SKRIPSI

Metode SVR Untuk Memprediksi Jumlah Pemakaian Air PDAM (Studi Kasus : Sub Unit Muntilan)

Jenis Skripsi:
Penelitian Eksperimental

Disusun Sebagai Salah Satu Syarat Memperoleh Gelar Sarjana Komputer (S. Kom.)

Disusun oleh:
Windi Wulandari
NIM. 20.0504.0030

Pembimbing:
Mukhtar Hanafi, S. T., M. Cs.
NIDN. 0602047502

Pembimbing:
Rofi Abul Hasani, S. Kom., M. Eng
NIDN. 0631079101

Program Studi Teknik Informatika
Fakultas Teknik
Universitas Muhammadiyah Magelang
Tahun 2025

Bab 1 Pendahuluan

1.1 Latar Belakang Masalah

Seiring perkembangan waktu dan kemajuan teknologi informasi mengubah cara hidup dan mempermudah dalam berbagai jenis pekerjaan manusia khususnya dalam penggunaan teknologi komputer seperti pengolahan data yang dapat mempengaruhi perusahaan dan memberikan manfaat kepada masyarakat (Nurul Noviyana & Muhammad Irwan Padli Nasution, 2024) . Penggunaan teknologi ini memiliki keterkaitan erat dalam perkembangan sebuah perusahaan yang memainkan peran besar dalam proses pengolahan data. Proses pengolahan data ditujukan untuk menganalisis sebuah data. Analisis ini akan berguna untuk memprediksi masalah yang mungkin timbul dalam perusahaan atau komunitas. Salah satu cara untuk menyelesaikan atau memperkirakan suatu masalah adalah dengan mengelola air bersih. Perusahaan Daerah Air Minum (PDAM) merupakan sebuah Perusahaan yang bergerak dalam bidang distribusi air minum di daerah kabupaten dan kota di seluruh Indonesia. Salah satunya yang terdapat di Kabupaten Magelang, Jawa Tengah, tepatnya yang berada di Sub Unit Muntilan. Perusahaan ini menyediakan air bersih dan mendistribusikannya ke rumah-rumah warga yang telah terdaftar di PDAM. PDAM Kabupaten Magelang didirikan berdasarkan peraturan daerah kabupaten daerah tingkat II Magelang nomor 11 tahun 1980. PDAM saat ini memiliki 19 sumber mata air, salah satunya terletak di Muntilan, yaitu Sumber Mata Air Semaren.

Air merupakan sumber daya alam yang memiliki fungsi penting dan tidak dapat lepas dari kehidupan manusia. Dengan semakin cepatnya pertumbuhan populasi, kemajuan pembangunan, dan biaya hidup, permintaan terhadap sumber daya air meningkat pesat baik dalam hal kualitas maupun kuantitas, sehingga pasokan air seringkali mengalami masalah (Agustina et al., 2023). Saat ini di Indonesia hal ini bisa teratasi dengan adanya Perusahaan PDAM. Sub Unit Muntilan yang merupakan bagian dari PDAM Kabupaten Magelang sering kali menemui kendala dalam mendistribusikan maupun mengelola pasokan air yang dipengaruhi oleh berbagai faktor, seperti demografi, musim, dan aktivitas masyarakat yang mengakibatkan distribusi ini terhalang dan tidak sejalan antara ketersediaan dan kebutuhan terkait kelangkaan air atau pemborosan air. Karena itu, pengelolaan analisis data harus dilakukan untuk menghindari ketidakefisienan dalam produksi dan distribusi.

Pengelolaan data merupakan salah satu usaha yang dapat dilakukan dengan memanfaatkan teknologi komputer saat ini. Pengelolaan analisis data dapat dilakukan untuk mencegah hal-hal yang tidak diinginkan terjadi seperti masalah diatas. Melalui pemrosesan data dan penerapan metode SVR, dilakukan prediksi penggunaan air di masa depan dengan memanfaatkan data historis pengeluaran air. Memprediksi air menggunakan metode SVR bertujuan sebagai salah satu penanggulangan sebelum terjadinya hal-hal yang tidak diinginkan.

Metode *Support Vector Regression* (SVR) adalah pengembangan dari *metode Support Vector Machine* (SVM). Metode ini sangat efisien untuk data non linier dengan memanfaatkan parameter utama RBF untuk menghitung manual dan *Grid Search* untuk menemukan parameter yang memiliki akurasi lebih tinggi dalam memprediksi suatu data (Ruliana et al., 2024). Algoritma *search grid* digunakan untuk menentukan parameter terbaik model SVR (Safira et al., 2023).

Berdasarkan latar belakang tersebut, dipilih metode SVR dengan menerapkan *Grid Search* untuk menentukan nilai parameter utama dalam memprediksi data PDAM Kabupaten Magelang Sub Unit Muntilan berdasarkan data jumlah air yang telah digunakan di masa lampau dari tahun 2020-2024 yang kemudian diproyeksikan ke masa depan melalui pemodelan sistematis. Oleh karena itu, hasil prediksi ini diharapkan dapat mencapai nilai train yang optimal serta memiliki tingkat akurasi yang cukup tinggi. Nilai akurasi diukur dengan melalui beberapa perhitungan, yaitu MSE, RMSE, MAPE, dan R^2.

Mean Square Eror (MSE), *Root Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R-Squared* (R^2). Perhitungan MSE dilakukan untuk menentukan rata-rata dari kuadrat selisih hasil prediksi. RMSE adalah akar kuadrat dari MSE, yang berfungsi untuk menghitung nilai prediksi model dengan nilai sebenarnya (Hikmah et al., 2023). MAPE digunakan untuk menghitung rata-rata selisih kuadrat antara nilai yang diprediksi, sedangkan R^2 dipakai untuk mengukur keterkaitan antara variabel independen (Najla et al., n.d.). Dengan memadukan berbagai perhitungan akurasi untuk mengukur beberapa tingkat model dan mengkombinasikan teknik sliding window untuk menemukan proyeksi data yang akan datang.

Teknik *sliding window* biasa digunakan untuk memprediksi sebuah data membentuk pola data sebagai variable input dan variabel target dengan mengkombinasikan ukuran jendela untuk mendapatkan hasil prediksi (Kartini et al., 2021).. Teknik ini dipilih karena mayoritas penelitian peramalan deret waktu umumnya menggunakan metode ini untuk menghasilkan data ramalan di masa depan (Radho et al., 2021).

Melalui pengelolaan analisis data yang menggabungkan metode, parameter, dan teknik dalam mencari hasil prediksi serta tingkat akurasi, diharapkan hasil tersebut mampu mendukung perusahaan PDAM untuk meningkatkan kinerjanya perusahaan dan meminimalisir terjadinya hal-hal yang tidak diinginkan, seperti terlepasnya sambungan pipa ketika mata air meluap ataupun ketika terjadinya kekeringan pada saat musim kemarau tiba. Serta diharapkan dapat memberikan solusi bagi masyarakat yang kerap mengeluhkan masalah terkait gangguan air yang sering muncul.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan diatas, berikut ini masalah penelitian dirumuskan adalah bagaimana metode SVR diterapkan dalam memprediksi data kebutuhan penggunaan jumlah air mendatang di PDAM Kabupaten Magelang Sub Unit Muntilan?

1.3 Tujuan

Berdasarkan rumusan masalah, tujuan penelitian yang akan dicapai adalah mengetahui cara kerja metode SVR dalam memprediksi jumlah kebutuhan air PDAM Kabupaten Magelang Sub Unit Muntilan yang akan datang berdasarkan data jumlah air yang telah digunakan mulai dari tahun 2020 sampai tahun 2024 serta mengetahui tingkat akurasi yang dihasilkan.

1.4 Manfaat

Berdasarkan tujuan penelitian yang telah disebutkan diatas, hasil dari penelitian ini adalah dapat memberikan manfaat mengenai penerapan metode SVR dalam menganalisis data di PDAM Kabupaten Magelang Sub Unit Muntilan dapat meningkatkan pertumbuhan perusahaan dalam menangani masalah yang akan datang, memungkinkan perusahaan dapat menanggulangi masalah mengenai pengelolaan air yang didistribusikan ke masyarakat. Dengan demikian, penelitian mengenai prediksi air yang akan datang menggunakan data yang telah digunakan sebelumnya ini dapat berkontribusi dalam pengembangan perusahaan dan memecahkan masalah masyarakat yang mengeluhkan sumber daya air.

Bab 2 Studi Literatur

2.1 Tinjauan Penelitian Terdahulu

Berikut ini adalah beberapa hasil penelitian sebelumnya yang relevan atau berhubungan dengan penelitian yang akan dilakukan oleh peneliti:

Menurut penelitian (Sepri et al., 2020), dengan judul "Prediksi Harga Cabai Merah Menggunakan *Support Vector Regression*" menyatakan bahwa prediksi SVR terhadap data waktu (*Time Series*) sangat efektif dalam melakukan prediksi terhadap harga cabai merah nasional. Data yang digunakan dalam penelitian ini mulai dari periode Januari 2017 sampai dengan Desember 2021. Keseluruhan data dibagi dua dengan pembagian sebanyak 80% data sebagai data latih dan 20% dari keseluruhan merupakan data uji, dengan memasukan parameter *hyperplane* SVR. Parameter *hyperplane* dilakukan dengan menggunakan *Grid Search* dengan *Kernel* RBF. Dengan Evaluasi prediksi menggunakan *Mean Absolute Percentage Error* (MAPE).

Penelitian ini memiliki parameter terbaik yang menunjukkan MAPE prediksi untuk data uji lebih besar dibandingkan MAPE prediksi data latih, dengan perbandingan MAPE data latih = 4,07% dan MAPE data uji = 9,11%, serta hasil turning parameter *hyperplane* dengan menggunakan algoritma *Grid Search* memiliki parameter terbaik, dengan C=1000, ε=5 dan γ =1. Namun, pemodelan ini memiliki nilai MAPE atau error yang cukup tinggi, oleh sebab itu hasil dari pemodelan ini kurang memuaskan.

Selanjutnya oleh (Akrom at al., n.d.) dengan judul "Prediksi Jumlah Air PDAM Menggunakan Metode ANN dengan Optimasi PSO" penelitian ini menyatakan bahwa prediksi jumlah air PDAM menggunakan ANN efektif dalam melakukan prediksi jumlah air. Data yang digunakan dalam penelitian ini antara lain jumlah penduduk, jumlah pelanggan, volume produksi, volume distribusi, air terjual dan kehilangan air. Data ini diperoleh selama 6 tahun terakhir yaitu mulai tahun 2008 hingga 2013. Serta menggunakan optimasi *Particle Swarm Optimization*. Jika dibandingkan dengan *neural network* biasa metode ini lebih bagus, sebab nilai RMSE yang dihasilkan sebesar 3,797 sedangkan nilai RMSE *neural network* biasa sebesar 4,943. Namun, nilai RMSE tersebut masih terbilang lebih besar.

Oleh (Thoriq et al., 2023.) dengan judul "Perkiraan Kebutuhan Air Bersih Menggunakan Metode Jaringan Syaraf Tiruan *Backpropagation*" dijelaskan bahwa prediksi menggunakan metode *Backpropagation*. Penelitian ini menggunakan data di PDAM pada masa lalu dengan rentang waktu 5 tahun mulai dari Januari 2018 hingga Januari 2023 yang terbagi atas 3 wilayah. Data ini berjumlah 60 data dimana 40 data digunakan untuk proses pelatihan dan 20 data untung mengecek keakuratan metode ini. Hasil prediksi pada pengujian ini terlihat bahwa hasilnya sangat baik, yaitu dengan 40 data latih menghasilkan 1,00 dengan tingkat kesalahan 0,00 dan data uji sebanyak 20 data mendapatkan hasil prediksi 0,92 dengan tingkat kesalahan 0,08.

Setelah meninjau beberapa penelitian diatas, bisa disimpulkan bahwa penelitian diatas membahas kasus serupa terkait prediksi data kebutuhan air PDAM, tetapi dengan menggunakan metode dan algoritma yang berbeda serta penelitian sejenis berupa prediksi, namun menggunakan objek penelitian yang berbeda. Kajian literatur yang telah dilakukan terhadap penelitian diatas menunjukkan bahwa metode SVR jarang diterapkan dalam memprediksi data kebutuhan air PDAM yang akan datang, maka penelitian ini tercipta sebagai salah satu pengembangan penelitian serupa, namun menggunakan metode yang berbeda serta mengkombinasikannya dengan beberapa teknik untuk mencari hasil prediksi dari jumlah kebutuhan air PDAM Kabupaten Magelang Sub Unit Muntilan pada tahun 2025.

2.2 Kajian Teoretis

2.2.1 *Machine Learning*

Machine Learning adalah aspek dari pemodelan *blackbox* atau sering dikenal sebagai basisdata, yang dibangun berdasarkan data historis dan berkaitan dengan matematika yang telah dikembangkan (BÎLBÎE & ZAHARIA, 2024). *Machine Learning* beroperasi melalui metode pemrograman komputer yang diberikan sejumlah dataset, yang kemudian akan dipelajari serta dilakukan proses train dan testing. Untuk mengklasifikasi polaritas dataset dengan menggunakan teknik fitur yang lebih kompleks (Hakim, 2021).

Machine learning merupakan rangkaian algoritma yang dipakai untuk meningkatkan kinerja sistem komputer dengan memanfaatkan data sampel yang digunakan untuk mencari data target yang akan dicari. *Machine learning* berkerja dengan teknik program komputer yang diberikan sekumpulan data latih atau data historis. Sistem akan secara otomatis belajar dari data tersebut untuk dapat mencapai suatu tujuan tertentu.

2.2.2 *Support Vector Regression*

Support Vector Regression (SVR) adalah teknik regresi yang didasarkan pada *Support Vector Machine* (SVM) yang digunakan untuk meramalkan nilai kontinu dengan mencari fungsi regresi yang meminimalkan kesalahan prediksi dan menjaga nilai *margin toleransi* atau *Epsilon* (ε) sekecil mungkin (Rais, 2022).

Model SVR merupakan model yang menghasilkan nilai parameter terbaik yang dihasilkan dengan melatih pemrograman komputer menggunakan data fitur (x) dan data target (y) yang digunakan. Data fitur (y) merupakan data variabel pendukung yang berkaitan dengan data target ataupun menggunakan data yang sama. Data target (y) merupakan data yang akan menjadi target hasil dari prediksi yang akan dilakukan.

Model SVR mempunyai rumus persamaan dalam menjelaskan cara menghitung deret waktu menggunakan data input yang berisi data historis metode regresi, proses prediksi ini dijabarkan kedalam sebuah fungsi (Saadah et al., n.d.).

Berikut rumus model SVR yang digunakan, adalah:

$$Yt = \omega t\emptyset(x_t) + b \quad \text{..... (2.1)}$$

Dimana:

a. Yt merupakan nilai prediksi *time series* (data target atau y),

b. xt merupakan adalah vektor input yang berisi data historis (data fitur atau x),

c. $xt = \{y_{t-D}, y_{t-D+1}, \dots, y_{t-1}\}^T$

d. $b \in R$ merupakan bias, $\omega \in R^M$ merupakan vektor bobot dari xt, dan $\emptyset(x_t) \rightarrow R^M(M > D)$ merupakan fungsi pemetaan fitur yang mengubah vektor input $xt \in R^D$ ke dalam ruang dengan dimensi yang lebih tinggi $\emptyset(x_t) \rightarrow R^M$.

Persamaan regresi linier dirumuskan sesuai dengan perumusan, sebagai berikut:

$$Yt = \sum_{(i=1,N)} \alpha_i(x_t - x_i)K(x_i, x_t) + b \quad \text{..... (2.2)}$$

Dimana:

a. $\alpha_i(i = 1, \dots, N)$) merupakan parameter *Lagrange* yang bernilai non negatif,

b. $K(x_t - x_i)$ merupakan fungsi kernel.

Proses persamaan regresi linier digunakan untuk menghitung parameter dengan menggunakan fungsi *kernel* yang akan digunakan dalam proses perhitungan SVR. Proses SVR mewajibkan proses perhitungan regresi linier. Model SVR tidak efektif jika data fitur (x) dan data target (y) tidak data linier.

Persamaan kernel yang sesuai dengan persamaan diatas:

$$K(x_t, x_i) = \exp(-\gamma||x_t - x_i||^2) \quad \text{..... (2.3)}$$

Proses perhitungan persamaan karnel merupakan proses perhitungan dalam SVR dengan menggunakan karnel RBF. Proses perhitungan karnel merupakan salah satu proses perhitungan yang diperlukan karena karnel merupakan penerjemahan atau proses perhitungan seberapa mirip kedua data dan memetakannya untuk dimanfaatkan dalam pembentukan model.

Dalam perhitungan karnel menggunakan *code*, model akan dikombinasikan dengan *grid search* dalam menentukan hasil parameter utama, sedangkan fungsi karnel RBF digunakan untuk melakukan prediksi nilai kontinu yang kemudian disesuaikan lewat *grid search*. *Grid search* adalah kombinasi dari parameter yang diuji untuk menemukan nilai kesalahan dalam klasifikasi. Dengan maksud untuk mengidentifikasi parameter terbaik dalam proses data *training*, hingga dapat memprediksi data. Dengan menggunakan prosedur *cross validation* menggunakan data *training* yang ditentukan (Purnama & Hendarsin, 2020).

Parameter SVR sendiri terdiri dari 3 parameter utama untuk optimasi *hyperlane* pada data training, antara lain nilai *C*, *Epsilon* dan *Gamma*. Nilai *C* merupakan *parameter regulasi* yang mengontrol *trade-off* antara kesalahan train dan kompleksitas model. Hasil nilai parameter *C* harus positif, dengan nilai *C* kecil memberi model margin lebar, dan nilai *C* besar akan mengurangi kesalahan yang dapat menyebabkan *overfitting*. Nilai *epsilon* berfungsi untuk menentukan margin di sekitar prediksi, dengan nilai kecil *epsilon* berarti model lebih sensitive terhadap kesalahan sedangan nilai *epsilon* besar memperikan toleransi terhadap kesalahan. Nilai *gamma* berfungsi untuk mengontrol pengaruh train, dengan defaultnya berupa "*auto*" dan "*scale*". Nilai "*auto*" pada *gamma* berarti model menggunakan *default* sederhana dan nilai "*scale*" atau skala berarti *gamma* menghitung berdasarkan jumlah fitur dan varian data yang berhubungan (Al-Ghifari Erza & Taqwa Prasetyaningrum, 2025).

Parameter ini lah yang akan menjadi hasil dari proses SVR. Parameter terbaik yang dihasilkan dalam proses SVR ini kemudian akan dilakukan untuk memprediksi data selanjutnya. Proses ini menghasilkan optimasi model dengan menemukan *hyperlane* atau proses pelatihan terbaik yang memprediksi nilai target.

2.2.3 *Mean Squared Error*

Pengukuran kinerja sistem menggunakan *Mean Squared Error* (MSE) memiliki tujuan untuk menghitung hasil prediksi dataset asli. Semakin kecil nilai MSE, semakin baik model prediksi karena kesalahan rata-rata semakin kecil.

Perhitungan MSE digunakan dengan menghitung rata-rata kuadrat selisih dari prediksi data. Pada perhitungan MSE digunakan data target dari proses SVR dan data target asli yang digunakan yang telah diolah sebelumnya. Perhitungan ini mengukur seberapa jauh hasil prediksi sebuah model dari nilai aktualnya.

Proses MSE memiliki sebuah persamaan dengan menghitung berdasarkan nilai data historis yang digunakan dalam proses SVR. Dengan rumus persamaan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - F(X_i))^2 \quad \text{.....(2.4)}$$

Dimana:

a. n, adalah jumlah total observasi.
b. yi, adalah nilai aktual data
c. $F(X_i)$, adalah nilai prediksi dari model untuk observasi ke-i (nilai data target yang telah diolah)
d. $(y_i - F(X_i))$, adalah selisih absolut antara nilai aktual dan nilai prediksi.
e. $\sum$, menunjukkan penjumlahan dari semua selisih absolut.
f. $\frac{1}{n}$, adalah kebalikan dari jumlah total observasi, yang merupakan rata-rata dari selisih absolut.

MSE memberikan indikasi seberapa baik model memprediksi nilai aktual, dan semakin rendah nilai MSE, semakin baik performa model yang akan dilakukan oleh model (Atika, 2024). Semakin kecil nilai MSE yang didapat maka semakin bagus model data yang digunakan dengan menghitung selisih dari nilai data target asli dan nilai data target prediksi yang telah dilakukan pada proses SVR sebelumnya. Hasil dari perhitungan MSE ini berupa nilai kontinu yang dengan satuan kuadrat.

2.2.4 *Root Mean Square Error*

Root Mean Square Error (RMSE) menyatakan kesalahan dalam bentuk satuan yang sama dengan target aslinya. Memberikan interpretasi yang langsung dalam skala yang sama dengan target semakin kecil nilai RMSE maka semakin baik model prediksi yang dihasilkan (Alleyda Fadhlullah et al., 2024).

Perhitungan RMSE digunakan untuk melihat rata-rata kesalahan berdasarkan dari nilai MSE yang telah dilakukan. Perhitungan ini menunjukkan seberapa jauh prediksi model dari nilai aktual target. Pada penelitian ini perhitungan PMSE digunakan untuk menunjukkan rata-rata kesalahan yang lebih mudah dipahami karena menggunakan satuan asli dan menjadi pembeda antara model perhitungan MSE yang terlalu besar.

Rumus persamaan RMSE dapat dinyatakan sebagai berikut:

$$\text{RMSE} = \sqrt{MSE} \quad \text{..... (2.5)}$$

Dimana:

a. MSE, adalah *Mean Squared Error*.
b. √, adalah fungsi akar kuadrat.

Nilai RMSE, atau mendekati 0 menunjukkan bahwa hasil peramaalaan sangat sesuai dengan data aktual atau data target. Semakin kecil nilai RMSE, semakin akurat prediksi yang dihasilkan (Nurani et al., 2023). Hasil nilai RMSE ini merupakan akar kuadrat yang menggunakan hasil dari nilai MSE yang terlalu besar. Menghasilkan nilai RMSE yang lebih kecil dan mudah dipahami.

Hasil dari perhitungan RMSE lebih mudah dipahami karena memiliki satuan yang sama dengan satuan data asli. Pada penelitian ini perhitungan RMSE digunakan untuk mengukur seberapa bagus model dibandingkan dengan perhitungan MSE yang menghasilkan nilai dengan satuan kuadrat.

2.2.5 Mean Absolute Precentage Error

Mean Absolute Percentage Error (MAPE) adalah pengukuran statistik tentang akurasi perkiraan menggunakan presentase kesalahan rata-rata secara mutlak (absolut). Pengukuran nilai MAPE memberikan gambaran mengenai berapa persen rata-rata prediksi model yang meleset dengan nilai aktual. Nilai MAPE yang digunakan dalam penelitian ini yaitu dengan menghitung nilai rata-rata dari data target berdasarkan nilai absolut data.

Pengukuran MAPE pada penelitian digunakan untuk melihat seberapa bagus tingkat akurasi model SVR dalam melakukan perhitungan yang digunakan dalam penelitian ini.

Berikut rumusan persamaan MAPE dapat dinyatakan sebagai berikut:

$$MAPE = \sum_{t=1}^{n} \left[\frac{y_i - \bar{y}}{\overline{y_1}}\right] \times 100\% \quad \text{..... (2.6)}$$

Dimana:

a. $\sum$, menunjukkan penjumlahan dari semua selisih absolut.
b. n, adalah jumlah total data.
c. t = 1, penjumlahan dimulai pada periode waktu dan berlanjut hingga periode waktu terakhir dalam kumpulan data.
d. $\frac{y_i - \bar{y}}{\overline{y_1}}$, adalah persentase kesalahan antara nilai aktual dan nilai yang diprediksi pada periode atau data ke-i

Rumus persamaan MAPE dengan menghitung nilai menggunakan rata-rata data target yang telah dihasilkan dan nilai aktual dari data target. Nilai MAPE memiliki 4 kategori atau *range* pengukuran yang menjadi interpretasi dari hasil pengukuran. Pengukuran kategori dari hasil nilai MAPE dapat diartikan dengan nilai yang ditampilkan pada tabel 2.1

Tabel 2.1 *Range* Nilai MAPE

Range MAPE	Arti Nilai
<10%	Kemampuan model peramalan sangat baik
10 - 20 %	Kemampuan model peramalan baik
20 – 50 %	Kemampuan model peramalan layak
>50%	Kemampuan model peramalan buruk

Pada tabel 2.1 merupakan tabel hasil *range* yang memiliki arti jika MAPE yang dihasilkan <10% memiliki tingkat akurasi peramalan tinggi. Nilai MAPE 10-20% tingkat akurasinya baik. Nilai MAPE 20-50% tingkat akurasinya layak, dan nilai MAPE >50% tingkat akurasinya buruk. Untuk menentukan nilai MAPE yang dihasilkan. Semakin rendah nilai MAPE, kemampuan dari model peramalan yang digunakan untuk perhitungan (Azman Maricar, n.d.).

Hasil dari perhitungan MAPE dinyatakan dalam bentuk presentase sehingga model mudah dibaca. Pengukuran ini digunakan untuk mempermudah dalam membaca pengukuran hasil model SVR karena memiliki tolak ukur untuk menginterpretasi data berupa pernyataan pendukung.

2.2.6 *R-Squared*

R-Squared (R^2) adalah nilai yang menunjukkan seberapa efektif variabel independen memengaruhi variabel independen yang lain. Angka yang dihasilkan R^2 berkisar sekitar 0 sampai 1 yang mengindikasi besarnya kombinasi variabel independen.

Pengukuan R^2 dalam penelitian digunakan untuk mengukur seberapa baik model dalam menjelaskan variasi data atau seberapa pengaruh variasi data yang digunakan. Pada penelitian ini digunakan untuk mengukur seberapa pengaruhnya data target yang telah di hasilkan dengan melalui proses SVR dengan data target aktual. Pengukuran ini menghitung dengan cara mengevaluasi nilai R^2. Berikut persamaan rumus mencari R^2 adalah

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad \text{..... (2.7)}$$

Dimana

a. yt adalah data yang diuji respon ke-t ($t = 1, \dots, n$),
b. $\hat{y}$ menunjukkan ramalan respon ke-t ($t = 1, \dots, n$),
c. $\bar{y}$ adalah rata-rata,
d. n adalah banyaknya data yang diuji

Proses persamaan diatas menjelaskan rumus pencarian R^2 dengan menggunakan "yt" sebagai data aktual target dikurangi dengan data target yang telah diolah pada proses sebelumnya dengan berdasarkan banyaknya data yang akan diuji. Pengukuran R^2 memiliki kategori interpretasi yang mengukur nilai dengan persamaan umum yang digunakan. Dengan arti hasil kombinasi hubungan variabel independen atau perhitungan R^2, yang akan dijelaskan pada tabel 2.2.

Tabel 2.2 Arti Nilai R^2

Interval koefisien	Hubungan
1-0,8	Sangat kuat
0,6-0,79	Kuat
0,4-0,59	Cukup kuat
0,2-0,39	Lemah
0-0,19	Sangat Lemah

Pada tabel 2.2 hubungan antara kombinasi variabel *independent* ditentukan untuk hasil nilai 1-0,8 berarti hubungan variabel independen dinilai sangat kuat. Nilai hasil R^2 0,6-0,79 berarti hubungan variabel independen kuat. Nilai hasil R^2 0,4-0,59 berarti hubungan variabel independent cukup kuat. Nilai hasil R^2 0,2-0,39 berarti hubungan variabel independen lemah dan nilai hasil R^2 0-0,19 berarti hubungan variabel independen sangat lemah (Nurani et al., 2023).

Hasil dari pengukuan R^2 berupa nilai antara 0-1 yang menunjukkan proporsi variasi data yang dilakukan pada mode SVR. Pada penelitian ini pengukuran R^2 dipilih untuk menujukkan seberapa kuat hubungan data target yang digunakan dengan menunjukkan hasil evaluasi akurasi terhadap model prediksi regresi SVR yang digunakan.

2.2.7 Sliding Window

Teknik *sliding window* digunakan untuk mendeteksi pola musim pada data deret waktu dengan membagi data menjadi segmen-segmen berukuran tetap (*window*), untuk setiap segmen digunakan sebagai input untuk memprediksi nilai berikutnya, Oleh karena itu, model ini sesuai untuk menangkap pola musiman serta tren jangka pendek dan panjang (Raykov, 2023)

Algoritma *sliding window* adalah metode untuk mengubah data deret waktu menjadi data potongan lintas berdasarkan hasil model sebelumnya dengan memanfaatkan ukuran jendela (*window size*). *Window size* memiliki peranan penting dalam data *time series*, hasil *window size* berperngaruh terhadap akurasi suatu model (Wahyuni et al., 2021). Teknik deret waktu ini melakukan ramalan berdasarkan data masa lampau dari target suatu model.

Teknik *sliding window* menghitung nilai prediksi lanjutan menggunakan jumlah segmen atau *window* yang digunakan. *window* atau segmen merupakan jumlah data yang akan digunakan dalam menyelesaikan proses pelatihan ini. Cara kerja teknik *sliding window* sendiri, yaitu dengan menggeserkan jendela (*window*) analisis.

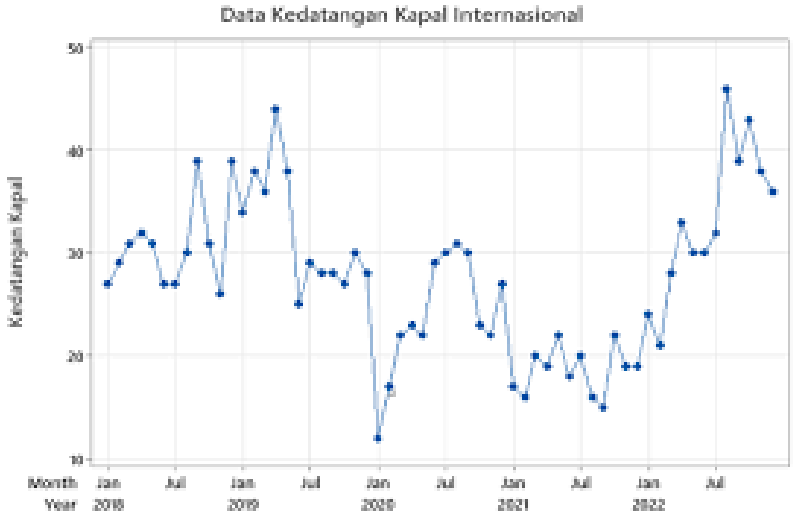
Dalam penelitian ini data window yang digunakan adalah data target yang telah diolah pada tahapan SVR. Teknik *sliding window* berfungsi untuk menghasilkan prediksi lanjutan yang telah dilakukan pada tahap SVR dengan mendeteksi pola lokal berdasarkan hasil model data target yang digunakan. Hasil dari proses *sliding window* pada penelitian ini berupa hasil dari prediksi data target.

2.2.8 Visualisasi Data

Visualisasi data merupakan menyajikan data dalam bentuk yang menarik dan mudah dipahami dengan tujuan dapat mempremudah dan membantu mempecepat dalam proses pembacaan berupa grafik (Irmayani, 2021). Dalam penelitian grafik garis dipilih dalam membantu visualisi data yang dihasilkan.

Visualisasi yang digunakan berisi data aktual target dan data prediksi yang telah dilakukan pada tahapan sebelumnya dengan. Visualisasi digunakan untuk mempermudah dalam membaca pola yang dihasilkan. Dengan menggunakan hasil prediksi dari proses *sliding window*. Pada penelitian ini digunakan diagram garis. Diagram garis (*line chart*) adalah jenis visualisasi grafik yang menampilkan data dalam bentuk titik-titk yang dibungkakan dengan garis lurus.

Diagram ini digunakan untuk menunjukkan perubahan data secara berurutan. Contoh visualisasi data yang digunakan seperti pada gambar 2.1 dibawah.



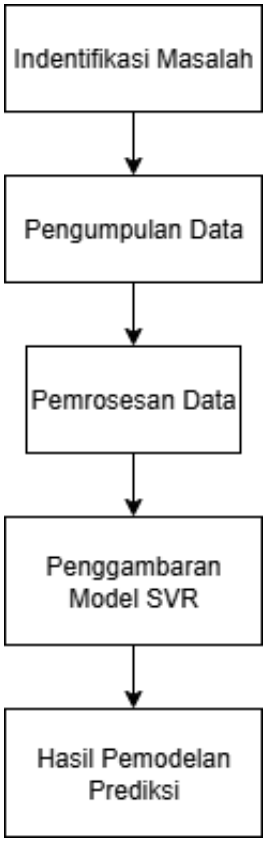
Gambar 2.1 Contoh Visualisai Daya yang akan Dibuat

Gambar 2.1 merupakan contoh visualisasi data yang akan dibuat, akan tetapi nanti data grafik yang akan menggunakan dua garis grafik dengan warna berbeda yang berisi data historis dan data prediksi yang telah dilakukan. Hasil visualisasi digunakan untuk menunjukkan tren (naik, turun, dan stabil) sebuah data dengan menampilkan pola siklus yaitu deret waktu.

Bab 3 Metode Penelitian

3.1 Prosedur Penelitian Eksperimental

Dalam melakukan prosedur penelitian maka metode yang diperlukan dalam penelitian ini berikut *flowchart* metode penelitian yang digunakan terdapat pada gambar 3.1



Gambar 3.1 Prosedur Penelitian Eksperimental

Pada gambar 3.1 merupakan tahapan penelitian ekperimentas pada sistem yang akan digunakan didalam penelitian ini.

3.1.1 Indentifikasi Masalah

Penelitian ini dilakukan karena sering terjadinya keluhan terhadap permasalahan air bersih oleh masyarakat sebagai konsumen dari PDAM. Hal ini lah yang menjadi penggerak utama dalam melakukan penelitian ini. Proses pengambilan data diawali dengan wawancara terhadap beberapa konsumen pengguna air bersih PDAM Kabupaten Magelang. Peneliti secara langsung menanyakan kepada para konsumen atas keluhan terhadap air bersih ketika musim kemarau tiba. Pada saat musim kemarau tiba banyak rumah masyarakat yang mengeluhkan air PDAM yang mati hingga beberapa hari menyebabkan kelangkaan terhadap air bersih pada waktu itu.

Dengan demikian, peneliti melakukan penelitian lebih lanjut mengenai masalah PDAM yang sering terjadi ini dengan menawarkan solusi agar dapat menjadi salah satu upaya penanggulangan terhadap permasalahan air bersih pada Perusahaan PDAM.

Tidak hanya itu, peneliti juga melakukan wawancara langsung dengan staff PDAM mengenai masalah apa saja yang sering menghambat kinerja pendistribusian air bersih ke masyarakat antara lain seperti meluapnya air ketika musim hujan tiba yang membuat pipa penyambung tempat mata air dan pendistribusian kerumah-rumah lepas. Serta mengecilnya mata air pokok ketika musim kemarau tiba yang sering menghambat kinerja distribusi air bersih ke masyarakat.

3.1.2 Pengumpulan Data

Diawali dengan melakukan wawancara untuk mengambil dataset yang diperlukan. Dalam pengambilan dataset ini, peneliti secara langsung melakukannya di kantor PDAM Kabupaten Magelang dengan menggunakan surat resmi yang dikeluarkan oleh pihak universitas (kampus). Dalam proses ini peneliti secara fisik berada di Kantor PDAM Kabupaten Magelang untuk mengumpulkan informasi yang relevan terkait dataset yang akan diambil.

Peneliti melakukan wawancara langsung terhadap staf yang mengurusi secara langsung mengenai masalah dataset ini, memungkinkan peneliti mendapatkan wawasan lebih luas mengenai produksi air, distribusi air, dan penangan terhadap beberapa masalah yang sering dihadapi di lapangan.

Data yang digunakan dalam penelitian ini mencakup data historis periode tahun 2020 hingga 2024, dengan variabel yang diambil dalam dataset ini berupa bulan dan tahun, tempat mata air terdekat, dan jumlah kebutuhan air yang telah dipakai. Mata air atau tempat produksi air di Kabupaten Magelang terdapat 19 mata air asal yang terdapat di berbagai daerah antara lain Mertoyudan, Grabag, Salaman, Borobudur, Secang, Muntilan, Banjarnegoro, Kalinegoro, dan Candi-rejo. Peneliti memfokuskan penelitiannya terhadap salah satu Lokasi agar penelitian ini dapat berjalan sesuai dengan yang diharapkan, yaitu Mata Air Semaren yang terletak pada Sub Unit Muntilan yang menjadi fokus dari penelitian ini. Hal ini akan ditampilkan pada tabel 3.1.

Tabel 3.1 Data Pengeluaran Air yang Telah Digunakan Tahun 2020-2024

No.	Tahun	Bulan	Sub Unit	Liter	M^3
1.	2020	Januari	Mertoyudan	160,67	430.337
2.	2020	Januari	Grabag	40,27	107.859
3.	2020	Januari	Salaman	45,31	121.358
….	….	….	….	....	….
538.	2024	Desember	Banjarnegoro	173,34	449.298
539.	2024	Desember	Kalinegoro	67,71	175.504
540.	2024	Desember	Candi-Rejo	85,01	220.347

Tabel 3.1 diatas menampilkan data jumlah air yang didistribusikan atau dikeluarkan Perusahaan PDAM Kabupaten Magelang di setiap bulannya periode 2020-2024. Berdasarkan data diatas jumlah keseluruhan data pengeluaran sebanyak 540 data. Data jumlah pengeluaran air yang digunakan dalam penelitian ini diperoleh dalam format PDF yang kemudian peneliti mengubahnya ke file excel dengan atribut berupa no, tahun, bulan, sub unit, liter, m^3.

Proses pengubahan data format PDF dilakukan agar lebih memudahkan dalam proses penelitian selanjutnya serta mempermudah dalam membaca data yang akan digunakan. Setelah itu, dilakukan studi literatur dengan cara mengumpulkan informasi baik berupa buku atau jurnal yang relevan terhadap penelitian yang akan dilakukan.

Studi literatur yang dikumpulkan berupa pengumpulan informasi yang berhubungan dengan prediksi kebutuhan air PDAM. Adapun cakupan lingkup studi literatur sangat luas tetapi dalam pemilihan jurnal yang relevan, peneliti melihat relevansi penelitian yang tidak lebih dari 5 tahun, sebagai acuan dalam mengkaji pemodelan sistem yang akan dibangun, dan debit produksi yang digunakan yaitu data pengeluaran air (m^3).

3.1.3 Pemrosesan Data

Pada tahap ini hasil proses pengubahan data mentah pada tabel 3.1 dari staff PDAM Kabupaten Magelang akan di sortir. Proses pertama data akan dipisah sesuai dengan sub unit yang akan dipakai, yaitu Sub Unit Muntilan lalu dilakukan penyesuaian isi variabel tahun. Dihasilkan tabel dengan variabel data yang dipakai berupa tahun, bulan, sub unit, liter dan tahun yang hanya menggunakan data Sub Unit Muntilan seperti yang ditujukkan pada tabel 3.2.

Tabel 3.2 Data Setelah Eliminasi

No.	**Tahun**	**Bulan**	**Sub Unit**	**Liter**	**m^3**
1.	2020-01	Januari	Muntilan	87,28	233.771
2.	2020-02	Februari	Muntilan	97,18	260.287
3.	2020-03	Maret	Muntilan	94,94	237.882
….	….	….	….	….	….
58.	2024-10	Oktober	Muntilan	85,36	221.253
59.	2024-11	November	Muntilan	91,21	244.297
60.	2024-12	Desember	Muntilan	91,85	238.076

Tabel 3.2 diatas merupakan tabel dataset setelah dilakukan eliminasi sesuai dengan Sub Unit Muntilan beserta dengan pengeluaran produksi (m^3). Dapat dilihat hasil dari tahap eleminasi menghasilkan sebanyak 60 data. Data ini merupakan data pengeluaran air setiap bulannya yang di produksi atau didistribusikan oleh Sub Unit Muntilan. Dengan menggunakan variabel tabel yang sama pada tabel 3.1. Data tabel ini lah yang akan digunakan dalam proses tahapan SVR dengan menggabungkan kombinasi lainnya.

3.1.4 *Support Vector Regression*

Setelah data diolah dan dilakukan kajian literatur, tahap berikutnya pemilihan model untuk dilakukan proses pengolahan data melalui pengkoden, sehingga model yang akan digunakan adalah model pengolahan data menggunakan metode SVR dengan mengkombinasikan pendekatan *sliding window*.

Metode SVR merupakan suatu program pemodelan pendekatan prediksi, yang pada penelitian ini digunakan *sliding window* sebagai algoritma yang mendeteksi pola musim untuk memprediksi nilai berikutnya. Data yang digunakan dalam metode ini, yaitu *train data* yang berasal dari data historis m^3. Metode SVR digunakan untuk mengukur parameter dengan menggunakan *karnel grid search* dalam pengolahannya. *Grid search* merupakan Teknik *hyperparameter tuning* yang digunakan untuk mencari kombinasi terbaik dari model SVR. Metode ini mencoba semua kombinasi parameter, lalu memilih yang memberikan kinerja prediksi terbaik untuk menghitung hasil matrik evaluasi.

Pada tahap ini, hasil dari algoritma prediksi menggunakan MSE, RMSE, MAPE dan R^2 akan terlihat. Hasil dari metode ini menunjukkan hasil metode SVR, beserta dengan nilai parameternya, yaitu *C*, *epsilon* (ε), *gamma* (γ). Hasil parameter ini lah yang akan menjadi hasil dari proses SVR menggunakan *grid search*. Proses kerja *Grid Search* SVR, yaitu dengan menentukan parameter dari data yang digunakan. Perhitungan SVR ini digunakan data fitur dan data target dalam menentukan hasil parameter. Dengan melatih model kinerja SVR untuk setiap kombinasi yang dipilih.

3.1.5 Hasil Pemodelan Prediksi

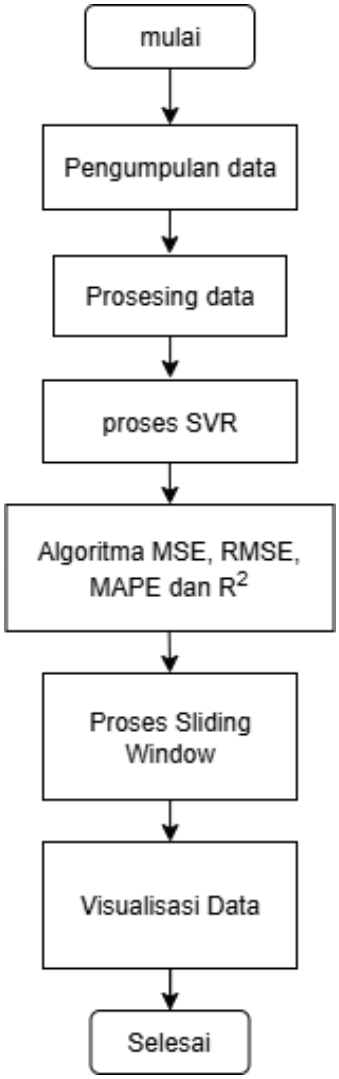
Hasil dari pemodelan prediksi ini menampilkan jumlah prediksi kebutuhan air PDAM Kabupaten Magelang Sub Unit Muntilan pada bulan berikutnya sebanyak 12 bulan, yaitu memprediksi data pada tahun 2025. Penelitian ini berfokus pada hasil prediksi dalam memprediksi jumlah kebutuhan air, serta menampilkan hasil evaluasi model prediksi data yang dihasilkan dari tahapan metode SVR dan pendekatan *sliding window* yang dilakukan.

Proses *sliding window* digunakan untuk memprediksi nilai lanjutan dengan menggunakan parameter terbaik dari proses SVR yang didapat. Proses *sliding window* menggunakan jumlah window sebanyak 12 atau menggunakan 12 data historis untuk memprediksi 12 data kedepan. Penelitian ini menggunakan *last window* atau data historis terakhir untuk dilakukannya proses selanjutnya dalam mencari data prediksi selanjutnya.

Hasil dari pemodelan ini kemudian akan diproyeksikan ke dalam bentuk tabel hasil dari pemodelan sebanyak 12 bulan mendatang, yaitu pada tahun 2025 berserta grafik garis berdasarkan data historis dan data prediksi yang dihasilkan .

3.2 Pemodelan Sistem

Tahap rancangan percobaan analisa sistem dalam penelitian ini dijelaskan pada gambar 3.2, terhadap rancangan percobaan berikut:



Gambar 3.2 Diagram Tahapan Analisa Sistem

Pada gambar 3.2 digambarkan tahapan analisa sistem yang akan digunakan dimulai dengan proses pengumpulan data yang dilakukan melalui wawancara dengan masyarakat dan staff Perusahaan PDAM dalam menentukan masalah air yang ada dan dilanjutkan dengan meminta data ke perusahaan PDAM Kabupaten Magelang. Setelah data didapat dilanjutkan dengan menginput data ke *excel* secara manual untuk mempermudah dalam proses analisa data dan memasukkannya ke dalam proses pengkodean.

Data yang digunakan dalam proses ini yaitu data historis penggunaan air yang telah di produksi sebelumnya. Langkah kedua, yaitu melakukan prosesing data. Dilakukannya tahap prosesing data berfungsi agar data yang digunakan siap untuk dilanjutkan ke tahap pemodelan. Setelah itu, masuk ke tahap proses SVR untuk mencari hasil parameter terbaik dari proses pemodelan ini, yaitu nilai parameter *C*, *Epsilon*, dan *Gamma*. Setelah itu, masuk ketahap pencarian algoritma MSE, RMSE, MAPE dan R^2 untuk melihat seberapa baik pemodelan ini dilakukan dalam memprediksi data PDAM. Dilanjutkan proses *sliding window* untuk mencari prediksi data selanjutnya, dan terakhir yaitu visualisasi yang berisi tabel hasil prediksi dan grafik data histogram pada *sliding window* berserta hasil prediksi yang dihasilkan.

3.2.1 Pengumpulan Data

Tahap pengumpulan data dimulai dari dilakukannya wawancara permasalahan yang sering dikeluhkan masyarakat yaitu sering terjadinya matinya air PDAM yang diakibatkan dengan beberapa faktor, seperti faktor musim ataupun geografi menyebabkan kelebihan air yang berimbas pada sering terjadinya pipa penyambung yang lepas dikarenakan kenaikan air yang terjadi tiba-tiba ataupun kekurangan air pada musim kemarau yang menyebabkan terjadinya kelangkaan akan sumber daya air.

Pengumpulan dataset dilakukan dengan menanyakan langsung ke Kantor PDAM Kabupaten Magelang menggunakan surat resmi. Data yang akan dipakai pada proses ini merupakan data historis penggunaan air mulai dari tahun 2020 hingga tahun 2024. Data ini merupakan rekapan produksi air perbulan yang dikeluarkan oleh PDAM. Data asli yang diberikan berupa data mentah berbentuk hardfile yang kemudian di scan untuk diberikan pada penelitian ini. Dataset yang terkumpul kemudian di input menggunakan *excel* secara manual oleh penulis dan terkumpul sebanyak 540 data dengan variabel atribut berupa kolom tahun, bulan, sub unit (sumber mata air) serta pemakaian air yang telah digunakan (L/dtk dan m^3) atau yang terlihat pada tabel 3.1 atau biasa disebut dengan data premier.

3.2.2 Prosesing Data

Selanjutnya data primer yang telah diubah akan dipisahkan guna untuk memfokuskan pada tujuan awal penelitian. Data yang ada kemudian dilakukan penyortiran menurut sub unit yang menjadi fokus utama penelitian, yaitu pada Sub Unit Muntilan. Data yang disortir akhirnya berjumlah 60 data dengan variable yang sama, seperti yang ditujukkan pada tabel 3.2. Data inilah yang akan menjadi sumber utama dalam proses prediksi yang akan dilakukan.

Setelah dilakukan pernyortiran data kemudian data akan disesuaikan dengan model, mulai dari mencari data kosong, dan mengubah data yang belum sesuai dengan model SVR dan dilakukan proses meratakan nilai data.

Data yang akan digunakan dalam proses ini yaitu, data numerik. Data tersebut adalah data kolom tahun, data kolom liter, dan data kolom m^3. Data yang telah di prosesing dan siap digunakan akan masuk ke tahap selanjurnya, yaitu proses SVR *Grid search* untuk mencari hasil parameter dari data yang telah diproses.

3.2.3 Proses *Support Vector Regression* (SVR)

Setelah dataset diolah kemudian data akan diproses kedalam metode SVR. Metode SVR merupakan metode pengembangan dari model SVM yang digunakan untuk kasus regresi data non-linier atau numerik. SVR bekerja dengan mencari fungsi regresi (*hyperplane*) yang paling sesuai dengan data namun masih dalam batas margin *epsilon* (ε) tertentu. Di dalam SVR digunakan fungsi kernel dalam menangani data non-linear (Algoritma et al., n.d.).

Penelitian ini menggunakan fungsi *kernel Grid Search* dikarenakan fungsi ini dapat menghasilkan model yang mengikuti pola data, meminimalkan error serta dapat memetakan data ke ruang dimensi yang lebih tinggi dengan demikian pola yang dihasilkan lebih kompleks dan bisa diprediksi dengan baik (Afifah et al., n.d.). Oleh karena itu, fungsi karnel ini dipilih untuk melakukan penelitian ini.

Selain menggunakan fungsi kernel di dalam model SVR terdapat pula parameter utama meliputi nilai *C* (*Cost atau Regularization Parameter*) yang berfungsi dalam memfokuskan model dan meminimalkan error. *Epsilon* (*ε*) yang berfungsi menentukan batas toleransi kesalahan pada fungsi regresi, namun error yang dihasilkan tidak dihitung sebagai kesalahan. Oleh karena itu nilai epsilon sangat berpengaruh dalam sensitivitas model. *Gamma* (*γ*) khusus untuk kernel non-linear, yang mengontrol pengaruh antara satu data terhadap data yang lain (Mulyawan et al., n.d.).

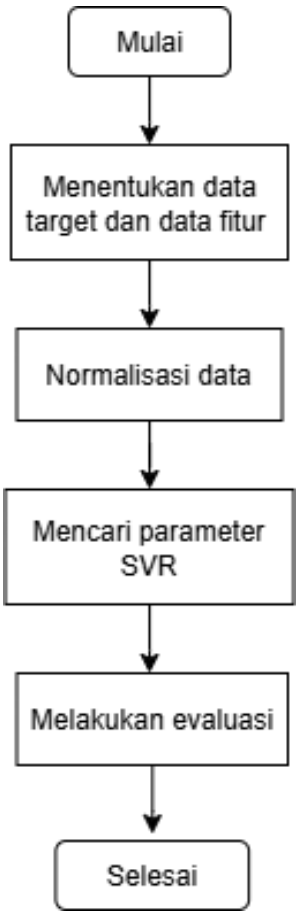
Dengan kata lain parameter nilai *C* berfungsi untuk mengatur seberapa ketat model dalam mengikuti data training yang diolah. Jika nilai *C* yang didapat besar kemungkinan model akan berusaha keras mengurangi kesalahan sekecil mungkin, bahkan beresiko sampai menghapal data (*overfitting*) dan jika nilai *C* yang didapat kecil maka model lebih santai, membiarkan adanya kesalahan asalkan model yang digunakan sederhana.

Parameter *Epsilon* berfungsi untuk menentukan batas toleransi kesalahan yang dianggap masih oke, tanpa dihitung sebagai *error*. Jika nilai yang didapat kecil, model lebih banyak kesalahan kecil yang diabaikan menyebabkan model jadi kurang sensitif terhadap detail kecill kesalahan. Jika nikai *Epsilon* ynag dihasilkan kecil, model hampir menghitung semua penyimpangan yang dihitung sebagai kesalahan menyebabkan model lebih sensitif.

Nilai parameter *Gamma* berfungsi untuk mengatur jarak pengaruh dari satu data terhadap data yang lain. Jika nilai Gamma besar, setiap data punya mengaruh sangat sempit dan hanya berpengaruh pada data yang ada disebelahnya, menyebabkan model bisa terlalu detai atau terjadi overfitting. Jika nilai *gamma* yang dihasilkan kecil, setiap data mempengaruhi area yang luas, menyebabkan model lebih halus.

Dapat diambil kesimpulan bahwa nilai parameter yang dihasilkan berpengaruh terhadap proses perhitungan prediksi selanjutnya. Dengan menghitung *error* hasil model SVR dengan menggunkan proses MSE, RMSE, MAPE ditambah proses perhitungan R^2 untuk memperkirakan seberapa besar pangaruh satu variabel dengan variabel lain yang digunakan.

Berikut tahapan atau *flowchart* dari proses pemodelan menggunakan metode SVR yang akan dilakukan, seperti yang ditampilkan pada gambar 3.3 sebagai berikut:



Gambar 3.3 Proses SVR

Pada gambar 3.3 diatas memperlihatkan tahapan dari proses SVR diawali dengan menentukan data fitur dan data target. Data fitur (x) yang digunakan dalama proses pemodelan SVR ini adalah data kolom "liter" dan data target (y) yang digunakan adalah data kolom "m^3" yang dimuat berdasarkan data "tahun". Dengan data yang digunakan merupakan data yang telah dieleminasi atau data Sub Unit Muntilan yang menjadi tujuan utama dari penelitian ini.

Setelah data ditentukan, langkah selanjutnya adalah dengan menormalisasi data, tahapan ini dilakukan untuk menyesuaikan data sesuai dengan model SVR sehingga akurasi yang akan didapat memiliki nilai yang lebih tinggi dan akurat. Proses normalisasi data yang digunakan adalah dengan menggunakan *StandardScaler*. *StandardScaler* dipilih karena dapat menyamakan skala data membuat semua data punya sekala seimbang. *StandardScaler* berfungsi untuk meningkatkan stabilitas proses perhitungan. Lebih efektif menghindari bias parameter dan membantu kinerja model SVR yang sangat sensitif terhadap jumlah data, hasilnya model lebih akurasi, stabil dan efisien saat dilatih.

Setelah itu dilakukan tahapan mencari parameter utama SVR untuk mempersiapkan model melakukan proses prediksi. Dan terakhir masuk ke tahap evaluasi. Tahapan ini berisi evaluasi sistem berupa *print* hasil parameter SVR.

3.2.4 Algoritma MSE, RMSE, MAPE dan R^2

MSE (*Mean Squared Error*) dan RMSE adalah (*Root Mean Squared Error*) berfungsi sebagai matrik evaluasi kinerja atau akurasi model. MSE berfungsi mengukur rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, sedangkan RMSE berfungsi mengukur seberapa baik model prediksi dengan dihitung sebagai akar dari rata-rata kuadrat dari selisih nilai prediksi dan nilai aktual (Febriana Yusuf, 2024). Pada proses akurasi digunakan MAPE dan R^2 yang memiliki fungsi berbeda-beda. MAPE berfungsi untuk mengukur akurasi menggunakan rata-rata kesalahan, dan R^2 berfungsi mengukur akurasi hubungan variabel model SVR.

Proses akurasi menggunakan algoritma diatas berfungsi untuk menghitung proses akurasi berdasarkan model SVR yang telah dilakukan MSE, RMSE, dan MAPE berfungsi untuk mengukur rata-rata kesalahan prediksi dengan aspek yang berbeda. Sedangkan pengukuran akurasi R^2 berfungsi untuk mengukur seberapa pengaruhnya variasi data yang digunakan terhadap data lainnya. Data yang digunakan untuk keempat proses pengukuran akurasi ini yaitu data target dari proses SVR yang telah diolah serta data target aktual yang telah dilakukan pembersihan data pada tahap awal preprosesing data.

MSE digunakan untuk menghitung rata-rata kuadrat selisih antara data aktual dan data prediksi SVR, dengan memberikan Gambaran seberapa besar *error* yang didapat dan tingkat sensiitifitas. RMSE adalah akar kuadrat dari perhitungan MSE, berfungsi untuk mempermudah dalam interpretasi atau menampilkan data karena satuan yang dihasilkan lebih kecil. MAPE adalah mengukur rata-rata presentase kesalahan data target prediksi SVR terhadap nilai aktual data target yang digunakan. R^2 digunakan untuk mengukur kombinasi antar variabel data. Jika hasill nilai mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang baik karena hubungan antar variabel data sangat cocok.

3.2.5 *Sliding Windows*

Sliding Windows digunakan sebagai pemrosesan data dalam potongan (*window*) yang berukuran tetap dan bergerak secara bertahap berupa *string*. Teknik ini digunakan sebagai fitur dalam membentuk dataset input dan output dari data deret waktu, dimana data histori digunakan untuk memprediksi nilai berikutnya (Wahyuni et al., 2021b). Teknik *sliding window* menggunakan data historis target dalam melakukan prosesnya. Dengan menetapkan ukuran *window* sebagai data input dari *sliding window*. Pembentukan struktur dari data *time series* dan segmen yang dilakukan untuk memperoleh *error* terkecil(Arsi et al., 2022).

Cara kerja proses ini, yaitu dengan menggeser jendela sesuai dengan nilai jendela yang telah ditentukan sebelumnya. Proses ini akan dilakukan terus menerus hingga hasil prediksi yang ditentukan selesai. Ketika dikombinasikan dengan model SVR, *sliding window* berkerja sebagai rekayasa fitur atau tempat pembuatan vitur, dan SVR sebagai model prediksi.

Proses *sliding window* dilakukan dengan mengtrain data target yang akan digunakan dengan menentukan jumlah *window* yang digunakan. Tahap ini digunakan sebanyak 12 *window* untuk menghasilkan sebanyak 12 data prediksi. Dengan menggunkan *last window* atau data terakhir histori dari data target untuk dilakukan proses train, misalnya jika dinggunakan data terakhir adalah y1, y2, y3, proses ini akan memprediksi data y4. Teknik ini akan dikombinasikan dengan hasil parameter pada proses model SVR menghasilkan data prediksi dari data yang digunakan yang menudian data akan dibagi menjadi set traing dan set testing.

3.2.6 Visualisasi Data

Hasil dari proses ini berupa data prediksi time series serta hasil dari evaluasi menggunakan MSE, RMSE, MAPE dan R^2. Tidak hanya tabel hasil data prediksi akan tetapi ada juga gambar berupa plot prediksi yang memudahkan dalam membaca visualisasi akurasi model dan pola prediksi terhadap data asli yang digunakan, seperti gambar 3.3 yang merupakan contoh plot hasil prediksi.

Proses visualisasi data merupakan proses interpretasi data prediksi. Pada proses visualisasi data yang akan ditampilkan sesuai dengan hasil dari prediksi data yang telah dilakukan. Data tersebut merupakan data histori dari data target yaitu periode 2020-2024 menghasilkan prediksi pada tahun 2025. Data prediksi yang dihasilkan kemudian dilanjutkan dengan mengubahnya kedalam bentuk diagram garis untuk mempermudah dalam melihat pola data dari data historis target yang digunakan dan prediksi data target yang dihasilkan berdasarkan berdasarkan data tahunnya. Serta menampilkan hasil dari MSE, RMSE, MAPE dan R^2 yang akan ditampilkan dalam proses menghitung akurasi SVR, tidak pada pencarian *sliding window*.

3.3 Simulasi Sistem SVR

Berdasarkan analisa sistem yang telah dijelaskan, berikut pengujian sistem dan pembuktian efektivitas dari metode SVR. Pada simulasi ini digunakan pengujian algoritma RBF dalam melakukan perhitungan manual dari pemodelan SVR. Data yang digunakan dalam pengujian ini merupakan data real empat bulan, yaitu bulan Januari hingga April yang memprediksi bulan Mei tahun 2023, seperti yang terlihat pada tabel 3.3.

Tabel 3.3 Tabel Data Input

No	Bulan	m^3
1.	Januari	231.815
2.	Februari	242.342
3.	Maret	217.341
4.	April	239.743

Dengan data aktual Mei = 227.530m^3

a. Bentuk vektor input historis (xt)

$xt = 222.367$ ..... (3.1)

Dengan data training:

- $x_1 = 231.815\,, y_1 = 231.815$
- $x_2 = 242.342,\ y_2 = 242.342$
- $x_3 = 217.341,\ y_3 = 217.341$
- $x_4 = 239.743,\ y_4 = 239.743$
- $x_5 = 222.367,\ y_5 = 222.367$

b. Tentukan *Kernel*

$K(x_t, x_i) = \exp(-\gamma||x_t - x_i||^2)$ ..... (3.2)

Asumsikan γ=1$\times\ 10^{-6}$ (karena nilai data besar dan y harus kecil, namun tidak membuat *kernel* menjadi 0.

Dihasilkan:

$k(x_t, x_1) = \exp\,(-10^{-6}.(222.367 - 231.815)^2)$ ..... (3.3)

$= \exp(-10^{-6}.(-9448))^2)$

$= \exp(-10^{-6}.89.268.704)$

$= \exp\,(-89{,}2687) \approx 1.7093524708 \times 10^{-39}$

$k(x_t, x_2) = \exp\,(-10^{-6}.(222.367 - 242.342)^2)$ ..... (3.4)

$= \exp(-10^{-6}.(-19975))^2)$

$= \exp(-10^{-6}.399.000.025)$

$= \exp(-399.000) \approx 5.20271779980 \times 10^{-174}$(≈ 0 numercal)

$k(x_t, x_3) = \exp\,(-10^{-6}.(222.367 - 217.341)^2)$ ..... (3.5)

$= \exp(-10^{-6}.(5026))^2)$

$= \exp(-10^{-6}.25.260.676)$

$= \exp\,(-25{,}2607) \approx 1.0701084762 \times 10^{-11}$

$k(x_t, x_4) = \exp\,(-10^{-6}.(222.367 - 239.743)^2)$ ..... (3.6)

$= \exp(-10^{-6}.(-17376))^2)$

$= \exp(-10^{-6}.301.908.736)$

$= \exp\,(-301{,}9087) \approx 7.50771528187 \times 10^{-132}$ (≈ 0 numercal)

Dibulatkan

- $K_1 \approx 1.70935 \times 10^{-39}$
- $K_2 \approx 5.20 \times 10^{-174}$ (praktis 0)
- $K_3 \approx 1.07011 \times 10^{-11}$
- $K_4 \approx 7.51 \times 10^{-132}$ (praktis 0)

Nilai parktis 0 artinya jarak x_t ke x_2 dan x_4 jauh sekali, dan nilai dan nilai γ tidak terlalu kecil, menyebabkan hasil perhitungan karnel hampir nol (0).

c. Masukan model SVR

$Yt = \sum_{(i=1,N)} \alpha_i(x_t - x_i) + K_i$ ..... (3.7)

Dimisalkan:

Dual coefficients : [100, -100, 100, -100]

Dihasilkan:

- $100 \times K_1 \approx 100 \times 1.70935 \times 10^{-39} \approx 1.7935 \times 10^{-37}$ ..... (3.8)
- $-100 \times K_2 \approx -100 \times 5.20 \times 10^{-174} \approx -5.20 \times 10^{-172}$..... (3.9)
- $100 \times K_3 \approx 100 \times 1.07011 \times 10^{-11} \approx 1.07011 \times 10^{-9}$ ..... (3.10)
- $-100 \times K_4 \approx -100 \times 7.51 \times 10^{-132} \approx -7.51 \times 10^{-130}$..... (3.11)

Ketika semua hasil di jumlahkan, dihasilkan:

$Yt \approx 1.0701084761711995 \times 10^{-9}$..... (3.12)

Dihasilkan nilai SVR adalah $1.0701084761711995 \times 10^{-9}$ nilai ini sangat kecil karena hampir mendekati nol.

d. Proses perhitungan prediksi

$$y_{pred} = Yt + b \approx 1.0701084761711995 \times 10^{-9} + 231055.05826358683$$
$$\approx 231055.05826 \text{.....} (3.13)$$

Dihasilkan nilai prediksi Juni berdasarkan data yang telah diproses adalah **231.055,06** m^3 Nilai b atau bias didapat dari hasil run perhitungan pada pustaka *scikit-learn interceprt_*. Atau secara manualnya dapat dihitung dengan persamaan:

$b = y - nilai\ dual\ C$ ..... (3.14)

e. Evaluasi MSE dan RMSE terhadap prediksi bulan Juni

- Juni aktual = 227.530 m^3
- Juni prediksi = 231.055 m^3 (dibulatkan)

$MSE = (227.530 - 231.055)^2 = 12.036{,}76$ m^{32}..... (3.15)

$RMSE = \sqrt{12.036{,}76} = 3.525{,}06$ m^3..... (3.16)

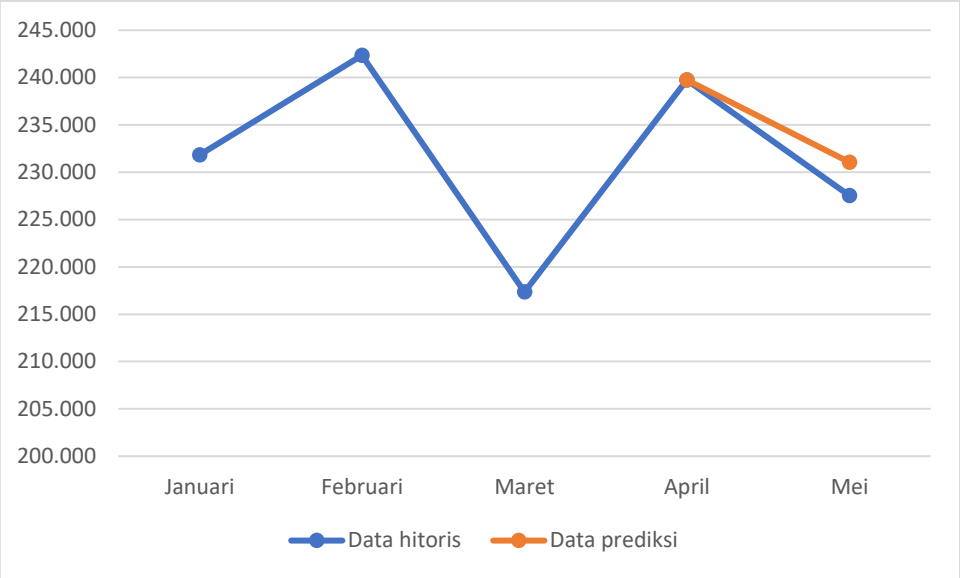
Evaluasi MSE dan RMSE digunakan untuk mencari nilai selisih dari data aktual dan data prediksi. MSE menghitung selisih dengan hasil yang memiliki satuan kuadrat, dan RMSE menghitung selisih dengan hasil satuan yang sama dengan data yang digunakan. Dihasilkan selisih kedua data sebesar 3.525,06 m^3.

f. Visualisasi Hasil Prediksi

Tahap ini merupakan tahap perbandingan data aktual bulan desember dan data prediksi yang dihasilkan pada percobaan simulasi perhitungan manual menggunakan tabel yang ditujukkan pada tabel 3.4 dan visualisasi grafik histogram pada gambar 3.4.

Tabel 3.4 Perbandingan Data

Juni Aktual(m^3)	Juni Prediksi(m^3)
227.530	231.055



Gambar 3.4 Visualisasi Perbandingan Data Aktual dan Data Prediksi

Pada tabel 3.4 dan gambar 3.4 menunjukan perbandingan hasil perhitungan manual sederhana antara data aktual, dan data prediksi, dimana data aktual dilambangkan dengan garis oranye dan data prediksi yang dihasilkan warna biru. Dengan data yang digunakan dalam perhitungan ini, yaitu data Bulan Januari, Februari, Maret, April sebagai nilai input yang memprediksi bulan Mei pada tahun 2023. Prediksi menggunakan SVR meramalkan nilai prediksi sebesar 231.055m^3, sedangkan data aktual sebesar 227.530 m^3. Dengan mencari selisih menggunakan MSE dan RMSE, yaitu MSE sebesar 12.036,76 m^{32} dan RMSE sebesar 3.525,06 m^3.

Akan tetapi fokus pengkodean penelitian ini terletak pada proses *grid search* yang mencari parameter terbaik nilai *C*, *Epsilon* dan *Gamma* dengan mengkombinasikan menggunakan *cross validation* yang ada di *param_grid*. Ini hanya perhitungan manual dari proses SVR untuk melihat seberapa bagus model SVR dalam mengasilkan prediksi hasil menggunakan data yang ada melalui proses nomalisasi dan *standard scaler* yang ada di pustaka *scikit-learn*.

Bab 5 Penutup

5.1 Kesimpulan

Berdasarkan dari penelitian yang telah dilakukan, dapat disimpulkan bahwa hasil dari penelitian adalah, sebagai berikut:

a. Metode SVR melakukan prediksi dengan mencari parameter terbaik yaitu dihasilkan nilai c=10, *epsilon*=0,1 dan *gamma*="*auto*", yang digunakan untuk memprediksi data selanjutnya menggunakan algoritma *sliding window*.
b. Tingkat akurasi yang dihasilkan berdasarkan nilai MSE sebesar 0,1115 dan RMSE sebesar 0,3340, sedangkan hasil dari pengukuran akurasi berdasarkan dari nilai MAPE sebesar 94,6612% dan nilai R^2 yang dihasilkan sebesar 0,8885.
c. Pada proses SVR menghasilkan parameter terbaik dan kemudian dikombinasikan dengan teknik sliding window digunakan jumlah jendela sebanyak 12 atau data yang digunakan untuk prediksi data ini sebanyak 12 data terakhir dari data target berhasil dilakukan, data yang dihasilkan, yaitu pada Bulan Januari sebanyak 229.294 m^3, pada Bulan Februari sebanyak 234.515 m^3, Bulan Maret sebanyak 227.326 m^3, Bulan April sebanyak 227.356 m^3, Bulan Mei sebanyak 230.241 m^3, Bulan Juni sebanyak 232.909 m^3, Bulan Juli sebanyak 225.334 m^3, Bulan Agustus sebanyak 214.598 m^3, Bulan September sebanyak 217.899 m^3, Bulan Oktober sebanyak 197.248 m^3, Bulan November sebanyak 205.513 m^3, dan Bulan Desember sebanyak 201.084 m^3.

5.2 Saran

Disarankan untuk penelitian lanjutan, yaitu menambah tingkat akurasi dengan menambahkan jumlah data yang lebih banyak lagi karena proses model SVR sangat berpengaruh dengan skala data, serta coba gunakan *sliding window* dengan *size* atau ukuran yang berbeda ataupun menguji kembali parameter SVR menggunakan karnel berbeda untuk melihat perbandingan hasil. Tidak hanya itu, lakukan pengujian menggunakan model lain sebagai perbandingan untuk mendapatkan hasil prediksi yang maksimal.

5.3 Keterbatasan Penelitian

Keterbatasan dari penelitian ini terletak pada jumlah data yang masih sangat sedikit. Keterbatasan jumlah data inilah yang membuat model SVR ini masih belum dikatakan maksimal karena sangat berpengaruh dalam proses penentuan akurasi dan prediksi selanjutnya.

Referensi

10065-31701-1-PB. (n.d.).

Agustina, D., Hafiyusholeh, Moh., Fanani, A., & Prasetijo, D. (2023). Prediksi Distribusi Air Perusahaan Daerah Air Minum (PDAM) Tirta Dharma Kota Pasuruan Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Jurnal PROCESSOR*, *18*(1). https://doi.org/10.33998/processor.2023.18.1.697

Al-Ghifari Erza, M., & Taqwa Prasetyaningrum, P. (2025). Analisis Kinerja Model Support Vector Machine dalam Prediksi Kasus HIV di Indonesia Berdasarkan Data Time Series. *Technology and Science (BITS)*, *7*(1). https://doi.org/10.47065/bits.v7i1.7365

Alleyda Fadhlullah, L., Purwadi, J., & Dahlan Yogyakarta, A. (2024). Optimalisasi Parameter Support Vector Regression dengan Algoritma Random search dan Algoritma Grid search Optimization of Support Vector Regression Parameters with Random search Algorithm and Grid search Algorithm. *Jurnal Ilmiah Matematika*, *11*(1), 50–61. https://doi.org/10.26555/konvergensi.29917

Arsi, P., Astuti, T., Rahmawati, D., & Subarkah, P. (2022). *Implementasi Sliding Window Algorithm pada Prediksi Kurs berbasis Neural Network*. *6*(1), 51–59. http://e-journal.unipma.ac.id/index.php/doubleclick

Azman Maricar, M. (n.d.). *Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ*.

BÎLBÎE, F., & ZAHARIA, L. (2024). *Flash Flood Forecasting Using Machine Learning Models: A Scientometric Analysis*. 1–10. https://doi.org/10.24193/AWC2024_01

Febriana Yusuf, B. (2024). IMPLEMENTASI ALGORITMA SUPPORT VECTOR REGRESSION DALAM MEMPREDIKSI HARGA SAYURAN KOMODITAS (STUDI KASUS : DESA HEGARMANAH). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 5).

Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, *4*(2). https://doi.org/10.30813/jbase.v4i2.3000

Hikmah, H., Asrirawan, A., Apriyanto, A., & Nilawati, N. (2023). Peramalan Data Cuaca Ekstrim Indonesia Menggunakan Model ARIMA dan Recurrent Neural Network. *Jambura Journal of Mathematics*, *5*(1). https://doi.org/10.34312/jjom.v5i1.17496

Irmayani, W. (2021). *VISUALISASI DATA PADA DATA MINING MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES*. www.bsi.ac.id

Kartini, D., Abadi, F., & Saragih, T. H. (2021). Prediksi Tinggi Permukaan Air Waduk Menggunakan Artificial Neural Network Berbasis Sliding Window. *Jurnal RESTI*, *5*(1), 39–44. https://doi.org/10.29207/resti.v5i1.2602

Najla, G., #1, A., & Fitrianah, D. (n.d.). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, *14*(2).

Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, *6*(1), 34–43. https://doi.org/10.24246/juses.v6i1p34-43

Nurul Noviyana, & Muhammad Irwan Padli Nasution. (2024). Penerapan Teknologi Informasi Untuk Efektivitas dan Efisiensi Pengolahan Data Mahasiswa. *JURNAL ILMIAH RESEARCH AND DEVELOPMENT STUDENT*, *2*(1), 152–160. https://doi.org/10.59024/jis.v2i1.578

Purnama, D. I., & Hendarsin, O. P. (2020). Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jambura Journal of Mathematics*, *2*(2), 49–59. https://doi.org/10.34312/jjom.v2i2.4458

Radho, A. E., Sugiartawan, P., & Santiago, G. A. (2021). Prediksi Jumlah Kasus COVID-19 Menggunakan Teknik Sliding Wondows dengan Metode BPNN. *Jurnal Sistem Informasi Dan Komputer Terapan Indonesia (JSIKTI)*, *4*(1), 12–23. https://doi.org/10.22146/jsikti.xxxx

Rais, Z. (2022). ANALISIS SUPPORT VECTOR REGRESSION (SVR) DENGAN KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK MEMPREDIKSI LAJU INFLASI DI

INDONESIA. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(1), 30–38. https://doi.org/10.35580/variansiunm13

Raykov, P. (2023). An Optimal Algorithm for Sliding Window Order Statistics. *Leibniz International Proceedings in Informatics, LIPIcs*, *255*. https://doi.org/10.4230/LIPIcs.ICDT.2023.5

Ruliana, R., Rais, Z., Marni, M., & Ahmar, A. S. (2024). Implementation of the Support Vector Regression (SVR) Method in Inflation Prediction in Makassar City. *ARRUS Journal of Mathematics and Applied Science*, *4*(1), 28–35. https://doi.org/10.35877/mathscience2608

Saadah, S., Zahra, F., & Haifa, H. (n.d.). *Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD (Support Vector Machine (SVM) To Predict Crude Oil Palm in Indonesia and Exchange Rate of EUR/USD)*. http://jcosine.if.unram.ac.id/

Safira, A. N., Warsito, B., & Rusgiyono, A. (2023). ANALISIS SUPPORT VECTOR REGRESSION (SVR) DENGAN ALGORITMA GRID SEARCH TIME SERIES CROSS VALIDATION UNTUK PREDIKSI JUMLAH KASUS TERKONFIRMASI COVID-19 DI INDONESIA. *Jurnal Gaussian*, *11*(4), 512–521. https://doi.org/10.14710/j.gauss.11.4.512-521

Sepri, D., Fauzi, A., Wandira, R., Riza, O. S., & Wahyuni, Y. F. (2020). Hummaira Hutagaol 6) 1) 2) 3) 4) 5) 6) UIN Imam Bonjol. *Jl. Prof Yunus*. http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis

SKRIPSI AULYA ATIKA MAHGFIRAH_E0220308 (pdf.io). (n.d.).

Thoriq, M., Syaputra, A. E., & Eirlangga, Y. S. (n.d.). *Perkiraan Kebutuhan Air Bersih Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation*.

Wahyuni, R. E., Provinsi, B., & Tengah, J. (2021a). OPTIMASI PREDIKSI INFLASI DENGAN NEURAL NETWORK PADA TAHAP WINDOWING: ADAKAH PENGARUH PERBEDAAN WINDOW SIZE? In *Technologia* (Vol. 12, Issue 3).