

BAB 2 LANDASAN KEPUSTAKAAN

2.1 Kajian pustaka

Penelitian mengenai peramalan telah banyak dilakukan oleh beberapa peneliti sebelumnya, contohnya adalah penelitian yang dilakukan oleh Du, et al. (2013), yaitu penggunaan algoritma SVM untuk peramalan permintaan hasil produksi pertanian yang tidak tahan lama. Hampir semua jenis hasil produksi pertanian yang tidak tahan lama, seperti telur, susu, dan daging memiliki waktu penyimpanan yang sangat pendek serta mudah membusuk. Oleh karena itu, peramalan ini sangat dibutuhkan oleh penyedia dan penjual untuk memperkirakan permintaan atas hasil produksi agar dapat mengendalikan biaya penyimpanan. Pada penelitian ini, peneliti melakukan langkah awal yaitu menangani data dengan dua cara. Cara yang pertama yaitu memproses data penjualan secara historis, yang terdiri dari pemrosesan pemerataan data (untuk menghilangkan nilai *singular*) dan pemrosesan normalisasi data (untuk menghindari data *overflow*). Cara yang kedua yaitu memproses informasi dinamis, seperti data cuaca, data temperatur, dan sebagainya yang berhubungan dengan data penjualan historis, dengan menggunakan metode *fuzzy*. Kumpulan data latih dibentuk setelah mengolah data awal. Kemudian, data latih tersebut dimasukkan ke dalam model SVM untuk menyesuaikan parameter dengan nilai optimal. Sistem akan memperkirakan permintaan masa depan setelah selesai proses *learning*. Selain itu, informasi yang relevan dengan peramalan perlu diolah dengan metode *fuzzy* dan hasilnya akan membentuk contoh peramalan. Kemudian, nilai hasil peramalan bisa diperoleh setelah contoh peramalan dimasukkan ke dalam model SVM yang telah terlatih sebelumnya (Du, et al., 2013).

Penelitian lain dilakukan oleh Yu, et al. (2013) untuk peramalan penjualan surat kabar atau majalah. Peramalan penjualan surat kabar ini sangat penting dilakukan oleh penerbit dalam mempersiapkan rencana pencetakan dan pendistribusian. Jumlah dari penjualan produk tersebut diasumsikan linier terhadap variabel-variabel yang relevan, misalkan demografis pelanggan atau jenis toko yang digunakan. Namun, asumsi seperti ini kurang tepat karena terdapat relasi *non linear* antara variabel penjelas dengan penjualan produk, sehingga algoritma SVR digunakan untuk mengatasi permasalahan tersebut. Pada saat melakukan pemrosesan data, data yang digunakan adalah 12 jenis toko berdasarkan lokasi toko tersebut, yang kemudian digunakan sebagai salah satu jenis variabel. Pekerjaan, pendidikan, umur, jenis kelamin, dan pendapatan juga digunakan sebagai variabel, dengan penjualan pada setiap toko sebagai variabel respons. Sebelum dilakukan pemodelan SVR, data penjualan tersebut dikelompokkan menjadi 3 kelas yaitu *low*, *media*, dan *High* dengan menggunakan 32 atribut data untuk mengetahui atribut yang paling penting pada setiap kelas. Selanjutnya, pemodelan SVR dilakukan dengan menggunakan 500 data (Yu, et al., 2013).

Pada penelitian ini, SVR akan digunakan sebagai solusi dalam meramalkan jumlah pemakaian air. Data pemakaian air akan dibagi menjadi data latih dan data uji, dengan data latih digunakan untuk proses *learning* sedangkan data uji digunakan untuk proses *testing*. Pada saat proses *learning*, terdapat beberapa parameter yang digunakan untuk membentuk sebuah model regresi. Model regresi yang sudah terbentuk dari hasil *learning*, selanjutnya akan diuji dan dihitung tingkat *error*-nya. Tujuan yang ingin dicapai adalah membentuk model regresi dengan nilai *error* terkecil.

Perbandingan objek dan metode penelitian yang telah dilakukan dengan penelitian yang akan dilakukan ditunjukkan pada Tabel 2.1.

Tabel 2.1 Perbandingan objek dan metode

No	Judul	Objek	Metode
1.	<i>Demand Forecasting of Perishable Farm Products Using Support Vector Machine</i> (Du, et al., 2013)	Permintaan hasil produksi pertanian	<i>Support Vector Machine</i>
2.	<i>Support Vector Regression For Newspaper/Magazine Sales Forecasting</i> (Yu, et al., 2013)	Penjualan surat kabar	<i>Support Vector Regression</i>
3.	Peramalan Jumlah Pemakaian Air di PT. Pembangkitan Jawa Bali Unit Pembangkit Gresik Menggunakan <i>Support Vector Regression</i>	Pemakaian air	<i>Support Vector Regression</i>

Untuk memperjelas perbandingan penelitian yang telah dilakukan dengan penelitian yang akan dilakukan, Tabel 2.2 berikut menunjukkan perbandingan *input*, proses, dan *output* dari masing-masing penelitian.

Tabel 2.2 Perbandingan data

Parameter	Penelitian 1	Penelitian 2	Penelitian 3
<i>Input</i>	<ul style="list-style-type: none"> Data historis penjualan apel dan pisang dari tiga pasar yang dipilih. Data latih berasal dari penjualan pada tanggal 1 Maret 2004 sampai dengan akhir Maret 2005. Data uji berasal dari penjualan pada tanggal 1 April 2005 sampai dengan akhir April 2005 	<ul style="list-style-type: none"> Data informasi penjualan surat kabar di beberapa toko pada tahun 2006 sampai dengan 2009 	<ul style="list-style-type: none"> Data pemakaian air periode Januari 2017 sampai dengan Juli 2017
Proses	1. Melakukan <i>preprocessing</i> data dengan pemerataan data untuk menghilangkan nilai	1. Melakukan <i>preprocessing</i> data dengan mengurangi dimensi data berdasarkan tingkat	1. Melakukan <i>preprocessing</i> data dengan melakukan normalisasi data dan menangani data

	<p><i>singular</i> dan normalisasi data untuk menghindari data <i>overflow</i>.</p> <p>2. Memasukkan sampel latih ke dalam model SVM untuk dilatih dan menyesuaikan parameter agar mencapai nilai optimal.</p> <p>3. Informasi yang relevan terhadap peramalan diproses dengan metode <i>fuzzy</i></p> <p>4. Melakukan proses peramalan</p>	<p>kepentingannya, sehingga dihasilkan variabel jenis toko dan demografis pelanggan.</p> <p>2. Melakukan transformasi data dengan memberikan skala 0 sampai dengan 1 pada semua variabel. Jenis toko dinilai dengan skala nominal dan demografis pelanggan diubah menjadi skala interval.</p> <p>3. Melakukan pengelompokan data menjadi tiga kelas yaitu <i>low</i>, <i>media</i>, dan <i>high</i>.</p> <p>4. Melakukan proses <i>learning</i></p> <p>5. Menguji model regresi yang telah dibentuk</p> <p>6. Menghitung tingkat <i>error</i></p>	<p><i>missing value</i></p> <p>2. Menentukan data pemakaian air yang akan digunakan sebagai data latih dan data uji</p> <p>3. Melakukan proses <i>learning</i></p> <p>4. Menguji model regresi yang telah dibentuk</p> <p>5. Menghitung tingkat <i>error</i></p>
<i>Output</i>	Permintaan hasil produksi pertanian	Penjualan surat kabar di setiap toko	Jumlah pemakaian air

2.2 Air

Air merupakan salah satu dari sumber daya yang menjadi salah satu dari berbagai jenis kebutuhan primer yang dibutuhkan oleh makhluk hidup. Ketersediaan air sangat mempengaruhi kelangsungan hidup tidak hanya untuk manusia, namun juga hewan dan tumbuhan. Hal tersebut membuktikan bahwa air digunakan untuk banyak tujuan dan dalam kuantitas yang besar. Dengan besarnya kebutuhan air, maka dalam penggunaannya perlu diadakan pengawasan agar tidak terjadi kekurangan air. Pada umumnya, manusia sering memanfaatkan dan menggunakan air untuk keperluan di toilet dan kegiatan rumah tangga lainnya. Menurut analisis yang dilakukan oleh Shan, et al. (2015), air paling sering digunakan untuk keperluan mandi. Selain itu, air juga digunakan untuk berkebun dan bertani (Shan, et al., 2015).

Sebelum digunakan untuk aktivitas manusia sehari-hari, air perlu disalurkan terlebih dahulu dari sumber-sumber air. Terdapat beberapa jenis sumber air yang menyediakan air untuk dapat digunakan antara lain sungai, danau,

penyimpanan air, air tanah, dan sumur (Magara & Morishita, 2007). Dari sumber air tersebut, selanjutnya air disalurkan ke PDAM yang kemudian dikelola agar dapat tersalurkan dengan baik ke konsumen.

Pada Peraturan Pemerintah No. 20 tahun 2009, disebutkan bahwa terdapat beberapa kelompok kualitas air yang dibagi sesuai dengan kegunaannya, yaitu sebagai berikut (Effendi, 2003):

- Golongan A: merupakan kelompok air yang dapat langsung dijadikan sebagai air minum, tanpa harus diolah terlebih dahulu.
- Golongan B: merupakan kelompok air yang menjadi bahan baku sebelum diolah menjadi air minum.
- Golongan C: merupakan kelompok air yang sering ditemukan untuk digunakan pada kegiatan yang berhubungan dengan perikanan dan peternakan.
- Golongan D: merupakan kelompok air yang digunakan pada bidang yang tidak dikonsumsi langsung oleh hewan ataupun manusia seperti pertanian, industri, dan pembangkit listrik tenaga air.

Berdasarkan pada hasil wawancara yang telah dilakukan dengan narasumber pada PT. PJB UP Gresik, salah satu faktor lokasi PT. PJB UP Gresik dekat dengan laut adalah agar dapat memanfaatkan air laut untuk digunakan dalam memproduksi energi listrik. Pada awalnya terdapat proses penyedotan air laut dan melakukan desalinasi untuk menghilangkan kandungan garam dari air laut tersebut. Setelah proses tersebut, air yang sudah terbebas dari kandungan garam disimpan di dalam tangki yang bernama *Raw Water Tank*. Air yang berasal dari *Raw Water Tank* digunakan untuk memenuhi kebutuhan operasional dari perusahaan. Selain itu, air dari *Raw Water Tank* juga diproses kembali untuk menghilangkan mineral-mineral yang dapat merusak mesin dan disimpan di dalam tangki yang bernama *Make Up Tank*. Air dari *Make Up Tank* yang akan digunakan untuk proses pembangkitan listrik.

Dengan adanya berbagai macam pemanfaatan air, keberadaan air menjadi sangat penting bagi perusahaan. Apabila terjadi kekurangan air, dapat mengakibatkan kerugian pada perusahaan. Hal tersebut dikarenakan air sangat berpengaruh pada kelancaran operasional perusahaan. Hasil dari wawancara dapat dilihat pada Lampiran A.

2.3 Peramalan

Peramalan adalah proses identifikasi dan memprediksi atau meramalkan sesuatu hingga mencapai suatu tingkat akurasi dengan menggunakan sebuah teknologi. Tujuan dilakukan peramalan adalah untuk membantu memperkirakan kejadian pada waktu yang akan datang berdasarkan fakta-fakta saat ini ataupun data-data dari waktu sebelumnya yang bisa menjadi dasar penentuan hasil. Terdapat beberapa penelitian mengenai peramalan yang telah dilakukan sebelumnya, contohnya peramalan tentang kemungkinan kerugian pada finansial

menggunakan *Non Linear Dynamics* dan SVR (Tung & Wong, 2009), peramalan nilai tukar mata uang menggunakan *Artificial Neural Network* (Divyapriya & Chezhian, 2013), dan peramalan mengenai cuaca per hari menggunakan *Artificial Neural Network* (Narvekar & Fargose, 2015),.

2.4 Support Vector Regression

Support Vector Regression (SVR) merupakan algoritma tingkat lanjut dari *Support Vector Machine* (SVM) dalam pembahasan tentang regresi. SVM adalah sebuah *learning machine* yang mengimplementasikan tentang bagaimana meminimalkan risiko struktural untuk mendapatkan generalisasi yang tepat pada jumlah yang terbatas pada *learning pattern*. Dengan mengimplementasikan algoritma *learning*, SVM sangat berguna untuk mengenali *pattern* yang samar atau halus di dalam kumpulan data kompleks. Pada algoritma ini, klasifikasi diskriminatif dilakukan berdasarkan contoh untuk memprediksi klasifikasi atas data yang sebelumnya tidak diketahui. SVM dapat menggeneralisasikan struktur data yang samar hanya dengan beberapa *support vector* (Basak, et al., 2007).

2.4.1 Model SVR

Prosedur regresi statistik sering disebut sebagai proses menurunkan sebuah fungsi $f(x)$ yang memiliki deviasi paling sedikit di antara respons yang terprediksi dan teramati secara eksperimen untuk semua contoh *training*. Salah satu karakteristik utama dari SVR adalah algoritma ini mencoba untuk meminimalkan batas *generalized error* dibandingkan dengan meminimalkan *training error* yang diamati, sehingga dapat mencapai *generalized performance*. Batas *generalization error* adalah hasil kombinasi dari nilai *training error* dengan *regularization term* yang bertugas untuk mengatur tingkat kompleksitas dari ruang hipotesis (Basak, et al., 2007).

Misalkan terdapat data latih $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\} \subset X \times R$, dengan X menyatakan ruang *input patterns* (seperti $X = R^d$). Tujuan dari ε -SV regresi adalah untuk menemukan fungsi $f(x)$ yang paling banyak ε deviasi dari target yang didapat sebenarnya y_i untuk semua data latih, dan pada saat yang sama dengan sedatar mungkin. Dalam kata lain, tidak perlu untuk memedulikan *error* yang bernilai kurang dari ε , namun tidak akan menerima deviasi lebih dari nilai ini (Scholkopf & Smola, 2004). Fungsi regresi $f(x)$ dinyatakan dalam bentuk persamaan umum yang ditunjukkan pada Persamaan 2.1.

$$f(x) = \langle w \cdot x \rangle + b \text{ dengan } w \in X, b \in R \quad (2.1)$$

Keterangan:

$f(x)$ = Nilai prediksi

b = Nilai bias, merupakan nilai kalkulasi dari *support vector*

w = Bobot

x = Data

$\langle \cdot \rangle$ = *Dot product* di dalam X .

Pada Persamaan 2.1, kedataran fungsi dicari dengan meminimalkan nilai w , yaitu dengan meminimalkan bentuk *Euclidean* seperti yang ditunjukkan pada Persamaan 2.2, dengan ε adalah nilai deviasi.

$$\text{Minimize } \frac{1}{2} \|w\|^2 \quad (2.2)$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

Asumsi pada Persamaan 2.2 adalah bahwa fungsi f benar-benar ada yang mendekati semua pasangan (x_i, y_i) dengan presisi ε , atau dengan kata lain, semua titik berada dalam rentang $f(x) \pm \varepsilon$.

Pada kondisi sebenarnya, ada beberapa titik yang mungkin berada di luar area $f(x) \pm \varepsilon$ sehingga perlu ditambahkan variabel *slack* ξ_i, ξ_i^* sebagai batasan baru untuk mengatasi kendala-kendala yang terdapat pada Persamaan 2.2. Kedataran fungsi dapat dicari dengan menggunakan Persamaan 2.3.

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} (\xi_i + \xi_i^*) \quad (2.3)$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Konstanta $C > 0$ merupakan harga batas toleransi antara kedataran fungsi f dengan jumlah deviasi yang nilainya melebihi dari nilai ε yang masih dapat diterima. Seperti yang disebut dengan ε -insensitive loss function $|\xi|_{\varepsilon}$ yang dinyatakan dalam Persamaan 2.4.

$$|\xi|_{\varepsilon} := \begin{cases} 0 & \text{jika } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{lainnya} \end{cases} \quad (2.4)$$

Setiap nilai kesalahan yang melebihi dari nilai ε akan dikenakan penalti sebesar C . Nilai C merupakan konstanta yang digunakan untuk menentukan penukaran kedataran fungsi dengan *error* pada saat proses *learning*.

2.4.2 Kernel Radial Basic Function (RBF)

Dalam penggunaan SVR untuk melakukan peramalan, kernel RBF sangat sering digunakan sebagai kernel dari SVR. Pada kernel ini, data non linier akan dipetakan ke dimensi yang lebih tinggi, sehingga dapat menangani permasalahan jika terjadi relasi antara kelas label dengan atribut non linier. Selain itu, kernel RBF memiliki parameter C dan γ yang digunakan untuk mentransformasikan data latih ke *feature space* (Yeh, et al., 2010). Proses perhitungan dengan menggunakan kernel RBF ditunjukkan pada Persamaan 2.5.

$$K(x, x') = \text{Exp}\left(\frac{-\|x - x'\|^2}{2\sigma^2}\right) \quad (2.5)$$

dengan $\|x - x'\|^2$; $x, x' \in R^d$ merupakan jarak antara dua data latih yang dihitung dengan menggunakan *euclidean distance* dan σ merupakan nilai varian data yang telah diinisialisasi sebelum proses *sequential learning* di dalam SVR.

2.4.3 Algoritma Sequential Learning

Proses *sequential learning* dilakukan untuk setiap perhitungan fungsi pada SVR. Langkah-langkah dari algoritma *sequential learning* dijabarkan sebagai berikut:

1. Melakukan inisialisasi pada nilai α_i dan α_i^* dengan masing-masing bernilai 0. Setelah itu dilakukan perhitungan matriks *hessian* seperti pada Persamaan 2.6.

$$[R]_{ij} = K(x_i, x_j) + \lambda^2 \text{ untuk } i, j = 1, 2, \dots, l. \quad (2.6)$$

Keterangan :

R_{ij} = Matriks *hessian* dengan baris i kolom j

λ^2 = *augmenting factor*

l = Banyaknya data latih yang digunakan

2. Untuk masing-masing data latih $i, j = 1, 2, \dots, l$, dilakukan:

2.1. Hitung nilai *error* dengan Persamaan 2.7.

$$E_i = y_i - \sum_{j=1}^l (\alpha_j^* - \alpha_j) R_{ij} \quad (2.7)$$

Keterangan:

E_i = Nilai *error* ke- i

y_i = Nilai aktual dari data latih

α_i dan α_i^* = Nilai *Lagrange Multiplier*

2.2. Melakukan perhitungan terhadap $\delta\alpha_i$ dan $\delta\alpha_i^*$ dengan Persamaan 2.8 dan 2.9.

$$\delta\alpha_i^* = \min \left\{ \max [\gamma(E_i - \varepsilon), -\alpha_i^*], C - \alpha_i^* \right\} \quad (2.8)$$

$$\delta\alpha_i = \min \left\{ \max [\gamma(-E_i - \varepsilon), -\alpha_i], C - \alpha_i \right\} \quad (2.9)$$

Keterangan:

$\delta\alpha_i, \delta\alpha_i^*$ = Variabel tunggal, merupakan perubahan nilai α_i dan α_i^*

ε = Nilai epsilon

C = Nilai kompleksitas

γ = Nilai *learning rate* yang didapatkan dari Persamaan 2.10

$$\gamma = \frac{\text{konstanta learning rate}}{\max(\text{diagonal matriks } R_{ij})} \quad (2.10)$$

2.3. Ubah nilai α_i dan α_i^* dengan Persamaan 2.11 dan 2.12.

$$\alpha_i^* = \delta\alpha_i^* + \alpha_i^* \quad (2.11)$$

$$\alpha_i = \delta\alpha_i + \alpha_i \quad (2.12)$$

3. Apabila proses sudah mencapai pada jumlah iterasi yang sudah ditentukan atau sudah memenuhi nilai $\max|\delta\alpha_i^*| < \varepsilon$ dan $\max|\delta\alpha_i| < \varepsilon$, maka algoritma selesai. Jika belum memenuhi syarat-syarat tersebut, maka ulangi kembali mulai dari langkah kedua.
4. Lakukan pengujian dengan menggunakan fungsi regresi seperti pada Persamaan 2.13.

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) (K(x_i, x) + \lambda^2) \quad (2.13)$$

5. Selesai.

2.5 Normalisasi Data

Data yang telah didapatkan untuk peramalan jumlah pemakaian air membutuhkan proses *preprocessing* data yaitu berupa normalisasi data. Dengan dilakukannya normalisasi pada data, data yang akan digunakan dalam perhitungan akan memiliki standarisasi sehingga selalu berada pada rentang tertentu. Proses normalisasi yang digunakan untuk penelitian ini adalah normalisasi *Min-Max* yang ditunjukkan pada Persamaan 2.14:

$$x' = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \quad (2.14)$$

Keterangan:

x' = Hasil dari normalisasi data

x = Nilai dari data yang akan dinormalisasi

x_{min} = Nilai terkecil dari seluruh data yang digunakan

x_{max} = Nilai terbesar dari seluruh data yang digunakan

Selain data bisa dilakukan normalisasi, proses pengembalian nilai data menjadi nilai semula juga bisa dilakukan dengan denormalisasi. Proses pencarian nilai semula (nilai x) ditunjukkan pada Persamaan 2.15:

$$x = (x'(x_{max} - x_{min})) + x_{min} \quad (2.15)$$

2.6 Mean Average Percentage Error (MAPE)

MAPE merupakan salah satu pengukuran dalam melakukan evaluasi terhadap akurasi dari hasil peramalan. Karena MAPE menyatakan kesalahan peramalan dari unit pengukuran yang berbeda menjadi kesalahan persentase pada hasil aktual, maka MAPE merupakan unit bebas. Dari berbagai macam pengukuran akurasi peramalan, MAPE sangat sering digunakan dalam peramalan yang menggunakan nilai aktual bersifat *non-negative*. Meskipun demikian, dalam peramalan praktis seperti peramalan *profit*, nilai aktual mungkin bernilai dengan nilai negatif (Ren & Glasure, 2009). Namun MAPE memiliki kelemahan yaitu menggunakan penalti yang besar pada eror positif dibandingkan dengan eror negatif (Hyndman & Koehler, 2006).

Untuk mengetahui apakah akurasi dari hasil peramalan masuk ke dalam kategori baik atau tidak, maka langkah pertama yang harus dilakukan adalah melihat tingkat akurasi dari metode peramalan sederhana (*naive forecasting*) kemudian membandingkan tingkat akurasi dari peramalan tersebut dengan tingkat akurasi yang dihasilkan jika menggunakan SVR. Metode peramalan sederhana dilakukan dengan cara menggunakan data nilai aktual pada waktu sebelumnya sebagai ramalan pada waktu selanjutnya. Apabila hasil akurasi dengan menggunakan SVR lebih tinggi dibandingkan dengan metode *naive forecasting*, maka dapat dikatakan bahwa hasil peramalan dari SVR yang telah dilakukan masuk dalam kategori kurang baik. Namun selain itu, bisa saja digunakan target performa dari tingkat akurasi yang biasa digunakan yaitu nilai $MAPE < 10\%$ termasuk dalam kategori sangat baik dan nilai $MAPE < 20\%$ termasuk dalam kategori baik (Gilliland, 2010). Proses perhitungan akurasi dengan menggunakan MAPE ditunjukkan pada Persamaan 2.16.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \times 100 \quad (2.16)$$