam. Jadi Copula dapat ditulis dalam bentuk $C(u) = C(u_1, \dots, u_d)$ untuk menyatakan fungsi distribusi multivariat. Jika F adalah suatu fungsi distribusi gabungan dengan fungsi distribusi marginal $F_1, \dots, F_d$, maka terdapat suatu Copula $C: [0,1]^d \rightarrow [0,1]$ sehingga untuk semua $x_1, \dots, x_d$ di $R = (-\infty, \infty)$ memenuhi:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (7)$$

dimana C adalah Copula d -dimensi standar yang unik. Jika $i = 1, \dots, d$ adalah kontinu pada F dan $F_1, \dots, F_d$ maka fungsi densitas bersama dari variabel acak $x_1, \dots, x_d$ dapat diperoleh dengan mendiferensialkan kedua ruas pada Persamaan (8) sehingga

$$f(x_1, \dots, x_d) = C(u_1, \dots, u_d) \prod_{i=1}^d f_i(x_i) \quad (8)$$

dimana $f_1, \dots, f_d$ adalah fungsi densitas marginal dan c disebut sebagai densitas Copula standar C yang diberikan oleh

$$C(u_1, \dots, u_d) = \frac{\partial^2 C(u_1, \dots, u_d)}{\partial u_1, \dots, \partial u_d} \quad (9)$$

Dengan demikian, jika diberikan fungsi distribusi marginal tertentu maka fungsi distribusi bersama dapat dituliskan pada bentuk berikut.

$$F(x_1, \dots, x_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (10)$$

dimana $F_i^{-1}(i = 1, \dots, d)$ adalah invers dari fungsi distribusi marginal.



Analisis Copula dilakukan dengan mentransformasikan variabel acak ke domain distribusi uniform $[0,1]$. Distribusi uniform menghasilkan nilai variabel acak memperoleh nilai peluang yang sama. Distribusi marginal dari masing-masing variabel acak X ditunjukkan persamaan berikut:

$$F(x) = \frac{1}{n+1} \sum_{j=1}^n I P(X_{i(j)} \leq x); i = 1, 2, \dots, n \quad (11)$$

dengan I adalah fungsi indikator bernilai satu jika $X_{i(j)} \leq x$. Transformasi data ke domain $[0,1]$ dengan membuat peringkat untuk setiap peubah acak sebagai berikut:

$$\left(\frac{R_{1(j)}}{n+1}\right), \left(\frac{R_{2(j)}}{n+1}\right), \dots, \left(\frac{R_{p(j)}}{n+1}\right), 1 \leq j \leq n \quad (12)$$

dengan $R_{1(j)}, R_{2(j)}, \dots, R_{p(j)}$ adalah peringkat data $X_{i(j)}$ ke- $i, i = 1, 2, \dots, n$.

1.4.6.2. Keluarga Copula

Copula secara garis besar terbagi ke dalam beberapa keluarga berdasarkan struktur dependensi yang terbentuk, diantaranya adalah Archimedean Copula, Elliptical Copula, Bivariate Extreme Value Copula, dan Marshal-Olkin Copula. Keluarga Copula yang paling populer adalah Archimedean Copula dan Elliptical Copula. Secara singkat, Copula Elliptical (normal dan Student's t) memiliki distribusi multivariat simetris, dimana copula Student's t memiliki keunggulan dibandingkan copula normal dalam hal menangkap ketergantungan ekor simetris. Copula Archimedean, yaitu Clayton dan Gumbel, memiliki distribusi asimetris di mana masing-masing menangkap ketergantungan pada ekor bawah dan ekor atas dari distribusi multivariat. Berbeda dengan copula Archimedean lainnya, copula Frank memiliki distribusi simetris dan relatif lemah dalam menangkap ketergantungan ekstrem. Dalam penelitian ini menggunakan Copula Archimedean karena lebih fleksibel menangkap berbagai struktur ketergantungan, termasuk ketergantungan yang asimetris.

1.4.6.3. Copula Archimedean

Copula Archimedean banyak dikaji dan dikembangkan karena merupakan Copula multivariat kontinu yang berbentuk sederhana dan memiliki range yang lebar dengan struktur dependensi, kemudian merupakan Copula bivariat yang sederhana dalam menggambarkan dependensi, dan pendekatan dependensi yang mudah diimplimentasikan. Copula Archimedean terdiri dari Copula Clayton, Copula Gumbel, dan Copula Frank. Copula Archimedean mempunyai bentuk distribusi dengan tail dependensi yang menghasilkan probabilitas kondisional pada daerah ekstrim. Copula Clayton memiliki tail dependensi bagian bawah, Copula Gumbel memiliki tail dependensi bagian atas, dan Copula Frank tidak memiliki tail dependensi.



Archimedean merupakan salah satu keluarga dari Copula yang dalam kasus bivariat. Secara sederhana dapat dijelaskan, jika $\varphi: [0, \infty] \rightarrow [0, 1]$ yang memiliki sifat monoton turun dan merupakan fungsi $\varphi(1) = 0$ dan $\varphi(0) = \infty$, dengan demikian maka dapat dinyatakan

$\varphi^{-1}: [0, \infty] \rightarrow [0, 1]$ (Nelsen, 2006). Fungsi Copula Archimedean $C: [0, 1]^m \rightarrow [0, 1]$ untuk copula berdimensi d dapat disajikan sebagai berikut

$$C(u_1, u_2, \dots, u_d) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_d)) \quad (13)$$

Keterangan:

φ : fungsi generator Copula dengan parameter θ

φ^{-1} : invers fungsi generator Copula

1. Copula Clayton

Copula Clayton adalah anggota keluarga Copula Archimedean yang dibentuk dari fungsi generator:

$$\varphi(u) = u^{-\theta} - 1 \quad (14)$$

Kemudian dari fungsi generator tersebut didapatkan fungsi distribusi kumulatif:

$$C^{Clay}(u_1, u_2, \dots, u_d) = \left[\left(\sum_{i=1}^d u_i^{-\theta} \right) - d + 1 \right]^{-\frac{1}{\theta}} \quad (15)$$

2. Copula Gumbel

Copula Gumbel adalah anggota keluarga Copula Archimedean yang dibentuk dari fungsi generator:

$$\varphi(u) = (-\ln(u))^\theta, \theta \geq 1 \quad (16)$$

Kemudian dari fungsi generator tersebut didapatkan fungsi distribusi kumulatif:

$$C^{Gu}(u_1, u_2, \dots, u_d) = \exp \left\{ - \left[\sum_{i=1}^d (-\ln u_i)^\theta \right]^{\frac{1}{\theta}} \right\} \quad (17)$$

3. Copula Frank

Copula Frank adalah anggota keluarga Copula Archimedean yang dibentuk dari fungsi generator:

$$\varphi(u) = -\ln \left(\frac{e^{-\theta u} - 1}{e^{-\theta} - 1} \right) \quad (18)$$

Kemudian dari fungsi generator tersebut didapatkan fungsi distribusi kumulatif:

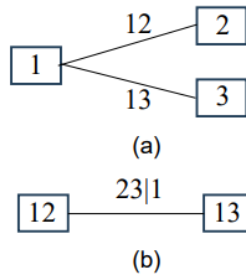
$$C^{Fr}(u_1, u_2, \dots, u_d) = -\frac{1}{\theta} \ln \left[1 + \frac{\prod_{i=1}^d [e(-\theta u_i) - 1]}{(e^{-\theta} - 1)} \right] \quad (19)$$



fungsi distribusi multi dimensi menjadi semakin rumit ketika bertambah. Vine Copula atau yang dikenal juga dengan Pair Copula adalah salah satu metode yang fleksibel dalam mengkonstruksi distribusi multivariat. Vine Copula yaitu menjelaskan bahwa copula multivariat dapat dibentuk dari pasangan-pasangan copula bivariate. Vine copula diterapkan jika kita ingin mengetahui fungsi bersamanya.

Dengan vine copula, fungsi densitas d dimensi dapat diperoleh dengan melakukan dekomposisi sehingga diperoleh hasil kali dari densitas marginalnya dengan dengan densitas copula berpasangan. Dalam melakukan pemilihan model yang tepat haruslah menentukan struktur (*tree*) keluarga copula, dan menduga parameter copula. Untuk menentukan struktur (*tree*) digunakan C-Vine copula.

C-Vine Copula memiliki struktur grafik (*tree*) dengan setiap tree memiliki node yang unik yaitu yang dihubungkan ke seluruh node. Jika terdapat d dimensi, maka banyaknya tree adalah $d - 1$. Misal, untuk 3 dimensi, maka diperoleh 2 *tree* sebagai berikut:



Gambar 1. C-Vine Copula 3 dimensi (a) *Tree 1*; (b) *Tree 2*

Gambar 1 menunjukkan bahwa variabel 1 bertindak sebagai variabel kunci dalam C-Vine yang berinteraksi dengan seluruh variabel dalam data. Oleh karena itu, pembentukan C-Vine akan menguntungkan ketika suatu variabel telah diketahui sebagai variabel kunci. Selanjutnya variabel kunci ini menjadi akar dari struktur CVine (Liu, 2011).

Sehingga fungsi densitas C-Vine menurut dekomposisi pada Gambar 1 adalah:

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{23|1}(F_2(x_2|x_1), F_3(x_3|x_1)) \quad (20)$$

Secara umum fungsi densitas, C-Vine berdimensi d adalah

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+1|1, \dots, j-1}(F(x_j|x_1, \dots, x_{j-1}), F(x_{j+1}|x_1, \dots, x_{j-1})) \quad (21)$$

Keterangan:

$f(x_i)$: fungsi densitas marginal dari X_i

$F(x_i)$: fungsi kumulatif dari X_i

c_{ij} : korelasi (X_i, X_j)

$c_{ij|k}$: korelasi (X_i, X_j) yang dikondisikan pada X_k

$c_{ij|kl}$: korelasi (X_i, X_j) yang dikondisikan pada X_k dan X_l



1.4.6.5. Estimasi Parameter Copula

Estimasi parameter Copula bisa didapatkan dengan *Maximum Likelihood Estimation* (MLE). MLE adalah metode untuk mengestimasi parameter Copula dengan memaksimalkan log *likelihood* dari fungsi kepadatan peluang bersama Copula. Nilai MLE dari sebuah parameter Copula digunakan memilih model Copula terbaik dengan mempertimbangkan nilai yang paling besar yang dihasilkan (Harsoyo, 2017). Nilai MLE dari parameter Copula akan digunakan untuk memilih model Copula mana yang paling cocok digunakan dengan mempertimbangkan nilai yang paling besar. Dasarnya, konsep dari MLE adalah mencari titik tertentu untuk memaksimalkan sebuah fungsi sehingga dapat digunakan sebagai acuan dalam pemilihan Copula dengan harapan bahwa MLE bisa menentukan model Copula mana yang paling baik digunakan disaat masing-masing fungsi berada dalam kondisi maksimal.

Berdasarkan teorema Sklar, fungsi densitas bersama f dari d -dimensi fungsi distribusi bersama f dengan fungsi distribusi marginal $F_1, F_2, \dots, F_d$ dapat direpresentasikan sebagaimana Persamaan berikut.

$$f(x_1, x_2, \dots, x_d) = c(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \cdot f_{x1}(x_1) \times \dots \times f_{xd}(x_d); x_d \in \mathbb{R} \quad (22)$$

Fungsi likelihood dapat ditulis seperti Persamaan berikut.

$$L(\theta) = \prod_{i=1}^n \left[c(F_{x1}(x_1^{(i)}), \dots, F_d(x_d^{(i)})) \cdot f_{x1}(x_1^{(i)}) \times \dots \times f_{xd}(x_d^{(i)}) \right] \quad (23)$$

Misalkan $\{x_{1i}, \dots, x_{di}\}$ merupakan sampel data matriks maka fungsi log *likelihood* sebagai berikut.

$$\begin{aligned} \ln L(\theta) = & \sum_{i=1}^n \ln c(F_{x1}(x_1^{(i)}), \dots, F_d(x_d^{(i)})) \\ & + \sum_{j=1}^m \sum_{i=1}^n \ln(f_j(x_j^{(i)})) \end{aligned} \quad (24)$$

dengan θ adalah kumpulan dari semua parameter marginal dan Copula. Oleh karena itu, diberikan fungsi probabilitas marginal dan Copula pada log *likelihood* sebelumnya dan dengan maksimisasi diberikan estimator *maximum likelihood* seperti Persamaan (27) berikut.

$$\hat{\theta}_{MLE} = \max_{\theta} L(\theta) \quad (25)$$

1.4.7. Bagan Kendali Multivariate Exponentially Weighted Moving Average

Bagan kendali *Multivariate Exponentially Weighted Moving Average* (MEWMA) digunakan untuk mendeteksi adanya rata-rata proses yang kecil secara multivariat.



MA adalah tahan terhadap asumsi distribusi normal, bila data nsi distribusi normal multivariat maka bagan kendali MEWMA an. Misalkan $X_1, X_2, \dots, X_n$ merupakan peubah acak berukuran distribusi normal multivariat dengan *vector mean* μ dan matriks sumsi masing-masing diketahui. Proses terkendali (*in control*) ak terkendali jika $\mu \neq \mu_0$ sehingga dibentuk persamaan:

$$\begin{aligned} \mathbf{Z}_0 &= \boldsymbol{\mu}_0 \\ \mathbf{Z}_i &= \lambda \mathbf{X}_i + (1 - \lambda) \mathbf{Z}_{i-1}; \quad i = 1, 2, \dots, n \end{aligned} \quad (26)$$

Keterangan:

- $\mathbf{Z}_i$: vektor statistik MEWMA pada observasi ke- i , berukuran $p \times 1$
 $\mathbf{X}_i$: vektor observasi aktual pada observasi ke- i
 λ : pembobot dengan nilai $0 < \lambda \leq 1$
 $\mathbf{Z}_{i-1}$: vektor statistik MEWMA pada observasi sebelumnya
 $\mathbf{Z}_0$: vektor 0
 p : banyak variabel yang digunakan

Adapun titik pengamatan yang akan diplot pada bagan kendali MEWMA adalah sebagai berikut:

$$T_i^2 = \mathbf{Z}_i^t \boldsymbol{\Sigma}_{\mathbf{Z}_i}^{-1} \mathbf{Z}_i \quad (27)$$

dengan T_i^2 adalah statistik Hotelling's T^2 untuk observasi ke- i , $\boldsymbol{\Sigma}_{\mathbf{Z}_i}$ adalah matriks varians dan kovariansi dari $\mathbf{Z}_i$ yang berukuran $p \times p$ yang didapat melalui persamaan sebagai berikut:

$$\boldsymbol{\Sigma}_{\mathbf{Z}_i} = \frac{\lambda(1 - (1 - \lambda)^{2i})}{2 - \lambda} \boldsymbol{\Sigma} \quad (28)$$

$\boldsymbol{\Sigma}$ merupakan matriks kovarian dari data. Nilai batas kendali atas (BKA) diperoleh berdasarkan nilai pembobot λ yang telah ditentukan dan nilai batas kendali bawah (BKB) untuk diagram kendali MEWMA sama dengan 0 karena nilai T_i^2 yang selalu positif sehingga batas pengendali bawah (LCL) yang paling minimum dari suatu nilai yang positif adalah 0. Proses dikatakan terkendali apabila nilai pengamatan lebih kecil dari batas kendali atas atau ($T_i^2 < \text{BKA}$) dan proses dikatakan tidak terkendali apabila nilai pengamatan lebih besar dari batas kendali atas atau ($T_i^2 > \text{BKA}$).

Pemilihan nilai pembobot (λ) yang optimal dapat dilakukan dengan menghitung selisih nilai maksimum MEWMA atau nilai T_i^2 tertinggi dengan batas kendali atas (BKA) dari bagan kendali MEWMA yang telah dihitung. Pembobot yang paling optimal dalam mengontrol nilai T_i^2 (titik pengamatan) pada bagan kendali MEWMA merupakan pembobot dengan nilai selisih yang paling kecil (Jayanti & Wibawati, 2014).

1.4.8. Average Run Length

Average Run Length (ARL) berfungsi untuk mengukur efektivitas kinerja bagan kendali dalam mendeteksi perubahan pada suatu proses. ARL merupakan nilai rata-rata banyak sampel yang harus diamati untuk dapat ditemukan out of control



dali yang lebih cepat mendeteksi sinyal *out of control* disebut lap perubahan proses. Secara umum persamaan untuk lah sebagai berikut:

$$ARL = \frac{1}{p} \quad (29)$$

abilitas suatu titik keluar dari batas-batas peta kendali.

Untuk ARL_0 , $p = \alpha$ = probabilitas kesalahan tipe I (menyatakan keadaan tidak terkendali padahal keadaan terkendali) atau probabilitas suatu titik rata-rata sampel berada di luar batas kendali pada saat proses terkendali, α disebut juga sebagai probabilitas *false alarm*. Untuk ARL_1 , nilai $p = 1 - \beta$ = probabilitas kesalahan tipe II (menyatakan keadaan terkendali padahal keadaan tidak terkendali) atau probabilitas suatu titik rata-rata sampel berada di dalam batas kendali pada saat proses tidak terkendali (Rasyid dkk., 2024).

ARL_0 adalah rata-rata banyak sampel yang harus diamati sampai ditemukannya pengamatan yang *out of control*, pada saat proses berada pada kondisi *in control*. Sedangkan ARL_1 adalah rata-rata banyak sampel yang diambil sebelum sinyal kontrol pertama kali mendeteksi bahwa proses mengalami perubahan ke keadaan *out of control* (Montgomery, 2012). ARL diharapkan memiliki nilai yang besar saat proses terkendali (ARL_0) dan ARL bernilai kecil saat proses tidak terkendali (ARL_1). Nilai ARL yang diperoleh dengan pendekatan hasil simulasi yang didefinisikan sebagai nilai rata-rata *Run Length* (RL) dari semua replikasi dalam simulasi. RL merupakan jumlah titik pengamatan hingga ditemukannya out of control yang pertama untuk masing-masing replikasi. Sehingga jika nilai ARL yang dihasilkan semakin kecil, maka sampel yang dibutuhkan untuk memberikan sinyal perubahan proses pun semakin kecil atau dengan kata lain semakin kecil nilai ARL maka semakin cepat pula diagram diagram kontrol dapat mendeteksi perubahan proses, sehingga diagram kontrol tersebut semakin efektif untuk mendeteksi perubahan proses (Van Delsen & Talakua, 2016).

1.4.9. Indeks Kualitas Udara

Indeks Kualitas Udara atau Indeks Standar Pencemaran Udara (ISPU) merupakan ukuran yang digunakan untuk menilai pencemaran udara di suatu daerah. Salah satu negara yang menggunakan ISPU sebagai satuan ukur kualitas udara adalah Indonesia. Indeks kualitas udara didefinisikan sebagai gambaran atau nilai hasil transformasi parameter-parameter individual polusi udara yang saling berhubungan, seperti SO_2 , NO_2 , PM, O_3 , CO menjadi satu nilai atau satu set nilai sehingga mudah dimengerti bagi masyarakat awam (Wahyudi dkk., 2023). Berdasarkan hasil penelitian Hasna dkk. (2024) terlihat bahwa 3 jenis polutan yang menjadi pemicu terbesar terjadinya pencemaran udara adalah Partikulat ($PM_{2,5}$) yang diikuti dengan Partikulat (PM_{10}), dan Karbon Monoksida (CO).

Partikulat dibagi menjadi partikulat kasar atau PM_{10} yang memiliki ukuran berdiameter $<10 \mu m$ dan partikulat halus atau $PM_{2,5}$ yang memiliki ukuran berdiameter $<2,5 \mu m$ menurut ukurannya di atmosfer. PM_{10} utamanya dihasilkan baik dari debu di permukaan jalan yang tersuspensi ke udara, industri dan agrikultur, atau bioaerosol. Sedangkan $PM_{2,5}$ ri emisi langsung dari proses pembakaran serta pembentukan kimiawi atmosfer sehingga umumnya terdiri dari lebih banyak dibandingkan dengan PM_{10} (Fan & Lin, 2011). PM_{10} menembus ke dalam paru-paru, merusak respon imun dan meletakkan dasar penyakit paru-paru. $PM_{2,5}$ menembus penghalang paru-paru dan



memasuki aliran darah, mengurangi kapasitas pembawa oksigen. Konsentrasi PM_{2,5} merupakan salah satu indikator paparan risiko Kesehatan (Hastiaty dkk., 2024). WHO sendiri memperbarui baku mutu untuk PM_{2,5} dan PM₁₀, dimana pada tahun 2005, baku mutu udara ambien harian untuk PM_{2,5} dan PM₁₀ adalah 25 µg/m³ dan 50 µg/m³, sedangkan untuk tahun 2021, baku mutu udara harian diturunkan menjadi 15 µg/m³ dan 45 µg/m³. Tujuan utama penurunan ambang batas ini adalah untuk menghindari dampak negatif PM_{2,5} dan PM₁₀ di udara pada kesehatan masyarakat (Wellid dkk., 2024).

Karbon monoksida merupakan pembunuh tersembunyi (*the silent killer*) yang tidak berwarna dan tidak berbau, namun bersifat racun. Gas karbon monoksida biasa terbentuk ketika pembakaran. CO merupakan polutan utama dari emisi kendaraan bermotor yang berdampak buruk bagi kesehatan manusia. Kemampuan CO yaitu mengikat hemoglobin darah sehingga dapat menurunkan kapasitas darah untuk mengikat oksigen (Kristianingsih dkk., 2024). Anak-anak termasuk objek yang rentan terhadap kualitas udara yang buruk disebabkan karna paru-paru mereka yang masih berkembang dan saluran udara paru-paru mereka yang kecil jika terjadi peradangan yang disebabkan oleh pencemaran udara maka saluran itu akan lebih menyempit ketimbang paru-paru milik orang dewasa. Baku mutu CO di Indonesia, menurut Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup, khususnya pada Lampiran VI tentang baku mutu udara ambien nasional yaitu untuk 1 jam maksimum 30 ppm dan 8 jam maksimum 9 ppm.



BAB II METODE PENELITIAN

2.1. Sumber Data

Penelitian ini menggunakan data sekunder berupa indeks kualitas udara atau *Air Quality Index* (AQI) di Kota Makassar. Data yang digunakan diperoleh dari website AQICN pada lokasi Makassar: <https://aqicn.org/station/indonesia-north-pisang-makassar-ujung-pandang/id/#/z/12>. Data yang digunakan merupakan data harian yang dibagi menjadi dua fase. Fase I pada periode 1 Januari 2024 sampai 30 Juni 2024 sebanyak 161 data, sedangkan fase II pada periode 1 Juli 2024 sampai 31 Juli 2024 sebanyak 31 data.

2.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini merupakan parameter utama yang dijadikan sebagai penentu kualitas udara adalah sebagai berikut.

Tabel 2. Variabel Penelitian

Variabel	Keterangan	Satuan	Definisi
X_1	<i>Particulate Matter</i> 2,5 (PM2,5)	$\mu\text{g}/\text{m}^3$	Partikel dengan diameter kurang dari 2,5 mikrometer. PM2,5 dapat masuk ke paru-paru dan aliran darah, sehingga dapat menyebabkan masalah kesehatan yang serius. PM2,5 berasal dari emisi pembakaran bensin, minyak, bahan bakar, dan kayu.
X_2	<i>Particulate Matter</i> 10 (PM10)	$\mu\text{g}/\text{m}^3$	Partikel dengan diameter kurang dari 10 mikrometer. PM10 dapat menembus jauh ke dalam paru-paru, namun tidak dapat memasuki aliran darah seperti PM2.5. PM10 dapat ditemukan di tempat pembangunan, pembuangan sampah, pertanian, kebakaran hutan, debu, serbuk sari, dan fragmen bakteri.
X_3	Karbon monoksida (CO)	ppm	Gas yang tidak memiliki bau, warna dan juga rasa sehingga indra manusia kesulitan untuk merasakannya dan CO memiliki sifat yang beracun ditambah dengan karakteristik CO yang tidak memiliki bau dan juga warna maka CO dijuluki dengan <i>silent killer</i> karena mengikat hemoglobin darah sehingga dapat menurunkan kapasitas darah untuk mengikat oksigen.



2.3. Tahapan Analisis Data

Adapun tahapan analisis yang dilakukan untuk mencapai tujuan dalam penelitian ini adalah sebagai berikut.

1. Melakukan analisis deskriptif data kualitas udara di Kota Makassar,
2. Melakukan uji dependensi pada data dengan menggunakan Persamaan (2)
3. Melakukan uji asumsi normalitas multivariat pada data asli dengan menggunakan Persamaan (3) dan Persamaan (5),
4. Mentransformasikan data ke domain Uniform $[0,1]$ dengan menggunakan Persamaan (12),
5. Menentukan struktur (*tree*) C-Vine Copula,
6. Menduga Parameter C-Vine Copula dengan keluarga Copula Archimedean menggunakan metode *Maximum Likelihood Estimation* (MLE)
7. Melakukan *fitting* C-Vine Copula dengan metode *Maximum Likelihood Estimation* (MLE) dan memilih Copula terbaik menggunakan Persamaan (23),
8. Mentransformasi data menjadi data variabel acak model Copula yang terpilih,
9. Melakukan analisis data fase I model C-Vine Copula menggunakan bagan kendali *Multivariate Exponentially Weighted Moving Average* (MEWMA). Langkah-langkah yang dilakukan adalah sebagai berikut:
 - a. Menentukan nilai pembobot (λ)
 - b. Menghitung nilai Z_i untuk masing-masing sampel menggunakan persamaan (26)
 - c. Menentukan nilai statistik T_i^2 dengan menggunakan Persamaan (27)
 - d. Membentuk bagan kendali MEWMA
 - e. Mengidentifikasi titik yang berada di luar batas kendali atau *out of control*.
10. Melakukan monitoring pada data fase II,
11. Menghitung nilai *Average Run Length* (ARL) bagan kendali MEWMA Klasik dan bagan kendali MEWMA pendekatan C-Vine Copula, kemudian membandingkan performa kedua bagan kendali tersebut.
12. Menarik Kesimpulan dari hasil analisis yang telah dilakukan.

