BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian oleh Pattanayak dkk. (2020) memperkenalkan metode peramalan deret waktu berbasis fuzzy yang menggabungkan nilai keanggotaan dengan data aktual dan menggunakan Support Vector Machine (SVM) untuk model peramalan. Hasil dari penelitian ini menunjukkan bahwa pendekatan yang diusulkan memberikan performa peramalan yang lebih baik dibandingkan dengan metode fuzzy time series konvensional. Metode ini memanfaatkan nilai keanggotaan fuzzy yang diperoleh dari data aktual, yang kemudian digunakan bersama dengan data itu sendiri sebagai input untuk model SVM. Pendekatan ini memungkinkan model untuk menangkap pola kompleks dalam data yang mungkin tidak terlihat jika hanya menggunakan data atau nilai keanggotaan saja. Secara keseluruhan, penelitian ini menunjukkan bahwa penggunaan nilai keanggotaan fuzzy bersama dengan SVM dapat meningkatkan akurasi peramalan untuk berbagai jenis data seri waktu.

Penelitian oleh Pattanayak dkk. (2023) menghasilkan bahwa metode *Hesitant Fuzzy Time Series Forecasting-Support Vector Machine* (HFTSF-SVM) dibandingkan metode lain dengan nilai *mean rank* terkecil yaitu 100,56 di *Symmetric Mean Absolute Percentage Error* (SMAPE) dan 129,69 di *Mean Absolute Scaled Error* (MASE) dengan jarak kritis 3,8. Dalam penelitian tersebut metode peramalan *Hesitant Fuzzy Time Series* dengan menggunakan *mean aggregated membership value* dalam menganggregasi *Hesitant Fuzzy Element* (HFE) dan *Support Vector Machine* (SVM) digunakan memodelkan data deret waktu. Dalam penelitian tersebut peneliti juga menggunakan metode *Length Based Discretization* (LBD) untuk menentukan banyak interval atau *Number of Interval* (NOI) bersama dengan

dua semesta pembicaraannya atau *Universe Of Discourse* (UOD) untuk deret waktu. UOD dalam penelitian tersebut digunakan untuk menghitung nilai nonkeanggotaan dan keanggotaan dari data observasi yang diagregasi untuk memperoleh HFE yang teragregasi. Metode LBD membantu secara otomatis dalam menentukan banyak dan panjang interval deret waktu sehingga dapat diaplikasikan untuk berbagai jenis data deret waktu.

Penelitian oleh Saini dkk. (2024) menunjukkan bahwa penggunaan model *Artificial Neural Network* (ANN) yaitu LSTM (*Long Short-Term Memory*) dalam memprediksi parameter polusi udara di kota Amravati, India baik di wilayah pemukiman, wilayah industri dan wilayah komersial. Dalam penelitian tersebut, dengan menggunakan data masa lampau selama 10 tahun mulai dari tahun 2008 sampai 2018, parameter polusi untuk tahun 2019 dihitung dengan menggunakan MATLAB. Hasil yang menunjukkan tingkat akurasi yang baik terhadap data aktual. Kualitas udara ambien kota Amravati dinilai dengan menggunakan tiga kriteria utama yaitu SO_2, NO_X dan RSPM (*Respirable Suspended Particulate Matter*). Akurasi prediksi untuk tahun 2019 di stasiun observasi GCE (*Government College of Engineering*) untuk SO_2, NO_X dan RSPM secara berurutan menghasilkan nilai MAPE (*Mean Absolute Percentage Error*) 1,229; 0,572 dan 6,006 serta untuk nilai MAE (*Mean Absolute Error*) yang dihasilkan adalah 0,012; 0,005 dan 0,061 serta nilai MBE (*Mean Bias Error*) yang dihasilkan adalah -0,147; 0,074 dan 4,204. Kemudian akurasi prediksi tahun 2019 di stasiun observasi MIDC (*Maharashtra Industrial Development Corporation*) untuk SO_2, NO_X dan RSPM secara berurutan menghasilkan nilai MAPE (*Mean Absolute Percentage Error*) 30,701; 19,907 dan 6,7149 serta untuk nilai MAE (*Mean Absolute Error*) yang dihasilkan adalah 0,307; 0,199 dan 0,067 serta nilai MBE (*Mean Bias Error*) yang dihasilkan adalah 4,013; 2,859 dan 6,379. Kemudian akutasi prediksi tahun 2019 di stasiun observasi RKC (*Rajkamal Chowk*) untuk SO_2, NO_X dan RSPM secara berurutan menghasilkan nilai

MAPE (*Mean Absolute Percentage Error*) 5,7935; 12,295 dan 17,798 serta untuk nilai MAE (*Mean Absolute Error*) yang dihasilkan adalah 0,057; 0,123 dan 0,178 serta nilai MBE (*Mean Bias Error*) yang dihasilkan adalah -0,811; 1,871 dan 18,154. Penelitian ini juga menunjukkan adanya korelasi antara polutan dengan AQI (*Air Quality Index*) atau indeks kualitas udara dengan besaran dan sifat yang berbeda di semua stasiun pemantauan di wilayah.

Penelitian oleh Marinov dkk. (2022) berfokus dalam memprediksi kualitas udara di Sofia, Bulgaria dengan menggunakan deret waktu dengan menggunakan model *Auto Regressive Integrated Moving Average* (ARIMA) pada parameter polutan $PM_{2.5}$, ozon (O_3), nitrogendioksida (NO_2) dan karbonmonoksida (CO) dengan interval waktu yang berbeda-beda yaitu dalam interval 1 jam, 3 jam, 6 jam, 12 jam dan 24 jam. Model ARIMA memberikan prediksi jangka pendek yang akurat, dengan MAPE antara 1.52 hingga 5.12 untuk granularitas 1 jam dan 2.53 hingga 14.12 untuk granularitas 3 jam. Dalam penelitian tersebut data dikumpulkan dari lima stasiun pemantauan kualitas udara antara tahun 2015 sampai 2019. Teknik analisis data lebih lanjut dalam mengatasi nilai-nilai hilang dalam dataset menggunakan metode rata-rata berdasarkan waktu serta dalam pemilihan parameter model ARIMA dipilih berdasarkan plot autokorelasi dan autokorelasi parsial serta menggunakan pencarian grid dengan pustaka yang ada di perangkat lunak python untuk memperoleh parameter yang optimal.

Penelitian oleh Hasnain dkk. (2022) memprediksi tingkat polusi udara dengan fokus pada parameter polutan $PM_{2.5}$, PM_{10}, ozon (O_3), nitrogendioksida (NO_2), sulfurdioksida (SO_2) dan karbonmonoksida (CO). Dalam penelitian tersebut model yang dikembangkan adalah model yang berbasis *Prophet Forecasting Method* (PFM) dengan menggunakan data yang diperoleh dari 72 stasiun pemantauan kualitas udara dari tahun 2016 sampai 2020 serta dataset publik dari *China National Environmental Monitoring*

Centre. Untuk mengevaluasi kemampuan model metrik statistik yang digunakan dalam penelitian tersebut menggunakan koefisien korelasi (R^2) , *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Hasil penelitian menunjukkan PFM menghasilkan tingkat akurasi lebih baik daripada model tradisional ARIMA dan SARIMA terutama dalam memprediksi jangka panjang. Temuan utama dalam penelitian ini adalah model PFM efektif dalam memprediksi $PM_{2.5}$ dan PM_{10} dengan nilai (R^2) berturut-turut 0,4 dan 0,52 serta nilai RMSE 12,07 dan MAE 8,22 dalam memprediksi $PM_{2.5}$.

Penelitian oleh Reikard (2019) membahas dampak emisi vulkanik dari Gunung Kilauea di Hawai'i terhadap kualitas udara dan kesehatan masyarakat. Penelitian ini mengkaji data emisi vulkanik, gas sulfur dioksida SO_2 dan partikel halus $PM_{2.5}$ untuk mengembangkan model peramalan menggunakan metode *time series* menggunakan model ARIMA (*Auto Regressive Integrated Moving Average*) dan regresi untuk meramalkan konsentrasi sulfur dioksida (SO_2) dan $PM_{2.5}$. Data emisi vulkanik dan kualitas udara dikumpulkan dari berbagai sumber, termasuk pengukuran harian dan jam dari sulfur dioksida (SO_2) dan $PM_{2.5}$. Penelitian ini juga mengevaluasi data baru yang dikumpulkan oleh *U.S. Geological Survey* (USGS) dan Departemen Kesehatan Hawai'i. Peramalan untuk sulfurdioksida (SO_2) dan $PM_{2.5}$ dilakukan dalam jangka waktu 1-24 jam dan 1-7 hari. Hasil menunjukkan bahwa model ARIMA (*Auto Regressive Integrated Moving Average*) memberikan hasil yang lebih baik dalam memprediksi SO_2, terutama dalam jangka pendek. Namun, hasil untuk $PM_{2.5}$ lebih ambigus, dan seringkali diprediksi lebih baik dengan regresi pada tingkat atau model persistensi sederhana. Peramalan emisi vulkanik tidak meningkatkan akurasi peramalan dan model ARIMA memberikan hasil yang lebih baik untuk memprediksi sulfur dioksida (SO_2) dalam jangka pendek.

Namun, kesalahan peramalan masih signifikan, terutama karena data emisi yang volatile dan non-stasioner.

Tabel 2.1 Penelitian terdahulu

No.	Peneliti	Hasil Penelitian	*Research Gap*
1.	Pattanayak dkk. (2020)	Menggabungkan nilai keanggotaan fuzzy dan data aktual sebagai input SVM. Hasilnya lebih baik dari FTS konvensional dalam memprediksi berbagai macam data.	Belum menggunakan teknik tertentu dalam fuzzifikasi seperti HFS dan model yang dikembangkan belum terfokus pada masalah polusi udara.
2.	Pattanayak dkk. (2023)	Memperkenalkan metode HFTS-SVM dengan LBD dan agregasi HFE. Menghasilkan Tingkat akurasi lebih baik daripada model prediki lain di Sebagian besar data.	Model yang fuzzy diintegrasikan dengan SVM, belum dicoba dengan algoritma deep learning dan belum terfokus pada masalah polusi udara.
3.	Saini dkk. (2024)	LSTM digunakan untuk memprediksi polusi udara (2008–2019). Akurat di berbagai wilayah (residensial, industri dan komersial).	Tidak melibatkan konsep fuzzy atau hesitant fuzzy. Tidak mengkaji efek perbedaan interval data. Fokus hanya pada LSTM dan data jangka panjang.

No.	Peneliti	Hasil Penelitian	*Research Gap*
4.	Marinov dkk. (2022)	ARIMA digunakan untuk memprediksi kualitas udara dengan interval berbeda (1–24 jam). MAPE cukup rendah untuk prediksi jangka pendek.	ARIMA kurang cocok untuk data non-linear dan volatil seperti SO_2. Tidak menggunakan pendekatan fuzzy atau machine learning.
5.	Hasnain dkk. (2022)	Prophet digunakan untuk prediksi polusi dari 72 stasiun di China. Lebih baik dari ARIMA/ SARIMA untuk prediksi jangka panjang.	Prophet tidak efektif untuk data sangat dinamis dan tidak stasioner seperti SO_2. Tidak menggunakan fuzzy atau pendekatan hybrid.
6.	Reikard, (2019)	ARIMA dan regresi digunakan untuk memprediksi emisi vulkanik SO_2 dan $PM_{2.5}$. ARIMA baik untuk SO_2 jangka pendek.	Performa rendah pada $PM_{2.5}$ dan data yang sangat volatile. Tidak gunakan fuzzy atau ML.

2.2 Dasar Teori

2.2.1 Polutan Udara

Menurut Persson dkk. (2022) polutan adalah zat atau energi yang masuk ke dalam lingkungan yang menimbulkan dampak yang tidak diinginkan atau berdampak buruk terhadap kegunaan suatu sumber daya. Polutan udara dapat diartikan sebagai zat atau bahan di udara yang dapat menimbulkan dampak terhadap ekosistem dan terhadap makhluk hidup termasuk manusia.

Menurut Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia (2020) dalam mengukur atau menggambarkan kondisi mutu udara ambien di lokasi tertentu yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya terdiri atas 7 parameter yaitu materi partikulat $PM_{2.5}$ dan PM_{10}, karbon monoksida (CO), nitrogen dioksida (NO_2), sulfur dioksida (SO_2), ozon (O_3) dan hidrokarbon (HC). Menurut Chan (2002) polutan utama yang berdampak terhadap kesehatan manusia diantaranya adalah nitrogen oksida, sulfur, materi partikulat (PM) tersuspensi yang dapat dihirup, materi partikulat (PM) dalam bentuk tersuspensi. Menurut petunjuk kualitas udara global oleh World Health Organization (2021) polutan utama antara lain adalah $PM_{2.5}$, PM_{10}, ozon (O_3), nitrogen dioksida (NO_2), sulfur dioksida (SO_2) dan karbon monoksida (CO). Berikut merupakan penjelasan polutan menurut World Health Organization (2021)

a) Materi Partikulat/ *Particulate Matter* (PM) merupakan campuran partikel-partikel kecil yang terdiri dari bahan kimia baik yang berasal dari kendaraan bermotor, industri, pembakaran biomassa dan debu alam yang meliputi sulfat, nitrat, amonia, natrium klorida, karbon hitam, debu mineral dan air yang dapat terhirup oleh manusia. PM diklasifikasikan berdasarkan ukurannya yaitu PM_{10} (PM dengan diameter kurang dari 10 mikrometer) dan $PM_{2.5}$ (PM dengan diameter kurang dari 2.5 mikrometer) sebagai yang paling umum dipantau karena ukurannya yang kecil dan kemampuannya untuk menembus ke dalam saluran pernafasan manusia yang dapat membawa dampak buruk terhadap kesehatan manusia.
b) Ozon (O_3) yang berada di permukaan bumi merupakan polutan udara yang berbahaya karena sifatnya yang sangat reaktif. Ozon terbentuk dari reaksi foto-kimia dengan polutan seperti senyawa organik yang mudah menguap, karbon monoksida, dan

nitrogenoksida (NO_X) yang dipancarkan dari kendaraan dan industri. Karena sifat foto-kimianya, tingkat ozon tertinggi terjadi selama cuaca cerah. Paparan ozon berlebih dapat menyebabkan masalah dalam bernafas, memicu asma, menurunkan fungsi paru-paru, dan menyebabkan penyakit paru-paru.

c) Nitrogen dioksida (NO_2) adalah gas berwarna cokelat kemerahan yang larut dalam air dan merupakan oksidan yang kuat. Nitrogen oksida merupakan gas beracun yang dihasilkan oleh pembakaran bahan bakar fosil pada suhu tinggi dalam proses seperti pemanasan, transportasi, industri, dan pembangkit listrik. Nitrogen dioksida merupakan salah satu polutan udara utama dan paparan nitrogen dioksida dapat menyebabkan iritasi saluran udara dan memperburuk penyakit pernafasan.

d) Sulfur dioksida (SO_2) merupakan gas tidak berwarna, mudah larut dalam air dan gas beracun yang dihasilkan dari pembakaran bahan bakar fosil yang mengandung sulfur terutama batu bara dan minyak bumi. Sulfur dioksida dapat menyebabkan iritasi pada saluran pernafasan, dan dalam jumlah tinggi dapat menyebabkan masalah kesehatan serius termasuk gangguan pernafasan, peningkatan risiko asma, dan penyakit kardiovaskular.

e) Karbon monoksida (CO) adalah gas tak berwarna dan tidak berbau yang dihasilkan dari pembakaran tidak sempurna bahan organik, seperti bahan bakar fosil dan kayu bakar. CO sangat berbahaya karena dapat mengikat erat dengan hemoglobin dalam darah, mengurangi kemampuan darah untuk membawa oksigen, yang dapat menyebabkan keracunan CO yang serius, bahkan kematian, jika terpapar dalam jumlah yang cukup besar.

2.2.2 Statistik Deskriptif

Menurut Sugiyono (2019) statistik deskriptif adalah statistik yang berfungsi untuk mendeskripsikan atau memberi gambaran terhadap obyek

yang diteliti melalui data sampel atau populasi sebagaimana adanya, tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum. Statistik deskriptif bertujuan untuk memberikan gambaran karakteristik dasar dari sebuah dataset serta memberikan wawasan yang berkaitan dengan distribusi data sehingga memungkinkan untuk memahami pola-pola yang mendasarinya. Secara umum, konsep dasar dalam statistik deskriptif berkaitan dengan ukuran pemusatan data serta berkaitan dengan sebaran data.

Dalam penelitian ini, statistik deskriptif yang akan digunakan untuk memberikan gambaran dataset adalah rata-rata, *maximum, minimum, range* dan standar deviasi. *Mean* atau rata-rata hitung merupakan suatu ukuran pemusatan data yang didasarkan atas jumlah semua nilai data dibagi dengan banyak data. Secara matematis nilai rata-rata dapat dinyatakan sebagai

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

dengan n menyatakan banyak data dan x_i menyatakan nilai data ke-*i*. *Maximum* (max) dan *minimum* (min) secara berturut-turut adalah nilai terbesar dan terkecil dalam dataset yang nilai selisihnya dapat dikatakan sebagai *range*. Secara matematis dapat dinyatakan sebagai

$$range = \max - \min \tag{2.2}$$

Standar deviasi merupakan ukuran statistik yang mengukur sebaran data dari nilai rata-rata sehingga memberikan informasi tentang variasi atau sebaran data dalam dataset. Secara matematis standar deviasi dapat dinyatakan sebagai

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.3}$$

dengan SD adalah nilai standar deviasi, n menyatakan banyak data x_i menyatakan nilai data ke-*i* dan $\bar{x}$ merupakan rata-rata hitung yang dapat ditentukan dengan menggunakan persamaan (2.1). Semakin besar nilai standar deviasi yang dihasilkan, semakin besar variasi antara titik data dengan

nilai rata-ratanya.

2.2.3 *Time Series*

Menurut Brockwell dan Davis (1996) *time series* (deret waktu) adalah kumpulan pengamatan x_t yang masing-masing dicatat berdasarkan selang waktu tertentu *t*. Sedangkan menurut Pattanayak dkk. (2023), *time series* (deret waktu) adalah suatu kumpulan data berurutan berdasarkan interval waktu yang sama dan menggambarkan suatu fenomena.

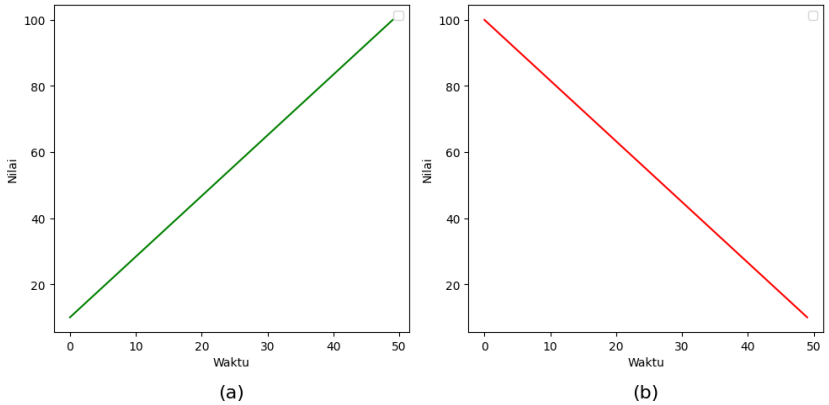
Menurut Levin dan Black dalam (Pasaribu dkk., 2021), misal terdapat suatu *time series* Y maka dalam *time series* umumnya dapat diuraikan secara matematis terdapat empat komponen dalam deret waktu yang menyebabkan gerakan atau variasi, yaitu *Trend* (T), *Cyclical* (C), *Seasonal* (S) dan *Irregular* (I). Secara matematis dapat dinyatakan sebagai hasil kali keempat komponen tersebut sebagaimana yang ditunjukkan dalam persamaan (2.4) selain itu beberapa juga menganggap bahwa *time series* merupakan hasil penjumlahan komponen tersebut sebagaimana yang dinyatakan dalam persamaan (2.5)

$$Y = T \times C \times S \times I \tag{2.4}$$

$$Y = T + C + S + I \tag{2.5}$$

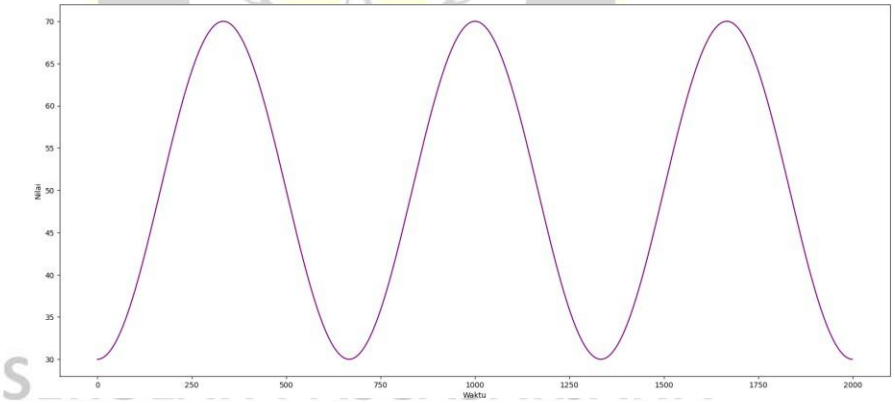
berikut merupakan penjelasan komponen-komponen dalam deret waktu menurut Pasaribu dkk., (2021)

a) *Trend*/ tren adalah suatu gerakan yang memiliki kecenderungan yang menunjukkan suatu perkembangan yang dilihat dari suatu arah yang menunjukkan kenaikan atau penurunan. Jika arah tren naik maka akan menunjukkan tren positif dan jika tren menurun maka menunjukkan tren negatif. Gambar 2.*1* sebagai berikut merupakan bentuk pola tren naik (a) dan bentuk pola tren turun (b).



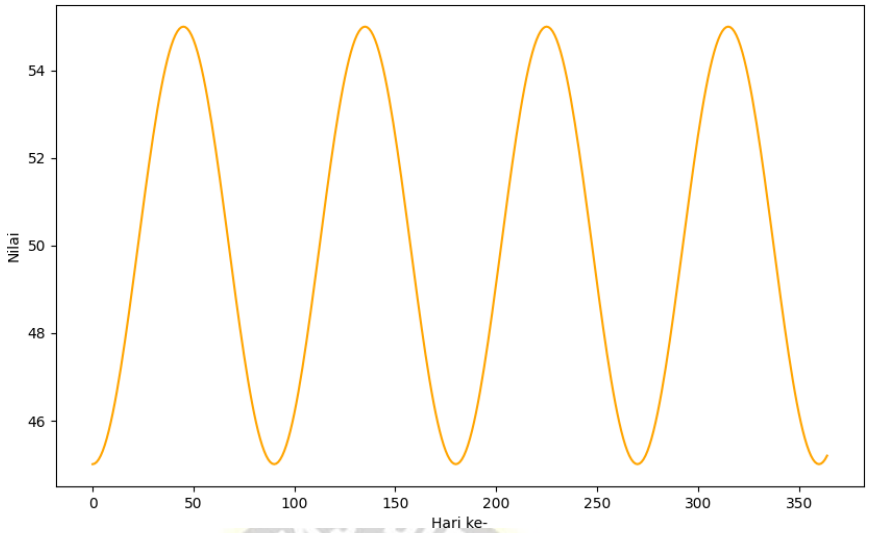
Gambar 2.1 Pola data tren naik (a) dan pola data tren turun (b)

b) *Cyclical*/ Siklis merupakan suatu variasi yang dipengaruhi oleh fluktuasi ekonomi jangka panjang dan terjadi berulang setelah jangka waktu tertentu sehingga membentuk suatu siklus yang memiliki gerakan naik atau turun secara periodik pada jangka waktu yang panjang. Gambar 2.*2* berikut merupakan pola siklus dalam jangka panjang.



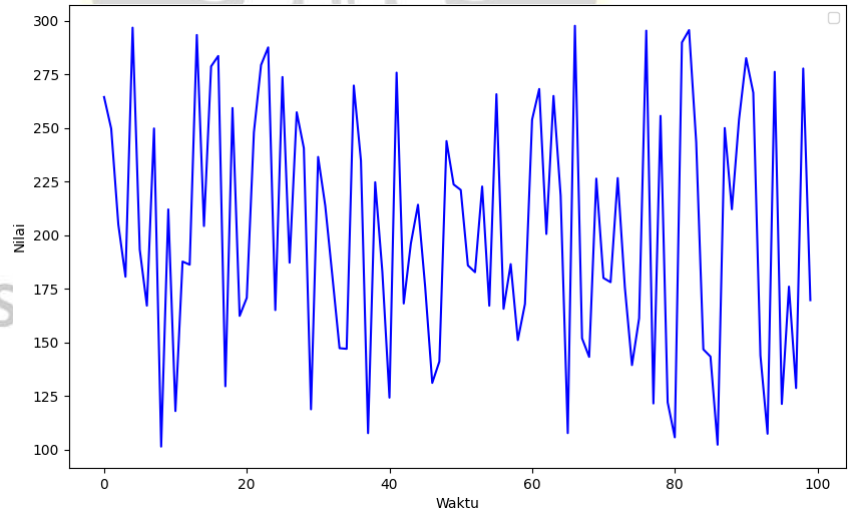
Gambar 2.2 Pola data siklis jangka panjang

c) *Seasonal*/ musiman adalah suatu pola gerakan yang menunjukkan perkembangan fluktuasi yang teratur meningkat atau turun pada waktu musiman atau bulan tertentu dalam jangka waktu kurang dari setahun. Gambar 2.*3* berikut merupakan pola data musiman dalam satu tahun.



Gambar 2.3 Pola data musiman dalam satu tahun

d) *Irregular*/ acak merupakan pola gerakan yang tidak teratur dan sulit untuk diprediksi. Pola *irregular* bersifat sporadis dan bisa disebabkan oleh faktor yang kebetulan namun sangat tidak merata dan tidak tentu. Gambar 2.*4* berikut merupakan contoh grafik pola data acak.



Gambar 2.4 Pola data acak

2.2.4 Hesitant Fuzzy Time Series (HFTS)

Fuzzy Time Series merupakan salah satu analisis deret waktu yang melibatkan prinsip-prinsip fuzzy. Beberapa konsep dasar yang berkaitan dengan FTS diantaranya adalah sebagai berikut

Definisi 2.1 (Song & Chissom, 1993) Diberikan $Y(t), \{\forall t \mid 1,2,3,...,n\}$ merupakan himpunan bagian bilangan riil yang merupakan semesta pembicaraan dengan himpunan fuzzy $f_i(t)$ terdefinisi. Jika $F(t)$ adalah kumpulan dari $f_1(t), f_2(t).f_3(t),...$ maka dapat dikatakan bahwa $F(t)$ merupakan FTS yang terdefinisi pada $Y(t)$.

Definisi 2.2 (Song & Chissom, 1993) Diberikan U merupakan semesta pembicaraan dengan elemen-elemen $\{u_1, u_2, u_3, ..., u_n\}$. Suatu himpunan fuzzy S pada semesta pembicaraan dapat dinyatakan sebagai

$$S = \frac{f_s(u_1)}{u_1} + \frac{f_s(u_2)}{u_2} + \frac{f_s(u_3)}{u_3} + ... + \frac{f_s(u_n)}{u_n} \tag{2.6}$$

Dengan S adalah himpunan fuzzy sedemikian sehingga $f_{S_i} : U \to [0,1]$ dan $f_{S_i}(u_i) \in [0,1]$. f_{S_i} menyatakan nilai derajat keanggotaan u_i di himpunan fuzzy S dan operator "+" menyatakan operasi gabungan.

Definisi 2.3 (Song & Chissom, 1993) Misalkan, nilai $F(t)$ diprediksi dari $F(t-1)$, kemudian relasi FTS dapat dinyatakan sebagai $F(t-1) \to F(t)$ dan model ini disebut dengan model FTS orde pertama. Jika $F(t-1) = S_i$ dan $F(t) = S_j$ maka relasi kedua observasi $F(t-1)$ dan $F(t)$ dapat dinyatakan dalam suatu *Fuzzy Logic Relationship* (FLR) yang dinyatakan sebagai $S_i \to S_j$.

Konsep *high-order fuzzy time series* merupakan pengembangan konsep FTS dengan melibatkan dua atau lebih data historis pada tahap menentukan FLR (Chen, 2002). Konsep *high-order fuzzy time series* dapat didefinisikan sebagai berikut (Li dkk., 2022).

Definisi 2.4 (Bas dkk., 2018) Misalkan F sebuah FTS, jika $F(t)$ diakibatkan oleh $F(t-1), F(t-2), ..., F(t-2),$ maka relasi yang terbentuk dapat dinyatakan sebagai

$$F(t-h),...,F(t-2),F(t-1)\rightarrow F(t) \tag{2.7}$$

relasi yang terbentuk pada persamaan (2.7) disebut dengan FTS orde ke-*h*.

Definisi 2.5 (Xia & Xu, 2011) Suatu himpunan fuzzy H dikatakan sebagai *Hesitant Fuzzy Set* (HFS) pada semesta himpunan $U=\{u_1,u_2,u_3,...,u_n\}$ adalah fungsi apabila jika diterapkan ke U maka menghasilkan suatu subset $[0,1]$. Secara matematis dapat dinyatakan sebagai

$$H=\left\{\left\langle u,h_H(u)\right\rangle \mid u\in U\right\} \tag{2.8}$$

dengan $h_H(u)$ merupakan himpunan beberapa nilai pada $[0,1]$ yang menyatakan kemungkinan nilai derajat keanggotaan dari elemen $u\in U$ ke himpunan H dan untuk setiap kemungkinan derajat keanggotaan disebut dengan *Hesitant Fuzzy Element* (HFE).

Sebagai contoh jika terdapat suatu himpunan $U=\{u_1,u_2,u_3\}$ dengan kemungkinan nilai derajat keanggotaan untuk $u_i(i=1,2,3)$ adalah $h_H(u_1)=\{0,4;0,5;0,3\}$; $h_H(u_2)=\{0,8;0,9;0,7\}$ dan $h_H(u_3)=\{0,2;0,1;0,3\}$ maka HFS yang terbentuk dari HFE tersebut adalah $H=\left\{\left\langle u_1,\{0,4;0,5;0,3\}\right\rangle,\left\langle u_2,\{0,8;0,9;0,7\}\right\rangle,\left\langle u_3,\{0,2;0,1;0,3\}\right\rangle\right\}$.

Dalam menentukan nilai derajat keanggotaan HFE, teknik yang akan digunakan adalah dengan menggunakan fungsi keanggotaan triangular. Fungsi keanggotaan triangular hanya memerlukan tiga parameter, yaitu batas bawah, nilai tengah dan batas atas sehingga memiliki kesederhanaan dalam perhitungan (Ross, 2010). Sedangkan menurut Pedrycz dan Gomide (1998), fungsi keanggotaan triangular memberikan keseimbangan yang baik antara presisi dan kesederhanaan sehingga cukup akurat untuk menggambarkan ketidakpastian dalam data. Secara umum fungsi keanggotaan triangular untuk suatu himpunan fuzzy *A* dapat didefinisikan sebagai berikut

$$\mu_A(y) = \begin{cases} \dfrac{y - x_1}{x_2 - x_1} & \text{jika } x_1 \le y \le x_2 \\ \dfrac{x_3 - y}{x_3 - x_2} & \text{jika } x_2 \le y \le x_3 \\ 0 & \text{lainnya} \end{cases} \tag{2.9}$$

dengan $\mu_A(y)$ adalah nilai derajat keanggotaan elemen y pada himpunan fuzzy A dan x_1, x_2 dan x_3 secara berturut-turut adalah nilai batas bawah, nilai tengah dan nilai batas atas.

2.2.5 Mean Aggregated Membership Value

Untuk menentukan HFS, HFE yang diperoleh untuk setiap y_i dalam himpunan Y diagregasi dengan menggunakan dengan menggunakan operator agregasi (Bisht & Kumar, 2016).

Diasumsikan bahwa H suatu HFS yang merupakan *Hesitant Fuzzy Element* (HFE) dari himpunan H dihitung dengan memetakan fungsi $h_H : U \to P[0,1]$ maka himpunan *hesitant fuzzy* teragregasi $H_A = \{\langle u, \Phi(h_H(u))\rangle\}, \forall u \in U$ suatu himpunan fuzzy dan nilai keanggotaan teragregasi ke himpunan ini dapat ditentukan dengan menggunakan persamaan berikut

$$\Phi : P[0,1] \to \text{sdh}, \Phi(\{u_1, u_2, ..., u_v\}) = 1 - \prod_{i=1}^{v} (1 - u_i)^{w_i} \tag{2.10}$$

sdh: $\min(\{u_1, u_2, ..., u_v\}) \le \Phi(\{u_1, u_2, ..., u_v\}) \le \max(\{u_1, u_2, ..., u_v\}) \mid \forall u_i \in [0,1]$

dengan i=1, 2, 3..., v dimana v merepresentasikan banyak subset yang terbentuk untuk setiap observasi untuk menghitung derajat keanggotaan pada rentang [0,1] untuk setiap interval himpunan fuzzy dan untuk setiap u_i dan bobot w_i dengan $\forall i = (1, 2, 3, ..., v)$ sedemikian sehingga $\sum_{i=1}^{v} w_i = 1$.

Sebagai contoh misalkan $U = \{u_1, u_2, u_3\}$ merupakan himpunan semesta dan $H = \{\langle u_1, \{0,4; 0,5; 0,3\}\rangle, \langle u_2, \{0,8; 0,9; 0,7\}\rangle, \langle u_3, \{0,2; 0,1; 0,3\}\rangle\}$ adalah

suatu HFS pada U. Pada contoh ini diketahui bahwa v=3, sehingga nilai bobot untuk setiap subset dapat ditentukan sebagai $w_1 = \frac{1}{3}$, $w_2 = \frac{1}{3}$ dan $w_3 = \frac{1}{3}$. Maka HFE yang teragregasi ke HFS dapat dihitung sebagai berikut

$$h(u_1) = 1 - \{(1-0,4)^{\frac{1}{3}}(1-0,5)^{\frac{1}{3}}(1-0,3)^{\frac{1}{3}}\} = 0,4056$$

dengan $0,4056 \geq \min(\{0,4;0,5;0,3\})$ dan $0,4056 \leq \max(\{0,4;0,5;0,3\})$

$$h(u_2) = 1 - \{(1-0,8)^{\frac{1}{3}}(1-0,9)^{\frac{1}{3}}(1-0,7)^{\frac{1}{3}}\} = 0,8183$$

dengan $0,8183 \geq \min(\{0,8;0,9;0,7\})$ dan $0,8183 \leq \max(\{0,8;0,9;0,7\})$

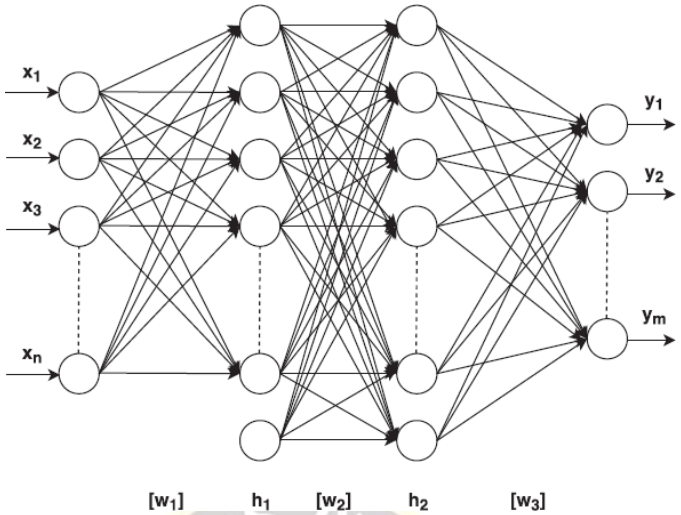
$$h(u_3) = 1 - \{(1-0,2)^{\frac{1}{3}}(1-0,1)^{\frac{1}{3}}(1-0,3)^{\frac{1}{3}}\} = 0,2042$$

dengan $0,2042 \geq \min(\{0,2;0,1;0,3\})$ dan $0,2042 \leq \max(\{0,2;0,1;0,3\})$

2.2.6 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) merupakan salah satu algoritma *supervised learning* yang mempelajari fungsi nonlinear dan memetakan input ke output dengan pelatihan yang ada pada dataset (Feng dkk., 2020).

Diberikan suatu himpunan input $X = x_1, x_2, x_3, ..., x_n$ dan output $Y = y_1, y_2, y_3, ..., y_n$ dengan n merupakan banyak input dan m merupakan banyak output. MLP mempelajari aproksimator nonlinear $f(.): X \to Y$ baik untuk masalah klasifikasi maupun masalah regresi. MLP terdiri atas tiga atau lebih lapisan (*layer*) yaitu sebuah *input layer, output layer* dan satu atau lebih *hidden layer*. Tiap titik di layer terhubung dengan layer selanjutnya dengan bobot tertentu. Secara umum struktur MLP dapat digambarkan sebagai berikut



Gambar 2.5 Struktur dasar MLP
(Chatterjee dkk., 2022)

Berdasarkan Gambar 2.5 dapat diketahui bahwa *input layer* terdiri atas kumpulan neuron dengan X merepresentasikan *input* sementara layer yang terletak di paling kanan merupakan *output* layer layer yang memperoleh informasi dari *hidden layer* sebelumnya yang ditransformasikan menjadi nilai *output*. Setiap neuron *hidden layer* mengakumulasikaan nilai dari *layer* sebelumnya sebagai jumlah linear bobot dengan bias, diikuti dengan fungsi aktivasi nonlinear. Persamaan dari titik di *hidden layer* dapat dinyatakan sebagai berikut

$$h_{1_j} = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right) \tag{2.11}$$

Dengan mengacu pada Gambar 2.5, h_{1_j} didefinisikan sebagai titik ke-*j* dari hidden layer h_1, w_{ij} merepresentasikan bobot yang berkaitan dengan input ke- x_i yang memasuki titik ke-*j* dari hidden layer h_1 dan b_j merupakan bias.

Model prediksi MLP juga menggunakan optimizer yang digunakan untuk meminimalkan *loss function* yang mengukur seberapa baik model dalam memprediksi serta berkaitan dengan kecepatan dalam proses pelatihan model. Optimizer merupakan suatu algoritma untuk *meminimalkan loss*

function untuk meningkatkan akurasi yang berbentuk fungsi matematika yang bergantung pada parameter model yang dapat dipelajari yaitu bobot dan bias (Sarno dkk., 2023). Dalam penelitian ini, teknik optimasi yang digunakan adalah adam yang merupakan gabungan dari teknik *Root Mean Square Prop* (RMSProp) dengan menggabungkan momen pertama (gradien rata-rata) dan *Adaptive Gradient Descent* (Adagrad) dengan momen kedua (variasi gradien) untuk menyesuaikan *learning rate* secara adaptif (Saputra & Kristiyanti, 2022). Teknik optimasi adam selain mudah diaplikasikan, teknik optimasi ini juga berfokus pada waktu komputasi yang cepat dan efisien serta membutuhkan memori yang kecil dan penyesuaian parameter yang lebih (Sarno dkk., 2023).

Dalam Multilayer Perceptron (MLP), tidak ada ketentuan khusus dalam menentukan lapisan *hidden layer* dan banyak neuron. Model ideal adalah yang memiliki akurasi tinggi, sehingga penyesuaian jumlah neuron dan *hidden layer* harus dilakukan dengan tepat, karena meskipun penambahan neuron atau *hidden layer* dapat meningkatkan akurasi, hal ini tidak selalu menjamin hasil yang optimal (Saputra & Kristiyanti, 2022). Penentuan jumlah *hidden layer* juga harus disesuaikan dengan kompleksitas data dan pola nonlinier yang dipelajari. Model dengan tiga layer atau lebih dapat digunakan untuk menangani masalah yang lebih kompleks, namun memerlukan regularisasi yang tepat untik mencegah *overfitting* pada model (Bengio dkk., 2021; Lecun dkk., 2015). Dalam penelitian ini, banyak *hidden layer* yang digunakan adalah dimulai dengan 2 lapisan sampai 5 lapisan serta jika terdapat kondisi *overfitting* pada model teknik regularisasi akan diaplikasikan.

Menurut Sarno dkk., (2023), regularisasi mengacu pada serangkaian teknik berbeda yang menurunkan kompleksitas model pada tahap pelatihan sehingga dapat mencegah *overfitting* dan yang paling populer dan efisien adalah L1, L2 dan dropout. Dalam penelitian ini, teknik regularisasi yang digunakan adalah L2. Regularisasi L2 atau yang juga dikenal sebagai ridge

regularization atau *weight decay* adalah teknik regularisasi dengan menambahkan pinalti terhadap bobot dalam jaringan saraf dengan tujuan untuk membatasi sehingga tidak terlalu besar yang dapat menyebabkan model *overfitting* (Goodfellow dkk., 2016). Regularisasi L2 dipilih jika diperlukan karena selain dapat mengurangi *overfitting* dan peningkatan generalisasi, regularisasi L2 hanya mengecilkan bobot dan tidak menghapusnya sama sekali.

Dalam persamaan (2.11) $f(.)$ merupakan fungsi aktivasi nonliear yang dapat berupa fungsi aktivasi ReLU, sigmoid maupun tanh. Rectified Linear Unit (ReLU) memiliki aktivasi yang bersifat linear untuk nilai input yang lebih besar dari nol, atau dapat dinyatakan sebagai

$$R(x) = \max(0, x) = \begin{cases} x, \text{ jika } x \geq 0 \\ 0, \text{ lainnya} \end{cases} \tag{2.12}$$

fungsi aktivasi sigmoid merupakan fungsi yang paling umum dengan memetakan masukan yang masuk ke rentang antara 0 dan 1, secara matematis dapat didefinisikan sebagai

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.13}$$

sedangkan fungsi tanh adalah fungsi hiperbolik tangen yang memtakan angka bernilai riil ke rentang $[-1,1]$ yang secara matematis dapat dinyatakan sebagai

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.14}$$

Menurut Nair dan Hinton (2010) fungsi aktivasi ReLU yang dinyatakan dalam persamaaan (2.12) memiliki performa yang lebih baik dibandingkan dengan fungsi aktivasi lain terutama untuk data yang berskala besar dan memiliki kompleksitas tinggi.

2.2.7 Metrik Akurasi

Metrik Metrik akurasi yang digunakan untuk mengetahui performa model dalam penelitian ini adalah *Mean Absolute Error* (MAE) dan *Symmetric Mean Absolute Percentage Error* (SMAPE). *Mean Absolute Error* (MAE) secara matematis dapat dinyatakan sebagai berikut (Ciulla & D'Amico, 2019)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.15}$$

Untuk menghitung nilai *Symmetric Mean Absolute Percentage Error* (SMAPE), secara matematis dapat dinyatakan sbeagai berikut (Makridakis & Hibon, 2000)

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)/2} \times 100\% \tag{2.16}$$

Menurut Purnama dan Hendarsin (2020) serta (Manuaba dkk., 2023) akurasi yang dihasilkan oleh persamaan (2.16) dapat direpresentasikan akurasinya sebagaimana ditunjukkan pada tabel berikut

Tabel 2.2 Representasi akurasi

Nilai (%)	Tingkat Akurasi
<10	Sangat akurat
10-20	Akurat
20-50	Kurang akurat
>50	Tidak akurat