

BAB II

KAJIAN LITERATUR

2.1 Polusi Udara

Pengertian pencemaran udara adalah masuknya ke lingkungan zat yang berbahaya bagi manusia dan makhluk hidup lainnya. Pencemaran adalah zat cair, padat, atau gas berbahaya yang diproduksi dalam konsentrasi yang lebih tinggi dari normal dan menurunkan kualitas lingkungan kita. Polusi udara dapat mengancam kesehatan. Kesehatan orang-orang yang rentan dan sensitif dapat terpengaruh bahkan dengan polusi udara yang rendah. Paparan jangka pendek terhadap polutan udara sangat terkait dengan PPOK (Penyakit Paru Obstruktif Kronis), batuk, sesak napas, asma, dan lain-lain [31]. Adanya gas yang dikeluarkan oleh kendaraan bermotor ke udara akan memengaruhi kualitas udara. Gas yang dikeluarkan oleh kendaraan bermotor adalah CO, SO₂ dan NO₂. Keberadaan gas-gas tersebut jika berlangsung dalam waktu lama dan dalam jumlah tertentu akan menyebabkan pencemaran udara [16].

2.2 Prediksi

Prediksi adalah memperkirakan informasi masa depan menggunakan data masa lalu. Prediksi memainkan peran penting dalam berbagai bidang kehidupan, salah satunya adalah perencanaan untuk memprediksi kondisi masa depan. Fungsi lain dari prediksi adalah sebagai acuan dalam proses pengambilan keputusan. Salah satu ciri keputusan yang baik adalah berdasarkan pertimbangan terhadap sesuatu yang telah terjadi. Masa depan tidak pasti, jadi prediksi tidak pernah 100% akurat. Namun, metode penggunaan dan *hyperparameter* yang sesuai dapat mengubah prediksi tingkat kesalahan yang rendah menjadi perkiraan yang baik [32].

2.3 Data Time Series

Data *time series* adalah serangkaian data yang diamati dari suatu variabel berdasarkan deret waktu dengan rentang waktu yang sama (jam, hari, minggu, bulan, atau tahun) [33].

Data *time series* di dalam ilmu sosial dapat mencakup informasi tentang pertumbuhan populasi, pola migrasi, tingkat kelahiran, dan variabel politik [34].

2.3.1 Plot Runtun Waktu

Menurut [20] ada 5 plot *time series* yang sering digunakan atau yang dapat membantu kita dalam melakukan analisis *time series*, yaitu :

1. *Time plot*

Plot ini adalah garis yang menunjukkan evolusi deret waktu dari waktu ke waktu. Kita dapat menggunakannya sebagai titik awal analisis untuk mendapatkan pemahaman dasar tentang data, misalnya dalam hal tren, musiman, *outlier*/anomali, dan lain lain.

2. *Seasonal plot*

Plot ini sangat mirip dengan *time plot*, dengan pengecualian bahwa data di plot digambarkan terhadap masing masing musim. Musim yang ingin digambarkan juga tergantung dari sang analis, analis bisa menggambarkan data tersebut berdasarkan musim per minggu, perbulan, atau per tahun.

3. *Polar seasonal plot*

Plot ini merupakan variasi dari plot musiman, tetapi yang membedakan adalah plot ini menggunakan koordinat kutub. Plot ini juga digunakan pada kasus kasus tertentu.

4. *Time series decomposition plot*

Plot ini secara keseluruhan memberikan model yang berguna untuk memikirkan deret waktu dan membuat data lebih mudah dipahami. Dekomposisi mengasumsikan bahwa deret waktu dapat dipecah menjadi kombinasi komponen berikut :

- a. level, merupakan nilai rata rata dari data atau seri.
- b. trend, merupakan pola meningkat dan menurun dari data atau seri.
- c. seasonality, merupakan siklus jangka pendek yang berulang dalam data atau seri.
- d. noise, merupakan variasi acak yang tidak dapat dijelaskan.

5. *Autocorrelation plot*

Plot ini digunakan pada saat seseorang ingin mengukur korelasi antara rangkaian waktu dan nilai nilai yang tertinggal atau nilai *lagged*-nya (dari titik waktu sebelumnya, itu disebut autokorelasi. Ada dua jenis grafik autokorelasi yang dapat kita gunakan, yaitu :

a. *Auto Correlation Function (ACF)*

ACF menunjukkan nilai koefisien korelasi antara deret dengan nilai *lagged*-nya. ACF melihat semua komponen deret waktu (disebutkan dalam dekomposisi) sambil menemukan korelasinya. Inilah juga yang menjadi sebab ACF disebut sebagai grafik autokorelasi penuh.

b. *Partial Auto Correlation Function (PACF)*

PACF berfungsi untuk mempertimbangkan korelasi antara sisa (sisa setelah menghilangkan efek yang dijelaskan oleh *lag* sebelumnya) dan nilai *lag* berikutnya. PACF juga secara efektif menghapus variasi yang ditemukan sebelum menemukan korelasi berikutnya. Dalam percobaannya, korelasi parsial yang tinggi menunjukkan bahwa terdapat informasi dalam residual atau sisa yang dapat dimodelkan dengan *lag* selanjutnya.

2.3.2 Pola Runtun Waktu

Dikutip dari [35] ada 4 tipe pola data runtun waktu, yaitu:

1. Pola Data Noise

Pola ini terjadi ketika data biasanya tidak tersedia dalam kumpulan data tertentu dan independen, tidak berkorelasi dengan nilai lain.

2. Pola Data Musiman

Pola ini terjadi ketika data mengalami perubahan sistematis yang diberikan selama periode waktu tertentu. Periode ini bisa berupa hari, minggu, bulan, atau kuartal tertentu. Misalnya penjualan minuman dingin biasanya tinggi setiap musim panas.

3. Pola Data Siklus

Pola ini terjadi ketika fluktuasi data yang diberikan tidak tetap dan pergerakan data tidak teratur. Makanya disebut dengan data siklik.

4. Pola Data Trend

Pola ini terjadi ketika data memiliki tren naik atau penurunan jangka panjang yang terdapat dalam data.

2.4 Analisis Time Series

Runtun waktu merupakan kumpulan objek atau data observasi yang berurutan berdasarkan waktu. Metode deret waktu merupakan metode yang menganalisa hubungan antara variabel yang akan diprediksi dengan variabel waktu. Data runtun waktu (*Time Series*) merupakan jenis data berupa variabel yang sudah terurut dalam rentang waktu tertentu misalnya mingguan, bulanan, tahunan ataupun periode lainnya[22]. Analisis data runtun waktu ini memungkinkan dalam mencari tahu perkembangan suatu kejadian beserta hubungan dan pengaruhnya terhadap kejadian lainnya [23].

2.5 SARIMA

SARIMA adalah metode peramalan data time series yang merupakan evolusi dari metode ARIMA karena adanya pengaruh pola musiman. Metode SARIMA telah dipelajari secara intensif oleh George Box dan Gwilym Jenkins dan terdiri dari empat langkah yaitu identifikasi, estimasi, uji diagnostik, dan prediksi. Metode SARIMA dibentuk dari *seasonal* atau musiman, *autoregressive* (AR), *moving average* (MA) dan *integrated sample* yang ditulis dalam bentuk $(p,d,q)(P,D,Q)_s$ [24].

Metode SARIMA merupakan metode yang cocok untuk melakukan peramalan untuk jangka pendek serta menarik hubungan linear dalam data yang bersifat runtun waktu (*time series*)[25]. Model SARIMA memiliki 4 komponen yaitu :

1. Komponen *non-seasonal* dan *seasonal Auto Regressive* (AR) yang dinyatakan dalam bentuk p dan P dalam model SARIMA seperti pada persamaan (1) dan (2)

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (1)$$

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{Ps} \quad (2)$$

2. Komponen *non-seasonal* dan *seasonal Moving Average* (MA) yang dinyatakan dalam bentuk q dan Q dalam model SARIMA seperti pada persamaan (3) dan (4)

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (3)$$

$$\Theta_Q(B^s) = 1 + \theta_1 B^s + \theta_2 B^{2s} + \dots + \theta_Q B^{Qs} \quad (4)$$

3. Operator differensial *non-seasonal* dalam bentuk d yang digunakan untuk menghilangkan trend seperti pada persamaan (5)

$$(1 - B)^d \quad (5)$$

4. Operator differensial *seasonal* dalam bentuk D yang digunakan untuk menghilangkan pola seasonal seperti pada persamaan (6)

$$(1 - B^s)^D \quad (6)$$

Parameter ϕ dan θ merupakan koefisien ARMA, Φ dan Θ merupakan koefisien *seasonal* ARMA, B sebagai operator *backshift*, dengan pengaruhnya terhadap deret waktu Y_t dapat di simpulkan dengan persamaan (7) dibawah

$$B^d Y_t = Y_{t-d} \quad (7)$$

Sehingga model umum SARIMA(p, d, q)x(P, D, Q) $_s$ untuk deret waktu Y_t dapat dituliskan dengan persamaan (8) dibawah [26]

$$\phi_p(B)\Phi_p(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (8)$$

Dimana :

s = panjang dari periode(seasonalitas)

ε_t = rangkaian *white noise*

2.6 Akurasi Model

Kegiatan peramalan (*forecasting*) yang dilakukan dalam penelitian akan memiliki unsur kesalahan (*error*). Semakin rendah tingkat kesalahan pada prediksi yang dilakukan, maka akan semakin baik hasil dari prediksi tersebut[27]. Oleh karena itu penting untuk mengetahui tingkat akurasi dengan menggunakan model prediksi. Pemilihan model prediksi yang paling baik dapat ditentukan berdasarkan beberapa nilai ukuran diantaranya[28]:

1. Mean Absolute Error (MAE)

Mean Absolute Error merupakan ukuran yang digunakan untuk mengukur ketepatan hasil ramalan dengan merata-ratakan kesalahan hasil peramalan. Metode ini bermanfaat ketika digunakan saat mengukur kesalahan hasil peramalan dalam unit yang sama. Nilai MAE dapat dihitung dengan menggunakan persamaan (9) berikut :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

Dimana :

n = jumlah data

y_i = nilai aktual

\hat{y}_i = nilai prediksi

2. Mean Square Error (MSE)

Mean Square Error merupakan ukuran lain yang digunakan untuk mengukur ketepatan hasil ramalan dengan merata-ratakan kesalahan hasil peramalan yang dikuadratkan. Nilai MSE yang rendah atau mendekati nol menunjukkan bahwa hasil ramalan konsisten dengan data aktual dan dapat digunakan untuk menghitung ramalan di periode waktu mendatang. Nilai MSE dapat dihitung dengan menggunakan persamaan(10) berikut :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

Dimana :

n = jumlah data

y_i = nilai aktual

\hat{y}_i = nilai prediksi

3. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error merupakan ukuran yang didapatkan dari nilai kesalahan absolut pada tiap periode pengamatan dibagi dengan banyaknya observasi pada periode tersebut, lalu merata-ratakan persentase kesalahan absolut tersebut. Pada metode ini, nilai yang semakin kecil juga akan memberikan arti bahwa ramalan yang dilakukan memberikan hasil yang baik. Nilai MAPE dapat dihitung dengan menggunakan persamaan (11) berikut :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| 100 \quad (11)$$

Dimana :

n = jumlah data

y_i = nilai aktual

\hat{y}_i = nilai prediksi

MAPE merupakan alat ukur yang lebih mudah untuk diinterpretasikan dibandingkan dengan alat ukur yang lainnya. Dalam memahami hasil akhir dari nilai mape dapat dilihat pada tabel 2.1 dibawah ini[29].

Tabel 2.1 Tabel Interpretasi hasil MAPE

Range	Keterangan
< 10 %	Model peramalan yang dihasilkan sangat baik
10 - 20 %	Model peramalan yang dihasilkan baik
20 - 50 %	Model peramalan yang dihasilkan cukup/layak
> 50 %	Model peramalan yang dihasilkan buruk

4. *Root Mean Square Error (RMSE)*

Root Mean Square Error merupakan alternatif lain ukuran yang dapat digunakan dalam melihat akurasi dari model peramalan yang digunakan. Metode ini didapatkan dengan menarik akar kuadrat dari nilai MSE. Metode ini menghasilkan skala yang lebih intuitif dibandingkan MSE sehingga dapat menjadi alternatif yang lebih baik. Nilai RMSE dapat dihitung dengan menggunakan persamaan (12) berikut :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

Dimana :

n = jumlah data

y_i = nilai aktual

\hat{y}_i = nilai prediksi

2.7 Penelitian Terdahulu

Adapun dalam pengerjaan skripsi ini, memiliki beberapa studi literatur yang diambil sebagai referensi dalam pengerjaannya. Beberapa studi literatur yang diambil dapat dilihat pada tabel 2.1.

Tabel 2.2 Studi Literatur terdahulu

No	Nama Peneliti	Model	Kontribusi	Area Aplikasi	Hasil
1	Fara Inka Durrah , Yulia, Tessa Prihartina Parhusip , Asep Rusyana [11]	SARIMA	Melakukan peramalan jumlah penumpang pesawat di Bandara Iskandar Muda dengan periode 12 bulan pada tahun 2017	Penerbangan	Penelitian menghasilkan model SARIMA dengan nilai MAPE 9,7%
2	Rifqi Fahrudin, Irfan Dwiguna Sumitra [12]	SARIMA dan SES	Melakukan peramalan inflasi di kota Bandung dengan periode Januari 2019-Juni 2019	Ekonomi	SARIMA memiliki nilai MAPE 0,72% dan SES memiliki nilai MAPE 2,86%
3	Regita Aprilia Maharani Yusuf, Teti Sofia Yanti [13]	SARIMA dan Fuzzy Time Series	Melakukan peramalan jumlah wisatawan di Bali pada periode Januari sampai dengan Maret 2020	Pariwisata	SARIMA memiliki nilai MAPE 4,64% dan Fuzzy Time Series memiliki nilai MAPE 5,96%
4	Ni Putu Nanik Hendayanti, Maulida Nurhidayati [14]	SARIMA dan SVR	Melakukan peramalan pada wisatawan mancanegara di Bali pada periode Januari sampai dengan Desember 2019.	Pariwisata	SARIMA memiliki nilai MAPE 5,33% dan SVR memiliki nilai MAPE 19,74%
5	Soumik Ray, Soumitra Sankar Das, Pradeep Mishra,Abdullah	SARIMA	Melakukan peramalan pada tingkat hujan serta temperatur di negara di Asia Selatan pada periode Januari	Agrikultur	Penelitian menghasilkan model SARIMA dengan rentang nilai MAE terendah 1,31 sampai

Mohammad Ghazi Al Khatib [30]		2011 sampai dengan Desember 2021		dengan tertinggi 4.3396
-------------------------------------	--	-------------------------------------	--	----------------------------



UNIVERSITAS MIKROSKIL