

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

##### **2.1.1 Pertumbuhan Ekonomi**

Pertumbuhan ekonomi adalah suatu proses kenaikan output perkapita dalam jangka panjang, persentase pertumbuhan output harus lebih tinggi dari pada persentase pertumbuhan jumlah penduduk dan ada kecendrungan dalam jangka panjang bahwa pertumbuhan itu akan berlanjut (Srihidayati, 2022). Pertumbuhan penduduk akan mempengaruhi pembangunan ekonomi. Pertambahan penduduk dapat memperluas pasar, perluasan pasar meningkatkan spesialisasi dalam perekonomian. Pertumbuhan ekonomi adalah kenaikan kapasitas dalam jangka panjang dari negara yang bersangkutan untuk menyediakan berbagai barang ekonomi bagi penduduknya (Simanungkalit, 2020).

Ada tiga faktor atau komponen utama pertumbuhan ekonomi, yaitu akumulasi modal, pertumbuhan penduduk, dan hal-hal yang berhubungan dengan kenaikan jumlah angkatan kerja yang dianggap secara positif merangsang pertumbuhan ekonomi. Guna mendorong pembangunan nasional yang berkelanjutan, maka harus didukung dengan pembangunan daerah yang dilaksanakan secara tepat. Pertumbuhan ekonomi daerah adalah peningkatan kapasitas suatu wilayah atau daerah tertentu dalam menghasilkan barang dan jasa dalam jangka waktu tertentu terfokus dalam dinamika ekonomi regional atau lokal. Pembangunan ekonomi daerah adalah suatu proses dimana pemerintah daerah dan masyarakatnya mengelola sumber sumber daya yang ada dan membentuk suatu pola kemitraan antara pemerintah daerah dan swasta untuk menciptakan lapangan kerja baru dan merangsang perkembangan kegiatan ekonomi dalam wilayah tersebut (Azulaidin, 2021). Laju pertumbuhan ekonomi daerah biasanya digunakan untuk menilai seberapa jauh keberhasilan pembangunan daerah dalam periode waktu tertentu. Pertumbuhan ekonomi daerah tersebut dapat ditunjukkan oleh kenaikan PDRB.

### 2.1.2 Konsep Produk Domestik Regional Bruto (PDRB)

Nilai tambah bruto seluruh barang dan jasa yang tercipta atau dihasilkan di wilayah domestik suatu negara yang timbul akibat berbagai aktivitas ekonomi dalam suatu periode tertentu tanpa memperhatikan apakah faktor produksi yang dimiliki residen atau non-residen disebut dengan Produk Domestik Regional Bruto (PDRB) (BPS, 2023). Menurut Charles *et al.* (2021) Produk Domestik Regional Bruto (PDRB) merupakan nilai bersih dari jasa dan produk terakhir yang dibuat oleh macam-macam kegiatan ekonomi di sebuah wilayah selama jangka waktu tertentu.

Produk Domestik Regional Bruto maupun agregat turunannya disajikan dalam 2 (dua) versi penilaian, yaitu atas dasar harga berlaku dan atas dasar harga konstan. Harga Berlaku adalah penilaian yang dilakukan terhadap produk barang dan jasa yang dihasilkan ataupun yang dikonsumsi pada harga tahun sedang berjalan. Harga Konstan adalah penilaian yang dilakukan terhadap produk barang dan jasa yang dihasilkan ataupun yang dikonsumsi pada harga tetap di satu tahun dasar (BPS, 2023). PDRB berdasarkan harga berlaku bertujuan untuk melihat struktur perekonomian suatu wilayah dengan menunjukkan kemampuan sumber daya ekonomi yang dihasilkan oleh suatu wilayah. PDRB berdasarkan harga konstan dapat digunakan untuk menunjukkan laju pertumbuhan ekonomi secara keseluruhan pada suatu wilayah. PDRB berdasarkan harga konstan adalah nilai tambah barang dan jasa yang dihitung menggunakan harga pada tahun tertentu sebagai tahun dasar yaitu menggunakan tahun dasar 2010.

Pemilihan tahun dasar 2010 dipilih sebagai tahun dasar baru, yang dipakai menggantikan tahun dasar 2000. Badan Pusat Statistik saat ini telah melakukan perubahan tahun dasar secara berkala sebanyak 6 (enam) kali yaitu tahun 1960, 1973, 1983, 1993, 2000, dan 2010. Terdapat beberapa alasan penggantian tahun dasar (BPS, 2023)

1. Perekonomian Indonesia tahun 2010 relatif stabil yang dipengaruhi beberapa faktor pertumbuhan ekonomi Indonesia pada tahun 2010 mencapai sekitar 6,1%, yang didorong oleh konsumsi domestik, investasi, dan ekspor komoditas.
2. Rekomendasi PBB tentang pergantian tahun dasar dilakukan setiap 5(lima) atau 10 (sepuluh) tahun.

3. Tersedianya sumber data baru untuk perbaikan PDRB seperti sensus penduduk 2010.
4. Tersedianya kerangka kerja *Supply and Use Table* (SUT) yang menggambarkan keseimbangan aliran produksi dan konsumsi.
5. Telah terjadi perubahan struktur ekonomi selama 10 (sepuluh) tahun terakhir terutama dibidang informasi dan teknologi serta transportasi yang berpengaruh terhadap pola distribusi dan munculnya produk-produk baru.

Perubahan tahun dasar PDRB ini memiliki manfaat diantaranya: menginformasikan perekonomian regional yang terkini seperti pergeseran struktur dan pertumbuhan ekonomi, meningkatkan kualitas data PDRB, dan menjadikan data PDRB dapat diperbandingkan secara internasional (BPS, 2023).

#### **2.1.2.1 PDRB Atas Dasar Harga Konstan (ADHK)**

PDRB ADHK (riil) dapat digunakan untuk menunjukkan laju pertumbuhan ekonomi secara keseluruhan setiap kategori dari tahun ke tahun. Pertumbuhan PDRB ADHK biasa disebut dengan pertumbuhan ekonomi, yang menggambarkan kenaikan produksi riil tanpa dipengaruhi inflasi. Disebut harga konstan karena penilaianya didasarkan kepada harga satu tahun dasar tertentu. Laju pertumbuhan produk domestik regional bruto diperoleh dari perhitungan PDRB atas dasar harga konstan (BPS, 2023). Produk domestik regional bruto atas dasar harga konstan menunjukkan nilai tambah barang dan jasa yang dihitung menggunakan harga pada tahun tertentu sebagai tahun dasar (Fediansyah, 2023).

#### **2.1.2.2 PDRB Atas Dasar Harga Berlaku (ADHB)**

PDRB atas dasar harga berlaku (nominal) menunjukkan kemampuan sumber daya ekonomi yang dihasilkan oleh suatu wilayah. PDRB ADHB dipengaruhi harga (inflasi/deflasi). Nilai PDRB yang besar menunjukkan kemampuan sumber daya ekonomi yang besar, begitu juga sebaliknya. Disebut sebagai harga berlaku karena seluruh agregat dinilai dengan menggunakan harga pada tahun berjalan. Laju pertumbuhan produk domestik regional bruto diperoleh dari perhitungan PDRB atas dasar harga konstan (BPS, 2023). Produk Domestik Regional Bruto atas dasar harga berlaku menggambarkan nilai tambah barang dan jasa yang dihitung menggunakan harga pada setiap tahun (Citra *et al.* 2023)

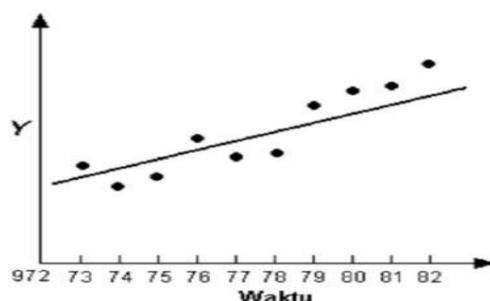
### 2.1.3 Analisis Deret Waktu (*Time Series*)

Data deret waktu (*time series*) adalah data yang dihasilkan dari pencatatan terus menerus dari waktu ke waktu, biasanya dilakukan dalam interval waktu yang sama. Analisis deret waktu memiliki tujuan agar dapat memahami pola, hubungan atau tren yang terjadi dari waktu ke waktu dengan urutan pengamatan yang berjarak sama dalam pada interval waktu tertentu, seperti harian, mingguan, atau bulanan. Analisis deret waktu dapat digunakan untuk membuat prediksi, perencanaan dan pengambilan keputusan.

Peramalan deret waktu adalah proses menggunakan model peramalan untuk memprediksi nilai masa depan suatu variabel berdasarkan nilai-nilai yang diamati sebelumnya (Ma'Fulloh *et al.* 2022). Dalam pengembangan analisis *time series* banyak digunakan dibeberapa bidang seperti bidang ekonomi, bidang keuangan, bidang transportasi, bidang pertanian dan masih banyak lagi untuk membuat model yang cocok untuk meramalkan data terdapat beberapa tahapan proses dalam analisis *time series* yaitu: stasioneritas data, *Parameter Estimation*, *Model Specification*, *Model Checking*, *Unit Root Test*, dan *Forecasting* (Asrini, 2020). Dalam *time series*, pola dibagi menjadi tiga yaitu, pola *trend*, pola siklus dan pola musiman (Wahyu & Hendrik, 2023). Dalam pemilihan metode ada beberapa pola data yang harus dipelajari antara lain:

#### 2.1.3.1. Pola *Trend*

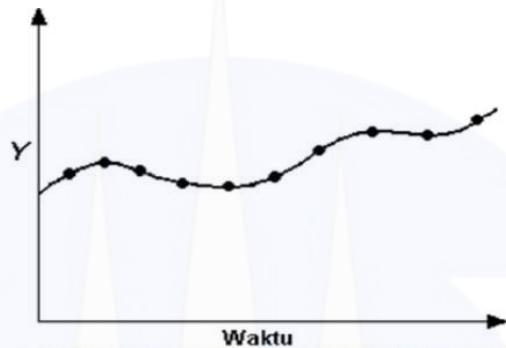
Pola atau kecenderungan yang terlihat dalam data seiring waktu yang menunjukkan arah pergerakan data disebut dengan pola data *trend*. Pola ini menunjukkan kecenderungan pergerakan data dalam jangka panjang, baik naik (*uptrend*), turun (*downtrend*), atau tetap stabil. Pola data *trend* terjadi jika terdapat kenaikan atau penurunan sekuler jangka panjang dalam data (Andini & Auristandi, 2020).



Gambar 2.1. Pola Data *Trend* (Ariyanto *et al.* 2020)

### 2.1.3.2 Pola Siklus

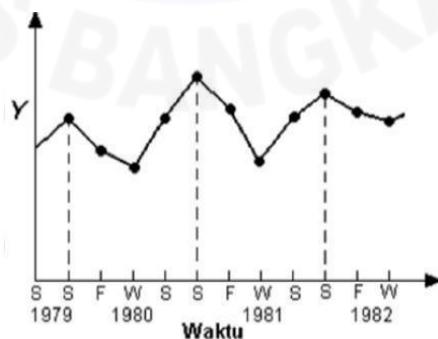
Pola data siklus merupakan pola dalam data yang berulang dalam jangka waktu panjang dan biasanya dipengaruhi oleh beberapa. Seperti faktor ekonomi, bisnis, atau faktor makro lainnya. Pola siklus ini berbeda dari pola musiman karena tidak memiliki interval waktu yang tetap dan sering kali berlangsung selama beberapa tahun. Pola data ini terjadi jika terdapat data yang dipengaruhi oleh fluktuasi ekonomi jangka panjang seperti yang berhubungan dengan siklus bisnis (Ariyanto *et al.* 2020).



Gambar 2.2. Pola Data Siklus (Ariyanto *et al.* 2020)

### 2.1.3.3 Pola Musiman

Pola data musiman adalah pola data yang terjadi secara berulang dan juga teratur dalam periode tertentu, seperti harian, bulanan, atau tahunan. Pola data ini terjadi karena faktor musiman seperti cuaca, hari libur, atau kebiasaan masyarakat. Pola data ini terjadi jika terdapat suatu deret data yang dipengaruhi oleh faktor musiman misalnya kuartal tahun tertentu, bulanan, atau hari-hari pada minggu tertentu (Ariyanto *et al.* 2020).



Gambar 2.3. Pola Data Musiman (Ariyanto *et al.* 2020)

#### 2.1.4 *Machine Learning*

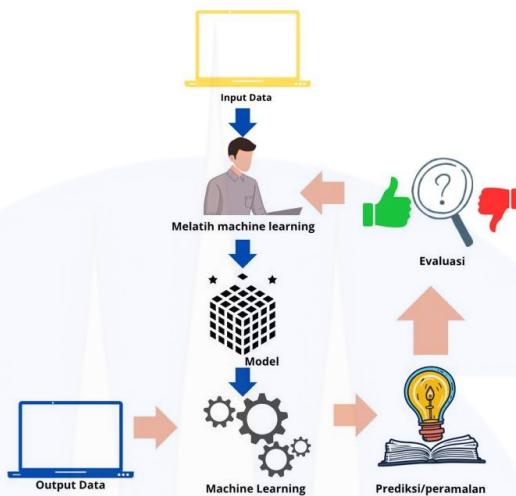
*Machine Learning* (ML) adalah salah satu aplikasi dari *Artificial intelligence (AI)* yang fokus kepada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus diprogram berulang kali. *Machine Learning* (ML) membutuhkan sebuah data (data *training*) sebagai proses *learning* sebelum menghasilkan sebuah hasil (Chazar & Erawan, 2020). *Machine Learning* adalah ilmu yang mempelajari tentang algoritma komputer yang bisa mengenali pola-pola didalam data, dengan tujuan untuk mengubah beragam macam data menjadi suatu tindakan yang nyata dengan sesedikit mungkin campur tangan manusia. *Machine Learning* (ML) dapat menciptakan mesin (komputer) yang belajar dari data yang ada, selanjutnya dia bisa membuat keputusan secara mandiri tanpa perlu diprogram lagi (Kurniawan, 2020:5).

Pada hasil evaluasi yang dihasilkan tidak memvalidasi akurasi algoritma dengan meningkatkan dan mengubah masing-masing pembagian data dalam algoritma *Machine Learning* (ML). Proses ini sangat penting dalam setiap algoritma *Machine Learning* (ML) untuk menghindari *underfitting* dan *overfitting*. *Overfitting* disebabkan oleh model yang terlalu kompleks dan terpengaruh oleh data latih, *overfitting* juga dapat dilihat pada nilai *error* data uji jika nilai *error* data uji lebih besar dibandingkan dengan nilai *error* data latih maka data tersebut dikatakan *overfitting*. *Underfitting* disebabkan oleh model yang terlalu sederhana dan tidak dapat memahami pola dari data. *Underfitting* dapat dilihat dari nilai *error* data latih, jika nilai *error* data latih lebih besar dari data uji maka data bisa dikatakan *underfitting* (Firmansyah *et al.* 2023).

Dataset terdiri *Machine Learning* dari dataset *training* dan dataset *testing*. Dataset *training* digunakan dalam mencari model yang cocok dalam melatih algoritma, sementara dataset *testing* akan dipakai untuk evaluasi dari kinerja dari model yang didapat (Theodorus *et al.* 2021). Kegunaan dari *Machine Learning* antara lain sebagai berikut (Pambudi *et al.* 2020). :

1. *Classification* (*Klasifikasi*) adalah metode *machine learning* yang digunakan untuk memprediksi suatu nilai/kelas sebuah individu dalam sebuah populasi.

2. *Similarity matching* (Pencocokan kemiripan) adalah metode *machine learning* yang digunakan untuk mengidentifikasi kemiripan antar individu berdasarkan data yang ada.
3. *Clustering* (Pengklasteran) adalah metode *machine learning* yang digunakan untuk mengelompokkan individu dalam grup yang sama berdasarkan kesamaan yang dimiliki.



**Gambar 2.4.** Cara Kerja *Machine Learning*

### 2.1.5 *Support Vector Machine (SVM)*

*Support Vector Machine (SVM)* dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory* (Putri, 2020). *Support Vector Machine (SVM)* merupakan algoritma pembelajaran mesin yang dapat digunakan untuk klasifikasi dan regresi, akan tetapi lebih sering digunakan dalam klasifikasi data. Metode *Support Vector Machine (SVM)* merupakan suatu metode klasifikasi dalam pemecahan masalah untuk dua kelas. SVM biasanya digunakan dalam mengklasifikasikan data yang bisa dipisah secara linier, untuk data yang tidak dapat dilakukan pemisahan secara linier maka digunakan fungsi kernel yang tujuannya adalah untuk memetakan data *input* ke data *fitur* (Amrozi *et al.* 2021).

*Support Vector Machine (SVM)* adalah salah satu metode *Machine Learning* yang bekerja berdasarkan prinsip *Structural Risk Minimization (SRM)* yang bertujuan untuk mencari *hyperplane* terbaik dengan cara memisahkan dua buah kelas pada input *space*. Tujuan penggunaan kernel yaitu untuk merubah data

ke ruang yang berdimensi tinggi dengan cara menjadikan data non-linier terpisah secara linier (Riyantoni *et al.* 2023). Menurut Ishlah *et al.* (2023) mengemukakan dengan sederhana konsep SVM, yaitu menemukan *hyperplane* terbaik yang dapat memisahkan dua kelas pada input *space*. *Hyperplane* berupa garis batas sebagai pemisah antara kedua kelas. *Hyperplane* yang letaknya di tengah-tengah antara dua set objek dari dua kelas disebut *hyperplane* terbaik.

SVM didasarkan pada konsep pengklasifikasi linear, yang memisahkan kasus-kasus yang dapat dipisahkan secara linear. Akan tetapi, SVM telah didesain untuk dapat menangani masalah non-linear dengan menggunakan konsep kernel pada ruang berdimensi tinggi (Sutrisno, 2018). *Linear Support Vector Machine* (SVM) bekerja dengan menempatkan *hyperplane* optimal sebagai pemisah untuk dua kelas di ruang input. Konsep ini terdiri dari *pattern* atau pola, *margin*, dan *support vector*. Pola adalah anggota dari dua kelas yang dipisahkan oleh *hyperplane* dengan nilai 1 dan kelas lain dengan nilai -1. Margin adalah jarak antara *hyperplane* dengan *pattern* terdekat dari masing masing kelas. Pola terdekat tersebut disebut *support vector* (Rahmadini *et al.* 2024). SVM yang biasanya menghasilkan nilai bulat dalam kasus klasifikasi.

### **2.1.6 *Support Vector Regression* (SVR)**

*Support Vector Regression* (SVR) adalah sebuah metode yang pertama kali diperkenalkan oleh Vladimir N. Vapnik, Harris Drucker, Christopher J. C. Burges, Linda Kaufman dan Alexander J. Smola pada tahun 1996 (Pura & Kurniawan, 2021). *Support Vector Regression* (SVR) adalah teori yang diadaptasi dari teori *Support Vector Machine* (SVM) untuk kasus regresi yang menghasilkan keluaran berupa bilangan riil. Konsep algoritma SVR dapat menghasilkan nilai peramalan yang bagus karena SVR mempunyai kemampuan menyelesaikan masalah *overfitting*. *Overfitting* adalah perilaku data saat data uji atau latih menghasilkan akurasi prediksi hampir sempurna (Sudarmin, 2022).

Metode Support Vector Machine (SVM) merupakan penerapan teori *machine learning* untuk kasus klasifikasi yang menghasilkan keluaran berupa nilai diskrit, sedangkan Support Vector Regression (SVR) merupakan penerapan SVM pada kasus regresi yang menghasilkan keluaran berupa bilangan riil (Maulana *et al.* 2019). Penggunaan metode SVR mempunyai kelebihan dibandingkan metode

*Linier Regression* (regresi linier) yaitu jika regresi linier menghasilkan fungsi linier yaitu berupa garis lurus, pada algoritma SVR menghasilkan *trend* data yang bergelombang mengikuti jalur data yang terbentuk, sehingga prediksi data yang dihasilkan lebih akurat. Fungsi regresi  $f(x)$  dengan batas deviasi ( $\epsilon$ ) sama dengan 0, merupakan fungsi regresi yang sempurna, sedangkan model SVR pada fungsi regresi non-linier, mengacu pada persamaan di bawah ini:

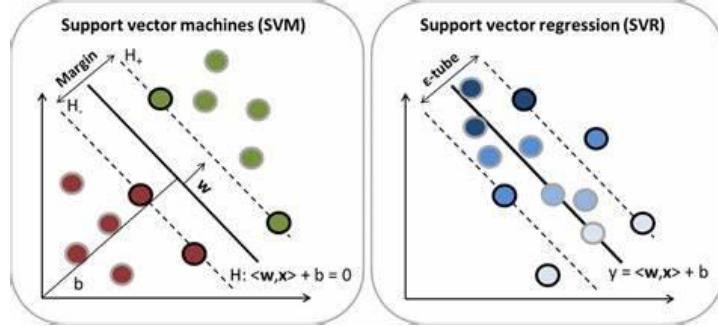
$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad [2.1]$$

Keterangan :

- $f(x)$  = fungsi regresi SVR
- $n$  = jumlah data pelatihan
- $\alpha_i - \alpha_i^*$  = pengali lagrange data ke- $i$
- $x_i$  = data input ke- $i$
- $x$  = data input yang diprediksi
- $b$  = bias

SVR dilatih dengan menerapkan algoritma pembelajaran untuk mendukung teori optimasi. Ide dasar dari SVR adalah mengubah ruang input menjadi ruang berdimensi tinggi melalui transformasi nonlinier, kemudian menemukan hubungan linier antara variabel input dan output pada ruang fitur tersebut. Besar kecilnya nilai akurasi yang akan dihasilkan pada model yang dibangun pada proses pelatihan dengan metode SVR sangat bergantung pada fungsi kernel dan parameter yang digunakan. Menurut furi, (2015) dalam Wahyudi *et al.* (2023) keunggulan SVR adalah kemampuan untuk mengatasi masalah data nonlinear dengan trik kernel. Konsep SVR didasarkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting*.

Metode ini juga mampu menemukan fungsi  $f(x)$  sebagai *hyperplane* untuk semua input data yang memiliki deviasi paling besar dari target aktual data training dan dapat membuat *error* setipis mungkin (Cahyono *et al.* 2022). Menurut Purnama, (2020) keunggulan SVR adalah kemampuan untuk mengatasi masalah data non-linear dengan trik kernel dan dapat mengatasi masalah *overfitting* dimana model yang dihasilkan hanya menghasilkan model yang baik untuk data *training* dan tidak untuk data *testing*.



**Gambar 2.5.** Hyperplane terbaik SVM dan SVR (Raquel, 2022).

### 2.1.7 Fungsi Kernel

Fungsi kernel dalam matematika adalah fungsi yang digunakan untuk mengubah data yang tidak linear (berkorelasi secara kompleks) menjadi data linear. Data yang linear dapat digunakan agar lebih mudah diolah oleh algoritma pembelajaran. Kernel berguna ketika kita memiliki data yang sulit dipisahkan dengan garis lurus, akan tetapi data tersebut bisa dipisahkan jika kita mengubah perspektif atau ruangnya. Fungsi kernel harus diatur dengan benar karena dapat mempengaruhi akurasi regresi (Purnama, 2020). Ada beberapa fungsi kernel yang sering digunakan khususnya dalam *support vector*, diantaranya adalah fungsi kernel Linear, fungsi kernel Polinomial, Fungsi *Gaussian Radial Basic Function* (RBF), dan fungsi kernel Sigmoid (Sarina, 2021). Berikut adalah rumus fungsi – fungsi kernel tersebut:

#### 2.1.7.1 Fungsi Kernel Linear

Fungsi kernel ini merupakan fungsi kernel yang paling sederhana, yang ditandai dengan perkalian titik dari,  $x_i \cdot x_j$ . Fungsi pada linier merupakan fungsi dengan variabel bebas berpangkat satu. Fungsi linier dapat digunakan dalam bentuk kernel yang digunakan sebagai sarana pemetaan (Martha & Sulistianingsih, 2019). Persamaan fungsi kernel linier adalah sebagai berikut:

$$K(x_i, x_j) = x_i \cdot x_j \quad [2.2]$$

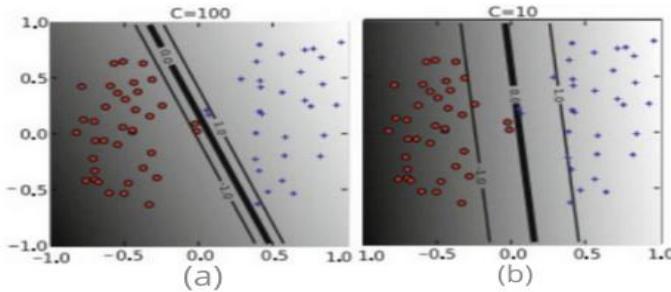
Keterangan :

$K(x_i \cdot x_j)$  = fungsi kernel linear

$x_i$  = vektor input data ke- $i$

$x_j$  = vektor input data ke- $j$

$x_i \cdot x_j$  = hasil perkalian titik



**Gambar 2.6.** (a) Kernel Linear C=100, (b) Kernel Linear C=10 (Ben-Hur & Weston, 2009).

Pada Gambar 2.6 jika nilai *cost* yang rendah (10) maka nilai *margin error* yang rendah, memperlebar nilai *margin* dan juga sebaliknya jika nilai *cost* yang tinggi (100) maka *margin error* besar dan mempersempit *margin*.

#### 2.1.7.2 Fungsi Kernel *Polynomial*

Fungsi kernel *Polynomial* ini biasa digunakan dalam SVR dan metode kernel lainnya, kernel ini mempresentasikan pemetaan vektor menggunakan fungsi *polynomial* untuk kasus non linier (Martha & Sulistianingsih, 2019). Persamaan fungsi kernel Polynomial adalah sebagai berikut:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^p \quad [2.3]$$

Keterangan :

$K(x_i, x_j)$  = fungsi kernel polinomial

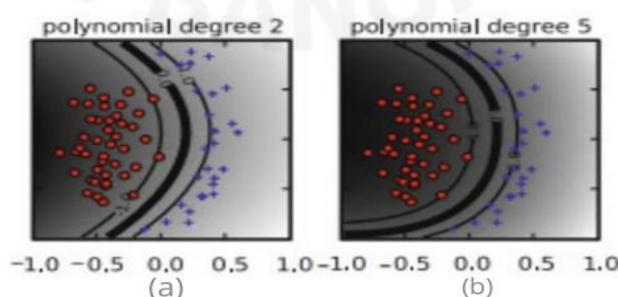
$x_i$  = vektor input data ke-*i*

$x_j$  = vektor input data ke-*j*

$x_i \cdot x_j$  = perkalian titik

$c$  = konstanta

$p$  = derajat polinomial

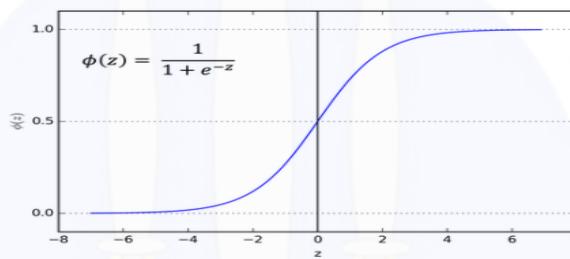


**Gambar 2.7.** (a) Kernel *Polynomial degree 2*, (b) Kernel *Polynomial degree 5* (Ben-Hur & Weston, 2009).

Derajat atau *degree* pada kernel polinomial dapat mengendalikan fleksibilitas dari hasil klasifikasi, semakin tinggi nilai derajat pada kernel polinomial memungkinkan *decision boundary* yang lebih fleksibel (Ben-Hur & Weston, 2009). Menggunakan derajat polinomial yang terlalu tinggi dapat mengakibatkan *overfitting* atau model yang digunakan terlalu fokus pada data training, sehingga pengujian dengan data testing yang berbeda akan menyebabkan penurunan akurasi (Dini, 2023).

### 2.1.7.3 Fungsi Kernel Sigmoid

Fungsi sigmoid merupakan fungsi trigonometri hiperbolik yang merupakan hasil kombinasi dari fungsi – fungsi eksponen, fungsi ini memiliki bentuk grafik berbentuk S ketika *diplot* (Martha & Sulistianingsih, 2019). Kernel sigmoid berasal dari bidang *Neural Network* di mana fungsi sigmoid sering digunakan sebagai fungsi aktivasi untuk *neuron* buatan (Satria, 2022).



**Gambar 2.8.** Grafik fungsi aktivasi Sigmoid (Tejakusuma, 2019)

Ada dua parameter yang dapat disesuaikan dalam kernel sigmoid. Nilai yang umum untuk *alpha* adalah  $1/N$ , dimana  $N$  adalah dimensi data. Kernel sigmoid didefinisikan menggunakan persamaan sebagai berikut:

$$K(x_i, x_j) = \tanh(y x_i \cdot x_j + r) \quad [2.4]$$

Keterangan:

$K(x_i, x_j)$  = fungsi kernel polinomial

$x_i$  = vektor input data ke-*i*

$x_j$  = vektor input data ke-*j*

$x_i \cdot x_j$  = perkalian titik

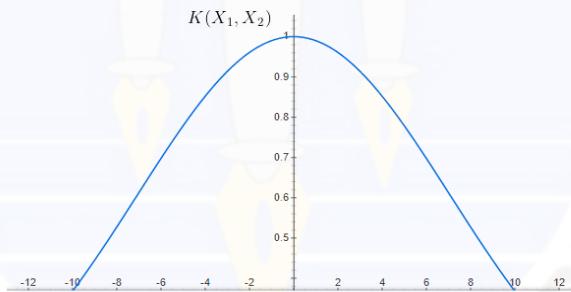
$y$  = parameter skala

$r$  = konstanta

#### 2.1.7.4 Fungsi *Radial Basic Function* (RBF)

Fungsi RBF merupakan fungsi yang berbentuk lonceng. Menurut Meyer & Wien, (2015) dalam Kemala, (2023) Kernel RBF mempunyai dua parameter, yaitu *Gamma* dan *Cost*. Parameter *cost* sebagai parameter yang digunakan untuk menghindari misklasifikasi pada setiap sampel dalam *training* dataset. Parameter *Gamma* dapat menentukan seberapa jauh pengaruh dari satu sampel *training* dataset dengan nilai rendah berarti jauh, dan nilai tinggi berarti dekat. Jika nilai *gamma* rendah, maka titik yang berada jauh dari garis pemisah yang logis dalam perhitungan untuk garis pemisah. Sedangkan, ketika *gamma* tinggi berarti titik-titik berada di sekitar garis yang logis akan dipertimbangkan dalam perhitungan.

Kernel ini dapat memiliki akurasi yang sama dengan linier dan *Polynomial*. Hal ini menjadikan kernel RBF sebagai kernel yang mewakili kernel lainnya (dapat digunakan dalam berbagai kasus) (Martha dan Sulistianingsih, 2019). Pada kernel RBF fungsi aktivasi memiliki nilai center di tengah dari lembar fungsi basis, sehingga yang dihitung adalah area antara nilai center pada fungsi dengan jarak antara titik dengan garis fungsi *gaussian*.



**Gambar 2.9.** Representasi grafis fungsi kernel RBF (Fauzan *et al.* 2011)

Fungsi kernel RBF menghitung antara dua vektor, rumus persamaan yang digunakan menggunakan rumus persamaan sebagai berikut:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad [2.5]$$

Keterangan:

$K(x_i, x_j)$  = fungsi kernel polinomial

$x_i$  = vektor input data ke-*i*

$x_j$  = vektor input data ke-*j*

$\|x_i - x_j\|^2$  = jarak Euclidean kuadrat antara  $x_i$  dan  $x_j$

$\gamma$  = parameter lebar kernel ( $\gamma > 0$ )

Pemilihan kernel berpengaruh terhadap tingkat akurasi serta *Root Mean Squared Error*. Berdasarkan penelitian, kernel yang sesuai untuk peramalan atau prediksi adalah kernel *Radial Basic Function* (RBF) (Assafat, 2015). Kernel RBF merupakan salah satu kernel yang digunakan untuk menentukan parameter terbaik (Ruliana, 2024).

### 2.1.8 Normalisasi dan Denormalisasi

Normalisasi data merupakan salah satu proses yang dilakukan pada fase *preprocessing* (pra proses) data. Dengan adanya *preprocessing* menjadikan penerapan algoritma menjadi lebih efisien. Metode normalisasi data adalah proses membuat beberapa variabel memiliki rentang nilai yang sama tidak ada yang terlalu besar maupun terlalu kecil sehingga dapat membuat analisis statistik menjadi lebih mudah (Kusnaldi *et al.* 2022). Fungsi tersebut akan membawa nilai input dengan *range* yang tak terbatas ke nilai *output* yang terbatas, yaitu dalam sebuah *range* 0 sampai 1. Agar dapat membawa *range* nilai *output* ke dalam *range input*, maka data input harus dilakukan normalisasi data ke dalam range 0 sampai 1, sehingga outputnya dapat di denormalisasi ke dalam range nilai input. Proses normalisasi data terdapat beberapa metode yang dapat digunakan. Pada penelitian ini digunakan metode normalisasi data *min max*. Normalisasi data *min max* adalah metode normalisasi dengan mentransformasi linier data asli sehingga menghasilkan nilai perbandingan yang berkorelasi antar data sebelum dan sesudah proses (Novianti *et al.* 2022). Adapun rumus dari metode *min max*:

$$x' = \frac{(x - x_{\min})}{x_{\max} - x_{\min}} \quad [2.6]$$

Keterangan:

$x'$  = nilai data hasil normalisasi

$x$  = nilai data asli

$x_{\max}$  = nilai maksimum data

$x_{\min}$  = nilai maksimum data

Data yang dinormalisasi ( $x$ ) dioperasikan dengan nilai minimum aktual ( $x_{\min}$ ) dan nilai maksimum aktual ( $x_{\max}$ ). Adapun *range* maksimum normalisasi dalam data disebut *Max range* dan *range* minimum normalisasi data disebut *Min range* (Novianti *et al.* 2022).

Sedangkan denormalisasi data dilakukan ketika sudah mendapatkan nilai prediksi dari proses pengujian. Tujuan denormalisasi data adalah mengembalikan data ke nilai aslinya seperti sebelum dinormalisasi. Rumus dari normalisasi data dapat dilihat pada persamaan (Novianti *et al.* 2022).

$$x = x'(x_{max} - x_{min}) + x_{min} \quad [2.7]$$

keterangan:

$x$  = nilai data asli

$x'$  = nilai data hasil normalisasi

$x_{max}$  = nilai maksimum data

$x_{min}$  = nilai minimum data

Data yang didenormalisasi  $x(x)$  dioperasikan dengan nilai maksimum aktual ( $x_{max}$ ) dan nilai minimum aktual ( $x_{min}$ ) (Novianti *et al.* 2022).

### 2.1.9 Mean Absolute Percentage Error (MAPE)

*Mean Absolute Percentage Error* (MAPE) merupakan rata-rata difrensiasi absolut antara nilai pramalan dan nilai actual, untuk perhitungannya sebagai rata-rata difrensiasi absolut antara nilai yang diramalkan dan aktual (Rusdy *et al.* 2022). Pengukuran persentase *error* memiliki kelebihan menjadi *scale-independent* dan sering digunakan untuk membandingkan performa peramalan antara kumpulan data yang berbeda. Pengukuran performa menggunakan MAPE akan menghasilkan nilai berupa persentase. MAPE memberikan hasil dalam bentuk persentase, yang lebih mudah dipahami. Persamaan yang digunakan untuk memperoleh nilai *Mean Absolute Percentage Error* (MAPE) yaitu:

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \quad [2.8]$$

Keterangan:

**MAPE** = *Mean Absolute Percentage Error*

$n$  = jumlah data

$A_t$  = nilai aktual pada periode ke- $t$

$F_t$  = nilai hasil peramalan pada periode ke- $t$

Adapun standar minimal untuk mengetahui kinerja model peramalan dapat dilihat pada tabel 2.1.

**Tabel 2.1** Kriteria *Mean Absolute Percentage Error* (MAPE)

MAPE (Persen)	Signifikasi
<10	Sangat baik
10-20	Baik
20-50	Cukup
>50	Buruk

### 2.1.10 Root Mean Square Error (RMSE)

*Root Mean Square Error* (RMSE) adalah sebuah metode yang digunakan untuk mengevaluasi keakuratan sebuah algoritma berdasarkan nilai *error*, metode ini biasa digunakan untuk hasil dari sebuah peramalan (Supriyanto *et al.* 2022). *Root Mean Square Error* (RMSE) merupakan ukuran yang baik, akan tetapi hanya untuk membandingkan kesalahan peramalan model yang berbeda untuk variabel tertentu dan tidak antara variabel, karena skala *dependent*. Fungsi dari pengukuran *Root Mean Square Error* (RMSE) untuk agregat besaran kesalahan dalam prediksinya dikarenakan beberapa kali menjadi ukuran tunggal daya prediksi. RMSE juga merupakan akar kuadrat rata-rata perbedaan kuadrat antara prediksi dan pengamatan aktualnya, dimana aturan penilaian kuadratnya juga mengukur besaran rata-rata kesalahan. Persamaan yang digunakan untuk memperoleh nilai pengukuran *Root Mean Square Error* (RMSE) yaitu:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad [2.9]$$

Keterangan:

**RMSE** = *Root Mean Square Error*

**n** = jumlah data

**y<sub>t</sub>** = nilai aktual pada periode ke-*t*

**ŷ<sub>t</sub>** = nilai hasil peramalan pada periode ke-*t*

### 2.1.11 Autoregressive Integrated Moving Average (ARIMA)

Metode ARIMA (*AutoRegressive Integrated Moving Average*) merupakan salah satu teknik dalam menganalisis data deret waktu (*time series*). Analisis deret waktu (*time series*) merupakan serangkaian data pengamatan yang terjadi berdasarkan indeks waktu secara berurutan dengan interval waktu tetap (Pujiati *et al.* 2022). Metode ARIMA digunakan untuk memodelkan dan juga meramalkan

data berdasarkan pola historisnya. Model ARIMA, atau *Box-Jenkin*, adalah metode statistik tanpa melibatkan variabel bebas dalam perumusannya. Dikembangkan oleh Box dan Jenkins pada 1976 (Ruliana, 2023). Metode ARIMA membuat model peramalan jangka pendek yang akurat yang sepenuhnya mengabaikan independen variabel dengan menggunakan data *time series* (Fauzani & Rahmi, 2023). George Box dan Gwilym Jenkins ialah orang pertama kali yang mengembangkan ARIMA untuk pemodelan analisis deret waktu.

Metode ARIMA ini terdiri atas dua metode yang digabung menjadi satu, yaitu AR (*Autoregressive*) dan MA (*Moving Average*), dimisalkan ARIMA (2, 1, 2) atau ARIMA (1, 0, 1) dan sebagainya. Angka pertama menunjukkan derajat AR, angka kedua menunjukkan derajat integrasi dan angka ketiga menunjukkan derajat MA. Model ARIMA ini dalam pengaplikasianya sering ditulis dengan ARIMA ( $p, d, q$ ) yang keterangan  $p, d, q$  berturut-turut sama dengan keterangan sebelumnya. Metode ini menggunakan pendekatan interatif dalam identifikasi terhadap suatu model yang ada (Dalimunthe, 2017).

Peramalan (*forecasting*) dengan menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*) merupakan salah satu dari banyaknya model peramalan data yang ada. Meskipun pendekatan dengan metode ARIMA dilihat secara teoritis dan statistik sangat menarik, akan tetapi kerumitan mereka menghalangi pemakaian secara luas sebagai dasar untuk peramalan di dalam organisasi-organisasi. Kelemahan dari metode ARIMA adalah output yang dihasilkan olehnya memiliki nilai *error* yang relatif tinggi sehingga diperlukan metode lain untuk memperbaiki nilai error pada hasil outputnya (Aizzah *et al.* 2021). Persamaan metode ARIMA:

$$\phi(B) (1 - B)^d Y_t = \theta(B) \varepsilon_t \quad [2.10]$$

keterangan:

$Y_t$  = nilai data pada waktu ke- $t$

$B$  = operator *backshift*

$d$  = orde diferensiasi

$\phi(B)$  = polinom Autoregressive (AR)

$\theta(B)$  = polinom Moving Average (MA)

$\varepsilon_t$  = *error* acak (*white noise*)

### 2.1.12 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA) merupakan pengembangan metode ARIMA yang memiliki efek musiman. Kelebihan dari metode ini dapat menerima semua jenis pola *time series* meskipun dalam prosesnya harus distasionerkan terlebih dahulu. SARIMA menggunakan nilai pada masa lalu dan masa sekarang dari variabel dependen untuk menghasilkan peramalan yang akurat (Dimashanti & Sugiman, 2021). Metode SARIMA merupakan model yang sangat kuat dan fleksibel diantara model analisis *time series* dan *forecasting*.

Data yang digunakan dalam penelitian termasuk data dengan pola musiman dengan kecenderungan mengulangi pola tingkah gerak dalam periode tertentu, biasanya satu tahun untuk data bulanan, maka model ARIMA yang dibutuhkan adalah model *Seasonal Autoregressive Integrated Moving Average* (SARIMA) (Khaulasari *et al.* 2021). SARIMA dinotasikan  $(p, d, q)(P, D, Q)^s$  bagian yang dinotasikan dengan huruf kecil dan huruf kapital, pada  $(p, d, q)$  adalah notasi dari bagian dari model yang tidak musiman. Sedangkan  $(P, D, Q)$  adalah notasi dari bagian model yang musiman. Pangkat  $s$  adalah notasi untuk jumlah periode permusim yang akan dihitung (Christie *et al.* 2022).

Kelebihan dari metode ini dapat menerima semua jenis pola *time series* meskipun dalam prosesnya harus distasionerkan terlebih dahulu. SARIMA menggunakan nilai pada masa lalu dan masa sekarang dari variabel dependen untuk menghasilkan peramalan yang akurat (Yuhana & Aulia, 2018). Kelebihan lain dari metode SARIMA adalah relatif mudah untuk digunakan dalam memprediksi data runtun waktu berpolanya musiman. Meskipun model SARIMA merupakan metode statistik yang andal untuk meramalkan data deret waktu musiman, model ini memiliki sejumlah keterbatasan yang penting untuk diperhatikan dalam konteks data modern yang kompleks. Salah satu kelemahan utama SARIMA adalah asumsi linearitas, di mana model ini mengasumsikan bahwa hubungan antara nilai masa lalu dan nilai saat ini bersifat linier. Selain itu, SARIMA menuntut data yang stasioner, keterbatasan lainnya adalah bahwa pemilihan parameter model SARIMA  $(p, d, q, P, D, Q, s)$  membutuhkan penyesuaian manual yang rumit dan rentan terhadap overfitting apabila tidak dilakukan secara hati-hati.

Langkah-langkah pemodelan dari SARIMA adalah yang pertama menstasioneritaskan data yang belum stasioner, kemudian mengidentifikasi data secara umum, selanjutnya melakukan identifikasi model sementara, lalu mengestimasi parameter yang ada, memeriksa diagnostik serta memilih model terbaik, kemudian selanjutnya dapat dilakukan peramalan (Utomo & Fanani, 2020). Perbedaan metode ARIMA dan SARIMA, metode ARIMA digunakan untuk data deret waktu non-musiman. Mempunyai tiga komponen utama: AR (*AutoRegressive*), I (*Integrated*), MA (*Moving Average*). Metode SARIMA perluasan dari ARIMA yang menangani komponen musiman dalam data, cocok untuk data dengan pola berulang secara periodik (misalnya per bulan, kuartal), menambahkan komponen musiman (*seasonal* AR, I, MA) ke dalam model ARIMA. Terdapat persamaan model SARIMA sebagai berikut:

$$\Phi_p(B^s) \Phi_p(B) (1 - B)^d (1 - B^s)^D Y_t = \theta_Q(B_S) \theta_q(B) \varepsilon_t \quad [2.11]$$

Keterangan:

$Y_t$  = nilai data pada waktu ke- $t$

$B$  = operator *backshift*

$d$  = orde diferensiasi nonmusiman

$D$  = orde diferensiasi musiman

$s$  = periode musiman

$\Phi_p(B^s)$  = polinom AR musiman

$\Phi_p(B)$  = polinom AR nonmusiman

$\theta_Q(B_S)$  = polinom MA musiman

$\theta_q(B)$  = polinom MA nonmusiman

$\varepsilon_t$  = *error acak (white noise)*

### 2.1.13 Stasioner

Stasioner adalah keadaan *mean* dan *varians* adalah konstan (Farosanti *et al.* 2022). Menggunakan data yang telah distasionerkan, model *time series* dapat dikatakan lebih stabil. Apabila data yang digunakan dalam model ada yang tidak stasioner, maka data tersebut dipertimbangkan kembali validitas dan kestabilannya, karena hasil regresi yang berasal dari data yang tidak stasioner akan menyebabkan *spurious regression*. *Spurious regression* adalah regresi yang memiliki  $R^2$  yang tinggi, namun tidak ada hubungan yang berarti dari keduanya. Kestasioneran suatu

data dapat dilihat dari dua hal yaitu stasioner dalam *mean* (rata-rata) dan stasioner dalam *varians* (Purnomo, 2019).

### 2.1.13.1 Uji *Augmented Dickey-Fuller* (ADF)

Uji *Augmented Dickey-Fuller* (ADF) untuk menguji stasioneritas secara statistik. Uji ADF digunakan untuk melakukan uji stasioner pada data yang sudah stasioner dalam ragam. Uji ini didasarkan pada pengujian hipotesis untuk mengevaluasi kestasioneran data deret waktu. Hipotesis pada uji ADF adalah (Wulandari & Yurinanda, 2021):

$H_0: \gamma = 0$  (Terdapat unit *root*, yang berarti data tidak stasioner pada rataannya)

$H_1: \gamma \neq 0$  (Tidak terdapat unit *root*, yang berarti data stasioner pada rataannya)

Tolak  $H_0$  jika  $p\text{-value} < \alpha=10\%$ .

Statistik uji:

$$t_{hitung} = \frac{Y}{sd(Y)} \quad [2.12]$$

Keterangan:

$t_{hitung}$  = nilai statistik uji *t*

$Y$  = nilai parameter

$sd(Y)$  = standar deviasi dari *Y*

Apabila data tidak stasioner terhadap rata-rata, maka perlu dilakukan proses pembeda (*differencing*) pada data asli agar menjadi stasioner.

*Differencing* adalah teknik dalam analisis deret waktu (*time series*) untuk mengubah data yang tidak stasioner (mengandung tren atau musiman) menjadi stasioner. Stasioner artinya pola data tidak berubah seiring waktu, rata-rata, varians, dan struktur korelasinya konstan. Proses pembedaan dilakukan jika data tidak stasioner dalam rata-rata. Pembedaan dapat dilakukan untuk beberapa periode sampai data stasioner. Proses pembedaan dilakukan dengan cara mengurangkan suatu data dengan data sebelumnya. Notasi *B* (operator *backshift*) digunakan dalam proses pembedaan.

$$Z_t^{(d)} = (1 - B)^d Z_t \quad [2.13]$$

Keterangan:

$Z_t^{(d)}$  = data setelah diferensiasi orde ke-*d*

$Z_t$  = data asli pada waktu ke- $t$

$B$  = operator backshift

$d$  = orde diferensiasi

### 2.1.13.2 Uji Box-Cox

Uji *Box-Cox* adalah metode transformasi data yang bertujuan untuk: Menstabilkan varians (kalau data makin besar, fluktuasinya makin besar juga). Membuat distribusi data mendekati normal, yang penting buat banyak analisis statistik dan pemodelan. *Box-Cox* adalah salah satu metode untuk proses stasioneritas data dalam varians yang dikenalkan oleh Box dan Cox (Julistio & Mahdy, 2024).

Transformasi *Box-Cox* adalah salah satu metode untuk proses stasioneritas data dalam varians yang dikenalkan oleh Box dan Cox. Transformasi *Box-Cox* juga sering disebut dengan transformasi kuasa. Secara matematis, transformasi *Box-Cox* dirumuskan sebagai berikut:

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda = 1 \\ \ln(Y_t), & \lambda \neq 1 \end{cases} \quad [2.14]$$

Keterangan:

$Y_t^{(\lambda)}$  = data hasil transformasi Box–Cox

$Y_t$  = data asli pada waktu ke- $t$

$\lambda$  = parameter transformasi

Notasi  $\lambda$  melambangkan parameter transformasi. Setiap nilai  $\lambda$  mempunyai rumus transformasi yang berbeda. Transformasi dilakukan jika belum diperoleh nilai  $\lambda=1$  yang artinya data telah stasioner dalam varians

### 2.1.14 Autocorrelative Function (ACF)

ACF adalah perhitungan autokorelasi berdasarkan varians dan kovariansnya. Autokorelasi merupakan cara untuk melihat adanya hubungan atau korelasi pada data yang sama antar waktu. *Varians* didefinisikan sebagai hubungan data pada waktu yang sama. Sedangkan *kovarians* didefinisikan sebagai hubungan data antar waktu. Kunci dari *time series* sendiri adalah terdapat pada koefisien korelasi. Ukuran korelasi pada time series dengan dirinya ada pada lag atau biasa dikenal dengan selisih waktu 0, 1, 2, sampai dengan lag pada data. Sedangkan plot

ACF berfungsi untuk mengenali kestasioneran data. Berikut persamaannya (Nasir, 2015).

$$\hat{p}_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad [2.15]$$

Keterangan:

$\hat{p}_k$  = koefisien autokorelasi pada lag ke-k

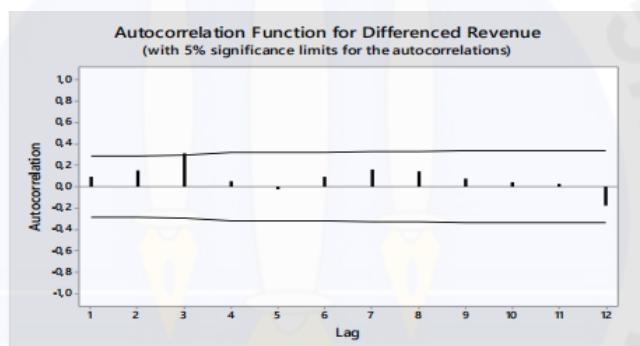
$Y_t$  = nilai data pada waktu ke- $t$

$\bar{Y}$  = rata-rata data

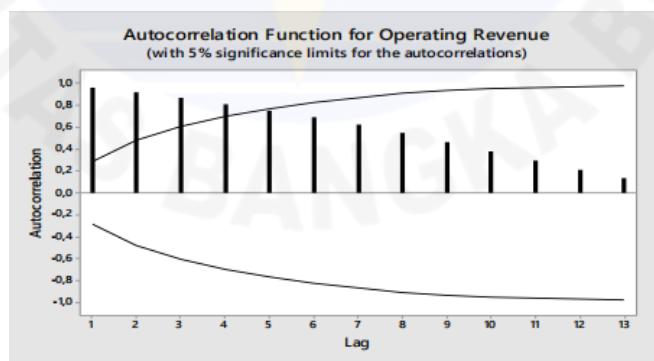
$n$  = jumlah data

$k$  = lag ke-k

Diagram ACF dapat digunakan sebagai alat untuk mengidentifikasi kesetasioneran data. Diagram ACF yang cendrung turun lambat atau turun secara linear, maka dapat disimpulkan data belum stasioner dalam rata-rata (Laga *et al.* 2006).



**Gambar 2.10.** Plot ACF data yang stasioner (Kataba, 2021)



**Gambar 2.11.** Plot ACF data yang belum stasioner (Kataba, 2021)

Jika data mengandung tren atau tidak stasioner, seringkali ACF tidak menunjukkan *cut off* tetapi justru menurun secara perlahan atau tetap tinggi di banyak lag.

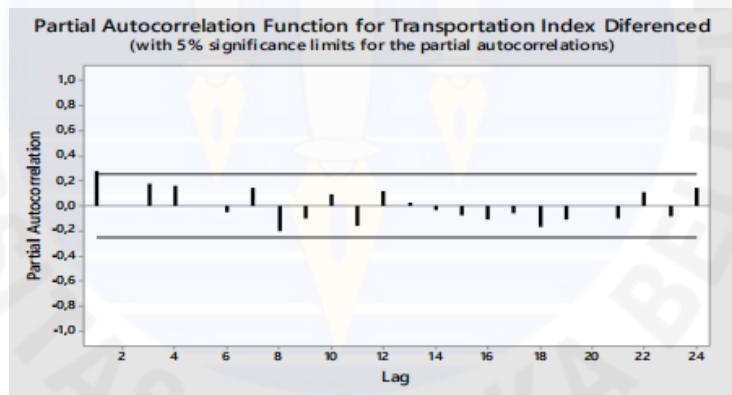
### 2.1.15 Partial Autocorrelative Function (PACF)

*Partial Autocorrelative Function* (PACF) adalah suatu persamaan yang berfungsi sebagai penghitung ukuran kekuatan dari variabel  $Y_t$  dan  $Y_{t+k}$ . Perbedaan PACF dan ACF adalah terletak pada model partialnya. Pada PACF, pengawalan perhitungan nilai dimulai dengan  $\hat{\Phi}_{kk}=\hat{p}_1$ , dimana  $\hat{p}_1$  adalah nilai dari autocorrelative lag pertama. Berikut ini adalah persamaan untuk menghitung nilai *partial autocorrelative function* lag ke- $k$  dengan menentukan hasil  $\hat{\Phi}_{kk}$  (Ilmiah, 2018).

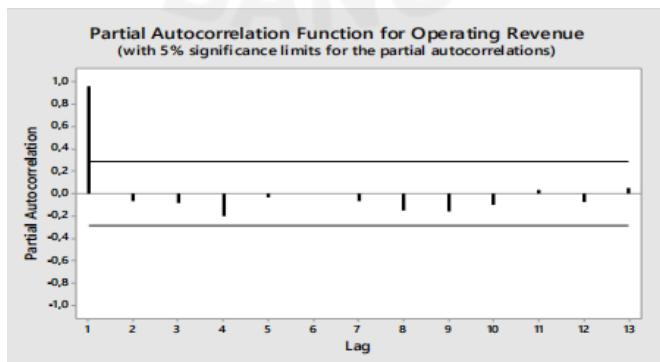
$$\hat{\Phi}_{kk} = \frac{\hat{p}_k - \sum_{j=1}^{k-1} \phi_{k-1,j} p_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} p_{k-j}} \quad [2.16]$$

Keterangan:

- $\hat{\Phi}_{kk}$  = koefisien PACF pada lag ke- $k$
- $\hat{p}_k$  = koefisien ACF pada lag ke- $k$
- $\phi_{k-1,j}$  = koefisien PACF orde sebelumnya
- $n$  = jumlah data
- $k$  = lag ke- $k$



Gambar 2.12. Plot PACF data yang stasioner (Kataba, 2021)



Gambar 2.13. Plot PACF data yang belum stasioner (Kataba, 2021)

Dalam data yang belum stasioner, *cut off* dalam ACF atau PACF dapat digunakan dalam mendeteksi apakah data masih mengandung tren atau sudah stasioner setelah *differencing*.

### 2.1.16 Prediksi

Prediksi adalah proses untuk meramalkan suatu variable di masa mendatang dengan berdasarkan pertimbangan data pada masa lampau, data yang sering digunakan untuk melakukan prediksi adalah berupa data kuantitatif (Novalyn *et al.* 2018). Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Adiguno *et al.* 2022). Prediksi memiliki manfaat yang luas dalam berbagai bidang kehidupan, seperti ekonomi, kesehatan, teknik dan lingkungan, serta di lembaga tertentu memungkinkan pengambilan keputusan atau kebijakan. Dengan prediksi dapat merencanakan masa depan lebih terarah, sumber daya dapat digunakan dengan lebih efisien dan juga dapat mengurangi resiko.

Peramalan dapat digunakan untuk membantu pengambilan keputusan yang lebih baik dengan memprediksi hasil yang mungkin terjadi di masa depan. Peramalan merupakan suatu kegiatan memperkirakan atau memprediksikan kejadian dimasa yang akan datang tentunya dengan bantuan penyusunan rencana terlebih dahulu, dimana rencana ini dibuat berdasarkan kapasitas dan kemampuan permintaan/produksi yang telah dilakukan perusahaan (Sihaloho *et al.* 2020).

## 2.2 Penelitian Terdahulu

Penelitian terdahulu adalah sumber lampau dari hasil penelitian yang nantinya dapat digunakan untuk membandingkan penelitian sehingga pelaksanaan penelitian kedepannya dapat terbantu (Afdal, 2024). Penelitian terdahulu mencantumkan berbagai hasil penelitian terdahulu yang berkaitan dengan penelitian yang akan dilakukan oleh penulis. Penelitian terdahulu selain berfungsi sebagai sumber inspirasi, juga berfungsi untuk memperkuat teori dan fenomena hubungan atau pengaruh antar *variable* untuk penelitian selanjutnya. Adapun penelitian terdahulu yang berkaitan dengan penelitian ini dapat dilihat pada Tabel 2.2.

**Tabel 2.2** Penelitian Terdahulu

No	Judul Penelitian	Nama Peneliti (tahun)	Hasil penelitian
1	Perbandingan Metode <i>Seasonal Autoregressive Integrated Moving Average</i> (SARIMA) dengan <i>Support Vector Regression</i> (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Mancanegara ke Bali	(Nanik hendayani & Nurhidayati, 2020)	Berdasarkan hasil yang diperoleh diketahui bahwa nilai MAPE, MSE, dan MAE yang dihasilkan pada model SARIMA lebih kecil dibandingkan dengan model SVR, peningkatan paling besar pada bulan Juli 2019 dengan jumlah kunjungan sebesar 647.230 dan serta kunjungan paling kecil terjadi pada bulan Januari 2019 sebesar 472.637 kunjungan.
2	Analisis Prediksi Harga Saham Pt. Astra International Tbk menggunakan metode <i>Autoregressive Integrated Moving Average</i> (ARIMA) dengan <i>Support Vector Regression</i> (SVR)	(Putra & Kurniawan, 2021)	Hasil penelitian menunjukkan, berdasarkan hasil perhitungan Evaluasi nilai kesalahan menggunakan <i>Root Mean Square Error</i> (RMSE), metode Arima unggul sebanyak 6 kali, sedangkan untuk metode SVR unggul sebanyak 7 kali
3	Peramalan Nilai Tukar Petani Subsektor Tanaman Pangan di Nusa Tenggara Timur Menggunakan Metode SARIMA	(Muzaki & Agustina, 2024)	Hasil peramalan ini menunjukkan di Peramalan NTPP dengan metode SARIMA diperoleh model terpilih ARIMA (01,0) (0,1,1)12 dengan ukuran kesalahan RMSE dan MAPE berturut-turut 2,5549 dan 2,3553 persen. Hasil ramalan NTPP periode selanjutnya Juli 2024–Juni 2025, diperoleh rata-rata nilai NTPP sebesar 9795.
4	Perbandingan Metode ARIMA dan SARIMA Dalam Peramalan Jumlah Penumpang Bandara Provinsi Kepulauan Bangka Belitung	(Febiola <i>et al.</i> 2024)	Metode terbaik untuk meramalkan jumlah penumpang di Bandara Provinsi Kepulauan Bangka Belitung adalah ARIMA dengan model terbaiknya yakni ARIMA (0,1,1) dan MAPE sebesar 9,41 persen. Peramalan jumlah penumpang di Bandara Provinsi Kepulauan Bangka

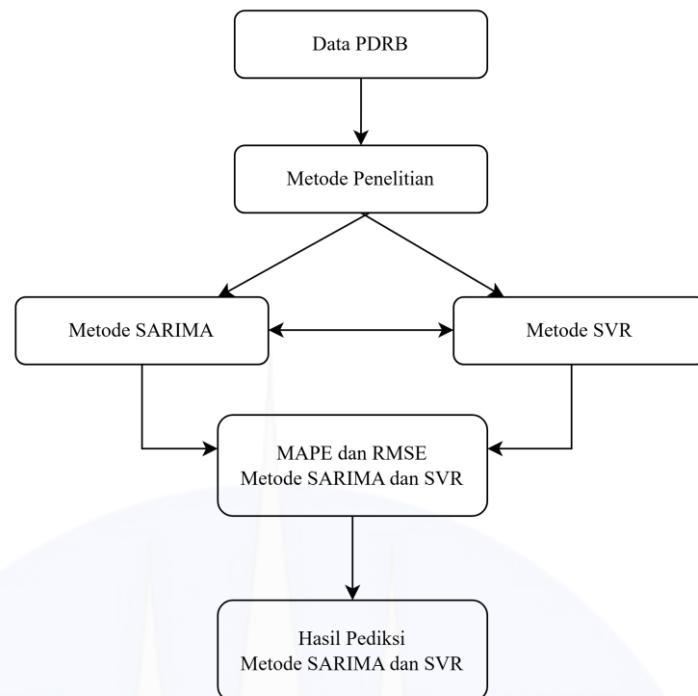
			Belitung dari bulan Mei 2024 sampai dengan bulan April 2026 terindikasi mengalami tren naik.
5	Aplikasi Model ARIMA dalam Peramalan Produk Domestik Regional Bruto (PDRB) Pertambangan dan Penggalian Kalimantan Selatan	(Hayati & Agustina, 2024)	Model terbaik untuk meramalkan nilai PDRB di Kalimantan Selatan adalah ARIMA (2,1,4) dengan tingkat akurasi MAPE 5,054 persen dan termasuk dalam kategori peramalan yang sangat akurat. Hasil peramalan nilai PDRB triwulan II tahun 2024 - triwulan IV tahun 2025 dengan model ARIMA (2,1,4) terus meningkat pada tahun 2024 - 2025.
6	Prediksi Penggunaan Bahan Bakar pada PLTGU menggunakan Metode Support Vector Regression (SVR)	(Novianti <i>et al.</i> 2022)	hasil prediksi dengan menggunakan kernel polynomial memperoleh hasil yang sangat baik dengan nilai $\epsilon$ yaitu 0.0266, nilai $b$ yaitu 0.0285 dan diperoleh MAPE sebesar 7.7513%. Hasil prediksi ini dikatakan sangat baik.
7	Pemodelan Matematis untuk Peramalan PDRB Atas Dasar Harga Konstan Di Sumatera Utara dengan Metode Polinomial dan Eksponensial	(Hutapea & Girsang, 2024)	Nilai R squared hasil nilai peramalan PDRB Sumatera Utara dengan regresi eksponensial sebesar 1,00 dan nilai MSE sebesar 0,809, hal ini menunjukkan bahwa tingkat akurasi perkiraan dari regresi eksponensial yang baik.
8	Peramalan Data Produk Domestik Regional Bruto Daerah Istimewa Yogyakarta Menggunakan Kombinasi Fuzzy Time Series Chen Dengan Particle Swarm Optimization	(Usmia & Noeryanti, 2021)	Hasil peramalan Fuzzy Time Series PDRB DIY diperoleh hasil prediksi sebesar 26167334.76 juta rupiah. Nilai MAPE dari metode FTS Chen adalah 4.33% dengan nilai RMSE sebesar 1092876.32. Sedangkan nilai MAPE dari metode FTS dengan PSO adalah 3.40% dengan nilai RMSE sebesar 875049.7
9	Peramalan Jumlah Penumpang Kereta Api di Indonesia Menggunakan	(Utomo & Panani, 2020)	Didapatkan hasil peramalan untuk periode 2020 dengan jumlah keseluruhan 36.941.500 orang penumpang. Perhitungan

	Metode <i>Seasonal Autoregressive Integrated Moving Average (SARIMA)</i>	error didapatkan MSE sebesar 0,046875, nilai MSE dibawah 1. Didukung dengan nilai MAPE sebesar 6,26%,
10	Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan <i>Support Vector Regression</i> (SVR)	Model SVR terbaik yang diperoleh adalah model SVR menggunakan kernel <i>Radial Basic Function</i> (RBF) dengan parameternya adalah $C = 2^{-1.5}$ , $y = 2^{-1.5}$ dan $\varepsilon = 0,03$ . Hasil menunjukkan kriteria peramalan yang akurat dilihat dari nilai MAPE data <i>training</i> sebesar 7,85 persen dan nilai MAPE data <i>testing</i> sebesar 18,24 persen.

Penelitian di atas, dapat disimpulkan bahwa SARIMA dan SVR memiliki keunggulan yang berbeda tergantung pada jenis data yang digunakan. Evaluasi berdasarkan nilai *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE) menunjukkan perbedaan berikut: SARIMA lebih unggul dalam data dengan pola musiman dan tren, SARIMA lebih cocok untuk data dengan pola musiman dan tren yang jelas karena memiliki MAPE dan RMSE lebih kecil dalam kasus wisatawan, transportasi. Jika tren musiman kuat, SARIMA lebih direkomendasikan. SVR lebih baik dalam data yang kompleks dan Non-linear dan pertanian. SVR lebih fleksibel dalam menangani data non-linear seperti harga saham dan konsumsi bahan bakar, tetapi sering memiliki MAPE lebih besar daripada SARIMA. Secara keseluruhan, pemilihan metode terbaik bergantung pada karakteristik data dan tujuan peramalan., tetapi jika data kompleks dan non-linear, SVR bisa menjadi alternatif yang lebih baik dengan tuning parameter yang tepat.

### 2.3 Kerangka Pemikiran

Kerangka pemikiran adalah landasan penelitian, sebagai pijakan peneliti agar penelitian menjadi kokoh dan memiliki landasan yang kuat sehingga penelitian tersebut dapat diandalkan (Kurniawati, 2021). Kerangka pemikiran dalam Analisis Data PDRB Menurut Lapangan Usaha Provinsi Kepulauan Bangka Belitung menggunakan metode SARIMA dan SVR ditunjukkan pada Gambar 2.14.



**Gambar 2.14.** Kerangka Pemikiran

PDRB memiliki peran penting dalam meningkatkan pertumbuhan ekonomi suatu daerah, dimana semakin tinggi PDRB maka dapat dikatakan bahwa pertumbuhan ekonominya juga tinggi (Putri, 2020). Pertumbuhan ekonomi tidak selalu diikuti oleh pertumbuhan pendapatan per kapita, melainkan sebagai peningkatan pendapatan nasional bruto yang tidak memiliki pengaruh terhadap besar atau kecilnya laju pertumbuhan penduduk (Rustam, 2022). PDRB menurut lapangan usaha dirinci menurut total nilai tambah dari seluruh lapangan usaha yang mencakup kategori pertanian, kehutanan, dan perikanan, pertambangan dan penggalian, industri pengolahan, pengadaan informasi dan komunikasi, jasa keuangan dan asuransi, real estat, jasa perusahaan, administrasi pemerintahan, pertahanan, dan jaminan sosial wajib, jasa pendidikan; jasa kesehatan dan kegiatan sosial, dan jasa lainnya (BPS, 2023).

Metode yang digunakan untuk menganalisis data PDRB Provinsi Kepulauan Bangka Belitung adalah metode SVR dan SARIMA. Membandingkan kedua metode tersebut untuk menentukan metode terbaik dalam melakukan prediksi yang dapat dilihat dari tingkat akurasinya yaitu hasil prediksi MAPE dan

RMSE. Semakin kecil tingkat akurasi MAPE maka prediksi dapat dikatakan sangat baik, begitu juga dengan nilai RMSE

*Support Vector Regression* (SVR) merupakan bagian dari *Support Vector Machine* (SVM) yang digunakan untuk kasus regresi dan prediksi SVR mampu mengatasi *overfitting*. Model SARIMA mempunyai akurasi yang lebih tinggi dalam memprediksi pola musiman dibandingkan dengan ARIMA (Siregar, 2022).

#### **2.4 Hipotesis Penelitian**

Hipotesis penelitian adalah pernyataan yang merupakan jawaban sementara terhadap rumusan masalah yang sifatnya menduga tetapi didasari oleh teori-teori atau temuan terdahulu (Zaki & Saiman, 2021). Menurut Dantes (2012) hipotesis adalah praduga yang harus diuji melalui data yang didapat dengan penelitian. Hipotesis yang diajukan dalam penelitian ini adalah sebagai berikut.

$H_0$ : Metode SARIMA tidak lebih baik dari metode SVR k untuk memprediksi data PDRB Provinsi Kepulauan Bangka Belitung.

$H_1$ : Metode SARIMA lebih baik dari metode SVR untuk memprediksi data PDRB Provinsi Kepulauan Bangka Belitung.