

SUPPORT VECTOR REGRESSION UNTUK PREDIKSI PRODUKTIVITAS KELAPA SAWIT DI PROVINSI RIAU

Mustakim¹, Agus Buono², Irman Hermadi³

¹ Jurusan Sistem Informasi, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
Email: mustakim@uin-suska.ac.id

^{2,3} Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor
Jl. Meranti Wing 20 Level 5-6, Bogor, 16680 Telp/Faks. +62 251 862558
Email: ²agusbuono@apps.ipb.ac.id, ³irmanhermadi@apps.ipb.ac.id

(Received: 7 April 2015; Revised: 2 Juni 2015; Accepted: 25 Juni 2015)

ABSTRAK

Krisis energi yang melanda wilayah Provinsi Riau dan sekitarnya memberikan dampak penurunan nilai ekonomi masyarakat, hal tersebut disebabkan salah satunya adalah negara selalu bergantung kepada sumber energi fosil. Upaya pemerintah untuk mengatasi krisis energi telah dilakukan dengan mengganti energi fosil dengan energi alternatif terbarukan dari limbah kelapa sawit. Produksi dan produktifitas kelapa sawit di Provinsi Riau memiliki peringkat terbesar di Indonesia, hal ini menjadi gambaran akan terwujudnya energi alternatif masa depan di Riau. Penelitian ini melakukan prediksi produksi dan produktifitas untuk kedepannya dengan menggunakan metode *Support Vector Regression*. Dari hasil percobaan yang dilakukan, diperoleh nilai prediksi untuk kedepannya mengalami penurunan hingga 52% dan hanya mengalami kenaikan 5 dari 74 lokasi sebesar 8% . Kernel SVR terbaik dari kombinasi percobaan ini adalah *Radial Basis Function*(RBF) dengan koefisien determinasi (R^2) sebesar 95% dan nilai *error* galat (MSE) sebesar 6% terdapat pada *fold* 1 dengan rentang $\gamma=2^0$ dan $C 2^3$. Hasil prediksi menunjukkan penurunan persentase produktifitas kelapa sawit untuk periode tahun berikutnya. Selisih data aktual dengan prediksi untuk rata-rata seluruh lokasi mencapai 10% dengan nilai penurunan hingga 52% dan 5 lokasi yang mengalami kenaikan sebesar 8% untuk periode selanjutnya.

Kata Kunci: Energi Terbarukan, Kelapa Sawit, *Radial Basis Function*, *Support Vector Regression*

ABSTRACT

The energy crisis hitting the Riau Province affects the economic impairment of society. One of the reason is that the country always depends on fossil energy resource. Government efforts to overcome the energy crisis has been done by replacing fossil energy with renewable alternative energy from palm oil waste. The production and productivity of palm oil in Riau Province has the highest rank in Indonesia, this becomes a future picture about realization of alternative energy in Riau. This research predicts on the future production and productivity by using Support Vector Regression. The results of experiments showed that the prediction value for the future have declined by 52% and increased only 5 out of the 74 locations at 8%. The best SVR kernel of this combination experiment are Radial Basis Function with a coefficient of determination (R^2) at 95% and an error value galat (MSE) at 6% in fold 1 ranging from $\gamma = 2^0$ and $C 2^3$. Research result for percentage prediction of productifity palm oil that descent next year. Between actual data and prediction data have average 10% for all location with descent 52% and sector 5 location have 8% for next period

Keywords: Oil Palm, *Radial Basis Function*, *Renewable Energy*, *Support Vector Regression*

Corresponding Author

Mustakim,
Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau,
Email: mustakim@uin-suska.ac.id

Pendahuluan

Perkembangan energi di Indonesia saat ini telah mengalami perubahan secara signifikan dari

negara pengekspor menjadi negara pengimpor. Berbagai kebijakan pemerintah yang ditetapkan untuk menyelamatkan energi terus dilakukan, namun hal itu dirasakan oleh berbagai pihak belum

secara maksimal. Krisis energi yang dialami sebagian besar wilayah pelosok Indonesia memberikan dampak negatif bagi pertumbuhan sektor ekonomi nasional. Setiap tahunnya krisis energi selalu melanda wilayah-wilayah penghasil energi utama di negara ini, tidak terlepas Pulau Kalimantan dan Provinsi Riau.

Gambaran pengelolaan energi nasional 2006-2025 menyatakan bahwa persoalan terbesar terkait energi di Indonesia yaitu struktur Anggaran Pendapatan dan Belanja Negara (APBN) masih bergantung kepada migas dan subsidi Bahan Bakar Minyak (BBM), industri energi yang belum optimal, keterbatasan infrastruktur energi, belum tercapainya harga ekonomi energi serta pemanfaatan energi belum efisien [5]. Hal demikian yang menyebabkan produktivitas energi dan perekonomian di Indonesia semakin melemah, terutama terkait permasalahan ketergantungan terhadap BBM yang semakin lama energi fosil akan semakin berkurang [18].

Terlepas dari hal di atas, berdasarkan Undang-undang No. 30/ 2007 tentang energi pasal 20 ayat 4 yaitu penyediaan dan pemanfaatan energi baru dan terbarukan wajib ditingkatkan oleh pemerintah pusat dan pemerintah daerah sesuai dengan kewenangannya. Energi terbarukan yang dimaksud dalam undang-undang tersebut salah satunya adalah energi biomassa dengan berbahan baku kelapa sawit [11]. Bahan baku kelapa sawit atau limbah kelapa sawit dan dengan teknologi konversi energi maka energi ini dapat digunakan untuk menghasilkan energi listrik sebagai sumber bahan bakar utama PLTU [13]. Energi listrik yang dihasilkan oleh konversi energi ini sangat bergantung dari limbah yang dihasilkan oleh pusat pengolahan minyak kelapa sawit [12].

Provinsi Riau merupakan sebuah Provinsi dengan luas 8,91 juta Ha sangat berpotensi untuk menghasilkan kelapa sawit terbesar di Indonesia karena memiliki lahan perkebunan seluas 2,26 juta Ha dengan rata-rata produksi sebesar 6,93 juta ton per tahun yang tersebar di berbagai kabupaten [4]. Produktivitas kelapa sawit di Provinsi Riau setiap tahun mengalami peningkatan baik produksi maupun luas lahan perkebunan. Hal ini menjadi gambaran akan terwujudnya energi alternatif berbahan baku kelapa sawit untuk masa mendatang dengan berbagai proyeksi limbah yang dihasilkan sebagai pengganti energi fosil. Dari sumber yang dirilis oleh Badan Pusat Statistik (BPS) terdapat nilai penurunan disuatu daerah, hal ini disebabkan karena perubahan/ penanaman ulang kelapa sawit yang telah mencapai batas usia produksi.

Keterkaitan antara energi alternatif masa depan dengan produksi kelapa sawit tidak terlepas dengan kondisi dan proyeksi sumber-sumber yang dihasilkan untuk jangka waktu yang lama. Kondisi

ini mampu memberikan gambaran efektifnya sebuah daerah dalam mengembangkan energi terbarukan, dengan sumber yang memadai dan tidak kurang pasokan bahan bakunya [11]. Oleh sebab itu, rencana strategis untuk mewujudkan pembangunan sumber bahan baku energi alternatif adalah dengan melakukan prediksi terhadap produktivitas bahan baku serta daerah penghasil bahan baku tersebut.

Perkembangan teknologi komputasi telah banyak menghasilkan model-model kecerdasan komputasional yang dapat diterapkan untuk penyelesaian berbagai kasus seperti prediksi. Berbagai riset telah banyak dilakukan dengan menerapkan metode *Support Vector Regression* (SVR) untuk menyelesaikan kasus prediksi dengan akurasi yang baik dibandingkan metode yang lainnya. SVR merupakan metode yang dapat mengatasi *overfitting*, sehingga akan menghasilkan performansi yang bagus [16], memberikan kesimpulan tentang keunggulan dan hasil akurasi [9] dan dapat diterapkan untuk berbagai kasus dengan data kontiniu [15]. Tahun 2009, produksi dan produktivitas kelapa sawit pernah diprediksi menggunakan metode *Artificial Neural Network* (ANN) akan tetapi tidak menyebutkan akurasi dari hasil prediksi tersebut [7].

Dalam penelitian ini akan dilakukan kombinasi percobaan berbasis *machine learning* menggunakan metode SVR untuk mengetahui prediksi kedepannya terkait produktivitas kelapa sawit sebagai sumber energi alternatif di Riau. Data training dan testing yang digunakan yaitu data produksi kelapa sawit yang berasal dari BPS Provinsi Riau dengan mengacu data 8 Tahun terakhir terdiri dari 74 Kecamatan. Tujuan dari penelitian ini yaitu mencari model terbaik untuk melihat pola data antara observasi dan prediksi, selain itu bagaimana perbandingan antara observasi data aktual dengan hasil prediksi yang diperoleh dari kombinasi percobaan pada kernel terbaik.

Metode Penelitian

Metode penelitian dimulai dari perencanaan, studi pustaka, pengamatan objek penelitian, wawancara kepada beberapa pihak yang terlibat, pengumpulan data. Kemudian melakukan percobaan metode SVR dengan data-data yang telah diperoleh, beberapa langkah dalam percobaan tersebut digambarkan pada Gambar 1.

Min-Max Normalization

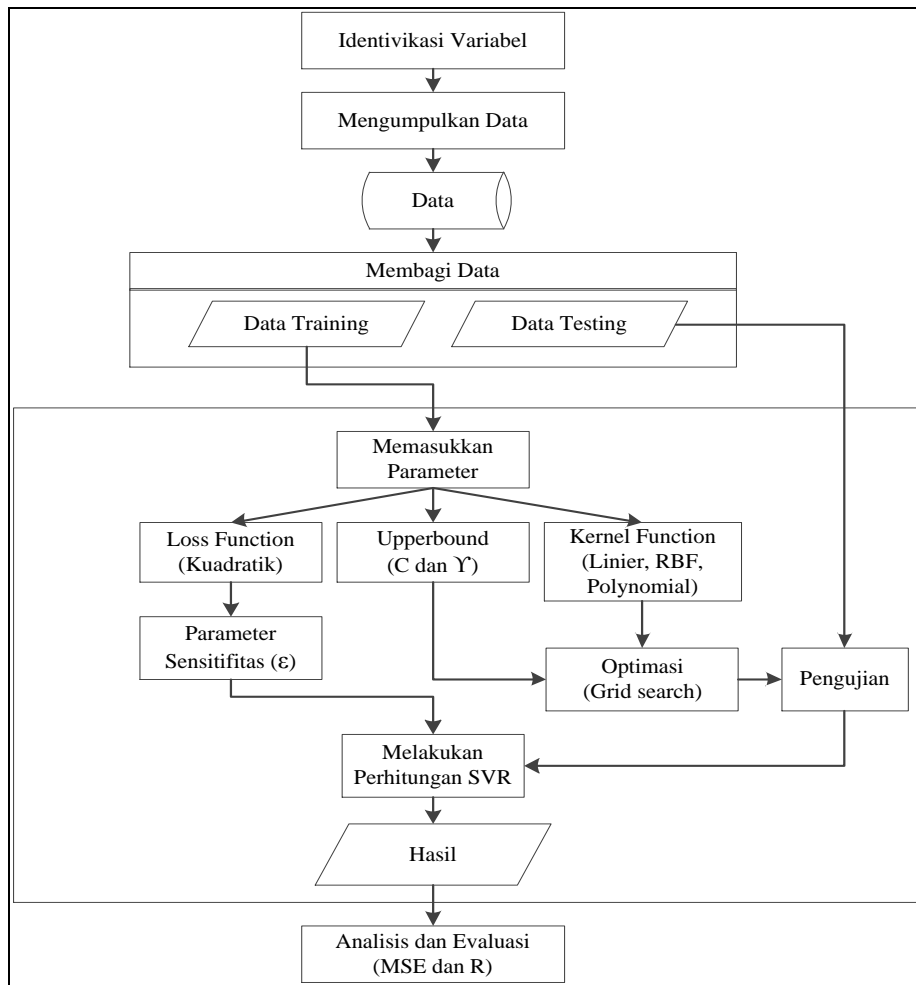
Normalization atau normalisasi data merupakan bagian dari data transformasi, yaitu teknik mengubah data menjadi nilai yang lebih mudah dipahami [14]. Tujuan normalisasi data

adalah mendapatkan bobot yang sama dari semua atribut data dan tidak bervariasi atau hasil dari pembobotan tersebut tidak terdapat atribut yang lebih dominan atau dianggap lebih utama dari pada yang lain.

Nilai rentang dari data memiliki ciri sangat bervariasi, sehingga nilai yang diperoleh harus diskalakan kedalam batas nilai tertentu supaya tidak terdapat dimensi data yang terlalu besar ataupun terlalu kecil yang dapat mempengaruhi hasil pengelompokan. Setiap nilai dalam data dikurangkan dengan nilai paling kecil dan dibagi

dengan nilai paling besar kurang nilai paling kecil, sehingga skala rentang nilai yang didapatkan berada pada $[0,0]$ hingga $[1,0]$. Min-max melakukan transformasi *linear* pada data, menggunakan nilai minimum dan nilai maksimum. Normalisasi min-max mempertahankan hubungan antar nilai data asli [10]. Dengan demikian Min-Max Normalisasi secara matematis dapat dituliskan dalam formula berikut:

$$v' = \frac{v - \min}{\max - \min} \quad (1)$$



Gambar 1. Langkah percobaan pada metode SVR

Support Vector Regression (SVR)

SVR merupakan penerapan *Support Vector Machine* (SVM) untuk kasus regresi. Dalam kasus regresi *output* berupa bilangan riil atau kontinyu. SVR merupakan metode yang dapat mengatasi *overfitting*, sehingga akan menghasilkan performansi yang bagus [22].

Smola dan Scholkopf (2004) menjelaskan mengenai SVR dengan memisalkan suatu kondisi yang terdapat λ set data training, (x_j, y_j) dengan $j=1,2,\dots,\lambda$ dengan input $x = \{x_1, x_2, x_3\} \subseteq \mathbb{R}^N$ dan output yang bersangkutan $y = \{y_1, \dots, y_\lambda\} \subseteq \mathbb{R}^N$. Dengan SVR, akan ditemukan suatu fungsi $f(x)$ yang mempunyai deviasi paling besar ϵ dari target aktual untuk semua data *training*. Maka dengan SVR ketika nilai ϵ sama dengan 0 akan

didapatkan regresi yang sempurna. Berdasarkan dari data, SVR ingin menemukan suatu fungsi regresi $f(x)$ yang dapat mengaproksimasi *output* ke suatu target aktual, dengan *error* toleransi- ϵ , dan kompleksitas yang minimal. Fungsi regresi $f(x)$ dapat dinyatakan dengan formula sebagai berikut [22]:

$$f(x) = w^T \varphi(x) + b \quad (2)$$

Dimana $\varphi(x)$ menunjukkan suatu titik didalam ruang fitur berdimensi lebih tinggi, hasil pemetaan dari input vektor x di dalam ruang input yang berdimensi lebih rendah. Koefisien w dan b diestimasi dengan cara meminimalkan fungsi resiko (*risk function*) yang didefinisikan dalam persamaan:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} L_{\epsilon}(y_i, f(x_i)) \quad (3a)$$

yang memenuhi:

$$\begin{aligned} y_i - w\varphi(x_i) - b &\leq \epsilon \\ w\varphi(x_i) - y_i + b &\leq \epsilon, i = 1, 2, \dots, \lambda \end{aligned} \quad (3b)$$

Dimana,

$$L_{\epsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \epsilon & |y_i - f(x_i)| \geq \epsilon \\ 0 & \text{untuk yang lain} \end{cases} \quad (3c)$$

Faktor $\|w\|^2$ dinamakan reguralisasi. Meminimalkan $\|w\|^2$ akan membuat suatu fungsi setipis mungkin, sehingga dapat mengontrol kapasitas fungsi. *Empirical error* diukur dengan ϵ -*insensitive loss function* yang diharuskan meminimalkan norm dari w agar mendapatkan generalisasi yang baik untuk fungsi regresi $f(x)$ [22]. Oleh karena itu diperlukan untuk menyelesaikan problem optimasi berikut:

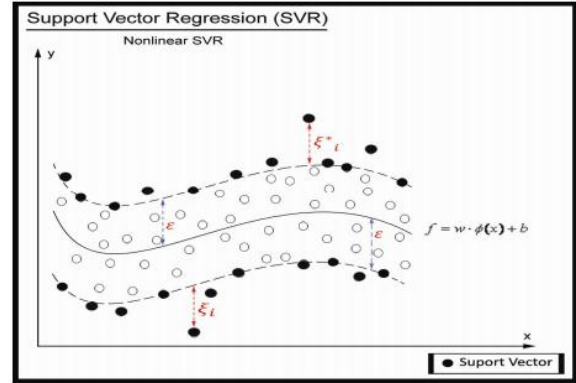
$$\min \frac{1}{2} \|w\|^2 \quad (4a)$$

yang memenuhi:

$$\begin{aligned} y_i - w\varphi(x_i) - b &\leq \epsilon \\ w\varphi(x_i) - y_i + b &\leq \epsilon, i = 1, 2, \dots, \lambda \end{aligned} \quad (4b)$$

Diasumsikan bahwa ada suatu fungsi f yang dapat mengaproksimasi semua titik (x^i, y^i) , dengan presisi ϵ . Dalam kasus ini kita asumsikan bahwa semua titik ada dalam rentang $f \pm \epsilon$ (*feasible*). Dalam hal ketidaklayakan

(*infeasible*), dimana mungkin ada beberapa titik yang mungkin keluar dari rentang $f \pm \epsilon$, perlu ditambahkan variable slack ξ, ξ^* untuk mengatasi masalah pembatas yang tidak layak (*infeasible constraint*) dalam problem optimasi seperti yang ditunjukkan pada Gambar 2 berikut.



Gambar 2 Penambahan variabel *slack* pada SVR
 (Sumber: Smola dan Schölkopf 2003)

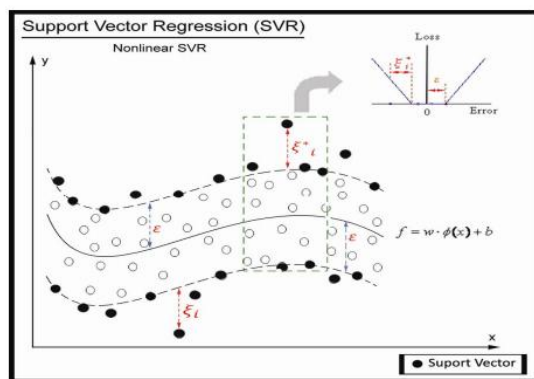
Selanjutnya problem optimasi di atas bisa diformulasikan sebagai berikut:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} (\xi_i, \xi_i^*) \quad (5a)$$

Dengan memenuhi:

$$\begin{aligned} y_i - w^T \varphi(x_i) - b - \xi_i &\leq \epsilon, i = 1, 2, \dots, \lambda \\ w\varphi(x_i) - y_i + b - \xi_i^* &\leq \epsilon, i = 1, 2, \dots, \lambda \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (5b)$$

Konstanta $C > 0$ menentukan tawar menawar (*trade off*) antara ketipisan fungsi f dan batas atas deviasi lebih dari ϵ yang masih dapat ditoleransi. Semua deviasi lebih besar daripada ϵ akan dikenakan pinalti sebesar C . Dalam SVR, ϵ setara dengan akurasi dari aproksimasi terhadap data training. Nilai ϵ yang kecil terkait dengan nilai yang tinggi pada variable slack ξ^* dan akurasi aproksimasi yang tinggi. Sebaliknya, nilai yang besar untuk ϵ berkaitan dengan nilai ξ^* yang kecil dan aproksimasi yang rendah seperti yang diilustrasikan oleh Gambar 3.



Gambar 3 Ilustrasi proses SVR
 (Sumber: Smola dan Schölkopf 2003)

Nilai yang tinggi untuk variabel *slack* akan membuat *empirical error* mempunyai pengaruh yang besar terhadap faktor regulasi. Dalam SVR, *support vector* adalah data *training* yang terletak pada dan diluar batas f dari fungsi keputusan, karena itu jumlah *support vector* menurun dengan naiknya ϵ (Smola dan Schölkopf 2003).

Terdapat 3 fungsi kernel pada model SVR, yaitu *Linear*, *Polynomial* dan *Radial Basis Function* (RBF). Fungsi-fungsi kernel tersebut dalam LIBSVM [21] diformulasikan sebagai berikut :

Kernel *Linear*

$$k(x,y) = x^T y + C \quad (6)$$

Kernel *Polynomial*

$$k(x,y) = (\alpha x^T y + C)^d \quad (7)$$

Kernel *Radial Basis Function* (RBF)

$$k(x,y) = \exp(-\gamma \|x - y\|^2) \quad (8)$$

Dalam regresi ada beberapa ukuran error yang sering dipakai untuk menilai performansi suatu fungsi prediksi. Jika y_i menyatakan nilai prediksi untuk data ke- i dan \hat{y}_i adalah nilai *output* aktual data ke- i dan m adalah banyaknya data, maka beberapa ukuran error yang sering dipakai adalah:

Mean squared error (MSE)

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad (9)$$

Mean absolute deviation (MAD)

$$MAD = \frac{\sum_{i=1}^m |\hat{y}_i - y_i|}{m} \quad (10)$$

Mean absolute percentage error (MAPE)

$$MAPE = \frac{\sum_{i=1}^m APE_i}{m} \quad (11)$$

Dimana

$$APE = \frac{\sum_{i=1}^m |\hat{y}_i - y_i|}{m} \times 100 \quad (12)$$

Dengan melihat salah satu atau lebih ukuran error diatas, dapat dipilih metode mana yang terbaik untuk suatu set data, semakin kecil nilai MSE, MAD atau MAPE maka semakin bagus[21].

K-Fold Cross Validation

K-fold cross validation dilakukan untuk membagi data menjadi *training set* dan *test set*. Inti dari validasi tipe ini adalah membagi data secara acak ke dalam k himpunan bagian yang diinginkan. *K-fold cross validation* mengulang k-kali untuk membagi sebuah himpunan secara acak menjadi k himpunan bagian yang paling bebas, setiap ulangan disisakan satu himpunan bagian untuk uji dan himpunan bagian lainnya untuk pelatihan [6].

K himpunan bagian dipilih satu himpunan bagian menjadi data pengujian dan (k-1) dijadikan sebagai data pelatihan. Namun, secara teori tidak ada tolak ukur yang pasti untuk nilai k. Keuntungan *k-fold cross validation* dibandingkan dengan variasi *cross validation* seperti *Repeated random sub-sampling validation* adalah semua data digunakan baik untuk data uji maupun data latih [3].

Grid Search Optimization

Metode *Grid search* merupakan salah satu metode yang sederhana untuk mengatasi permasalahan optimasi [19]. Metode ini melibatkan penyusunan *grid* yang cocok dalam suatu ruang dimensi, mengevaluasi fungsi objektif dari seluruh titik *grid*, dan menemukan titik *grid* yang sesuai dengan fungsi objektif yang memiliki nilai optimum. Prinsip kerja metode ini

secara sederhana adalah dengan menentukan beberapa nilai parameter pada rentang tertentu, kemudian memilih parameter pada nilai terbaik pada rentang tersebut dan melakukan pencarian berulang pada *grid*[8].

Hasil dan Pembahasan

Pemilihan Data dan Prediktor

Pemilihan data alternatif dilakukan dengan praposes data Kecamatan di Riau yang mengacu kepada aturan Standar Minimum Produksi (SMP) dari setiap wilayah. Beberapa perusahaan dan Dinas menetapkan SMP yang digunakan sebagai target untuk kelayakan bahan baku energi terbarukan adalah 1.000 Ton/ Periode atau minimum rata-rata HPP 1.000 Ton/ Tahun. Dari 142 Kecamatan hanya 74 Kecamatan yang memenuhi standar ketetapan minimum produksi yang berada diatas 1.000 Ton/ Periode. Dengan demikian data inilah yang dijadikan sebagai acuan dalam penelitian ini sebagai data prediktornya.

Setelah menetapkan dan memperoleh titik-titik data yang akan digunakan untuk melakukan prediksi, langkah selanjutnya adalah membagi data menjadi dua bagian yaitu *training set data* dan *test set data*. Pembagian ini dilakukan berdasarkan *k-fold cross validation* dengan membagi data secara acak ke dalam k himpunan bagian dengan semua data digunakan baik untuk data uji maupun data latih [3]. Data akan dibagi menjadi *4-fold* yang masing-masing *fold* terdiri dari 24 data latih dan 8 data uji. Ilustrasi data latih dan data uji terdapat pada Gambar 4 berikut:

<i>Fold-1</i>	Latih (1-8)	Latih (9-16)	Latih (17-24)	Uji (25-32)
<i>Fold-2</i>	Latih (1-8)	Latih (9-16)	Uji (17-24)	Latih (25-32)
<i>Fold-3</i>	Latih (1-8)	Uji (9-16)	Latih (17-24)	Latih (25-32)
<i>Fold-4</i>	Uji (1-8)	Latih (9-16)	Latih (17-24)	Latih (25-32)

Gambar 4. Ilustrasi data latih dan data uji pada *K-fold cross validation* yang digunakan

Penggunaan data yang akan diproses menggunakan SVR berdasarkan data normalisasi yang memprediksi dengan data dirinya sendiri dalam bentuk *time series*. Riset yang sama telah dilakukan oleh Ayuning Pramesthi Pintoasari tahun 2011 untuk prediksi alokasi *check-incounter* bandara Soekarno-Hatta 10 tahun kedepan, riset ini menggunakan data *time series* untuk memprediksi jumlah penumpang ditahun atau periode selanjutnya menggunakan data-data 10 tahun terakhir sebagai prediktornya [17].

Optimasi dan Fungsi Kernel SVR

SVR membutuhkan parameter yang sesuai dengan kernelnya dalam melakukan pelatihan. Untuk mendapatkan kernel yang optimal, pada saat pelatihan dilakukan optimasi menggunakan *grid search*. Terdapat 2 Parameter yang dioptimalkan dengan *grid search* yaitu parameter C dan parameter γ . Parameter C merupakan nilai pinalti terhadap galat model SVR, sedangkan parameter γ adalah parameter yang digunakan sebagai masukan untuk fungsi-fungsi kernel yang akan digunakan. Kernel RBF dan *polynomial* memerlukan parameter C dan γ , sedangkan pada kernel *linear* hanya memerlukan parameter C [1]. Untuk mencari nilai optimum dari parameter C dan γ dilakukan proses kombinasi percobaan pelatihan dan pengujian sebanyak 220 kali kombinasi percobaan untuk RBF, 44 kali kombinasi percobaan untuk *linear* dan 220 kali kombinasi percobaan untuk *polynomial* dengan variasi nilai parameter C dan γ sehingga dihasilkan sebuah model yang optimal. Selain dari parameter C dan γ , dilakukan pengujian dengan menerapkan parameter *New-SVR* dengan nilai 4. Kinerja model fungsi kernel dapat diketahui melalui nilai *corelation coefitient* (R) dan nilai galat *Mean Squared Error* (MSE). Model yang terbaik adalah model dengan nilai R terbesar (mendekati 1) dan nilai galat MSE terkecil (mendekati 0).

Simulasi yang dilakukan untuk mencari akurasi terbaik pada kernel RBF dan *polynomial* adalah dengan kombinasi C antara 2^{-6} sampai dengan 2^5 dan kombinasi γ antara 2^{-1} sampai dengan 2^4 dengan 220 kombinasi percobaan. Demikian juga untuk kernel *linear* yaitu antara 2^{-6} sampai dengan 2^5 kombinasi C. Model kombinasi ini juga dilakukan oleh Hendra Gunawan untuk mencari akurasi terbaik pada kasus prediksi produksi padi pada tahun 2012 yang menghasilkan akurasi diatas 95% [6].

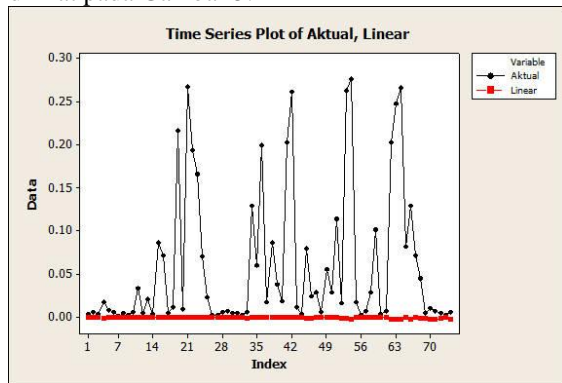
Kombinasi Percobaan Kernel SVR

Sesuai dengan simulasi yang jalankan dengan melakukan optimasi *grid search* pada *4 fold cross validation* diperoleh nilai MSE dan korelasi (R) beragam. Dari keempat *fold* memiliki nilai R atau dinyatakan dalam bentuk koefisien determinasi (R^2) maksimum 0,949855 pada *fold* 1 kernel RBF serta memiliki nilai MSE minimum 0,013948 pada *fold* 2 kernel RBF. Percobaan ini terus dilakukan sampai mendapatkan model terbaik.

Percobaan pada Kernel Linear

Pada kernel *linear* beberapa tahapan dan langkah yang dilakukan adalah dengan mengoptimalkan paramater C. Sesuai dengan penelitian-penelitian sebelumnya yang pernah dilakukan oleh beberapa peneliti, kernel *linear* merupakan kernel yang paling sederhana dibandingkan dengan yang lain. Melakukan kombinasi percobaan yang berada pada rentang 2^{-6} sampai dengan 2^5 menghasilkan MSE minimum 0,053308 atau 5% dengan R^2 maksimum 0,921253 atau 92%. Hal ini cukup memberikan akurasi tinggi pada penelitian ini jika tanpa membandingkan hasil percobaan kernel yang lain.

Prediksi diambil berdasarkan model terbaik hasil kombinasi percobaan yang telah dilakukan. Hubungan yang kuat antara observasi dan prediksi menunjukkan korelasi yang semakin kuat dan semakin kecil pula ukuran galat antara nilai yang diamati dan prediksi. Untuk lebih jelasnya dapat dilihat pada Gambar 5.

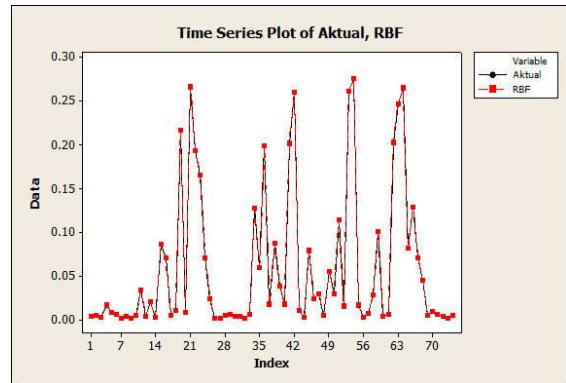


Gambar 5. Perbandingan hasil prediksi kernel *linear* dengan observasi pada produksi kelapa sawit dalam bentuk normal

Percobaan pada Kernel RBF

Berbeda dengan kernel *linear*, kernel RBF akan mengotimalkan nilai γ pada rentang 2^{-1} sampai dengan 2^4 dan parameter C pada rentang yang sama dengan kernel liner dapat dipeloreh nilai nilai galat terkecil pada *fold* 2 sebesar 1,4% dengan koefisien determinasi terbesar pada *fold* 1 sebesar 95%.

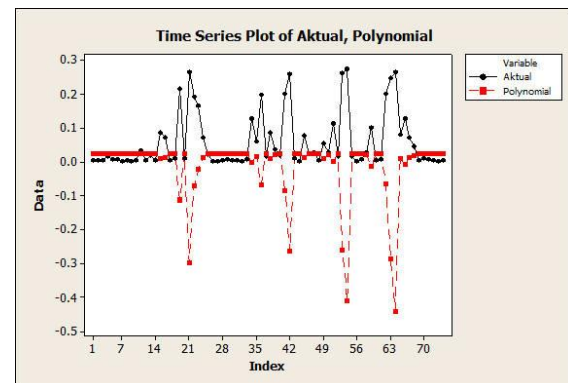
Hubungan antara observasi dan prediksi pada kernel RBF memiliki pola yang lebih baik dibandingkan dengan kernel *linear*, hal ini disebabkan bahwa kernel RBF merupakan kernel yang menghasilkan nilai koefisien kolerasi terbaik diantra yang lainnya. Gambar 6 berikut merupakan perbandingan antara hasil prediksi dan observasi pada kernel RBF.



Gambar 6. Perbandingan hasil prediksi kernel RBF dengan observasi pada produksi kelapa sawit dalam bentuk normal

Percobaan pada Kernel Polynomial

Sama halnya dengan RBF, kernel *polynomial* mengotimalkan nilai γ dan C dengan rentang yang sama seperti pada RBF. Percobaan terbaik pada *polynomial* terdapat pada *fold* 1 pada rentang γ dan C antara 2^{-1} dan 2^0 dengan nilai galat 18% dan koefisien determinasi sebsar 62%. Hubungan antara observasi dan prediksi pada kernel *polynomial* ditunjukkan pada Gambar 7 berikut:



Gambar 7. Perbandingan hasil prediksi kernel *polynomial* dengan observasi pada produksi kelapa sawit dalam bentuk normal

Apabila dilihat pada Gambar 7 diatas bahwa hasil prediksi semakin menjauhi dari nilai observasinya, makin tinggi nilai observasi maka semakin rendah nilai prediksi yang dihasilkan. Beberapa penelitian sebelumnya dengan kasus berbeda, kesimpulan yang dihasilkan dari prediksi *polynomial* merupakan kernel terburuk dibandingkan dengan dua kernel RBF dan *linear*.

Kernel SVR Terhadap Model Terbaik

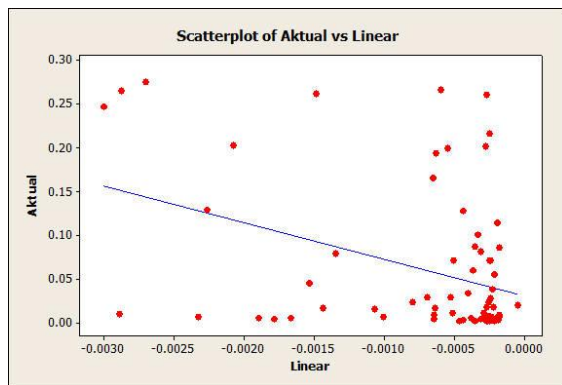
Pada kasus ini kernel yang paling tidak direkomendasikan adalah *polynomial* dan *linear*, karena kedua kernel ini memiliki hasil prediksi

yang memiliki rentang sangat jauh dengan observasinya. Sedangkan pada kernel RBF menghasilkan model terbaik diantara kernel yang lain. RBF memiliki pola yang mengikuti nilai observasinya serta menghasilkan selisih paling sedikit antara obeservasi dan prediksinya hal tersebut telah ditunjukkan dan dibuktikan pada Gambar 5, 6 serta 7 diatas. Nilai galat dan koefisien determinasi dari ketiga kernel tersebut dapat dilihat pada Tabel 1 berikut:

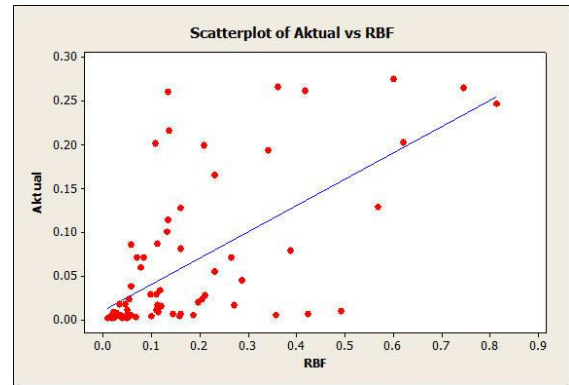
Tabel 1. MSE dan R² pada masing-masing kernel

Kernel	MSE	R ²
Linear	0,10	0,92
Radial Basis Function	0,06	0,95
Polynomial	0.18	0,62

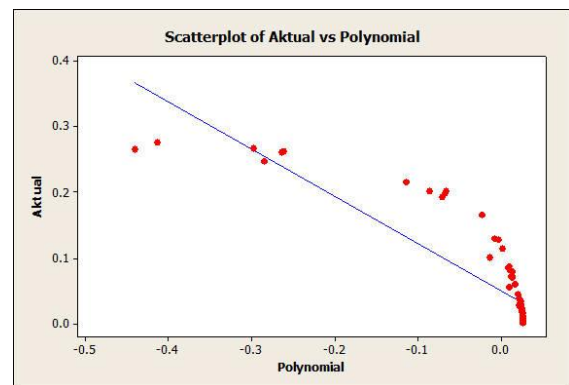
Dapat dilihat bahwa ketiga kernel tersebut memiliki rentang tidak terlalu jauh antara yang satu dengan yang lain. Pada nilai galat antara ketiganya memiliki selisih dibawah 15% dengan nilai terendah pada kernel RBF, sedangkan koefisien korelasi antara kedua kernel masih terbilang cukup tinggi dibandingkan kernel RBF yaitu memiliki selisih dibawah 35%. Penyebab ketiga kernel memiliki rentang korelasi yang masih tinggi dan hanya memiliki selisih yang tidak terlalu jauh yaitu terkait permasalahan data latih dan data uji yang kecil. Kurangnya kumpulan data yang besar dapat menyebabkan pola yang dibentuk tidak berfariatif. Berdasarkan skaterplot pola yang dibentuk dari ketiga kernel tersebut dapat ditunjukkan pada Gambar 8(a), 8(b) dan 8(c) berikut:



Gambar 8(a). Scatterplot hasil prediksi kernel *linear* dengan observasi



Gambar 8(b). Scatterplot hasil prediksi kernel *RBF* dengan observasi



Gambar 8(c). Scatterplot hasil prediksi kernel *Polynomial* dengan observasi

Jika ditelaah lebih mendalam, model prediksi yang menunjukkan tingkat korelasi tertinggi dan nilai galat terendah dari percobaan adalah dengan menggunakan kernel RBF. Hal ini sesuai dengan panduan SVM yang menjelaskan bahwa kernel RBF memang lebih unggul dalam kasus-kasus *machine-learning* pada umumnya [8].

Dari hasil percobaan yang dilakukan terkait kasus prediksi untuk periode selanjutnya terdapat 52% secara keseluruhan dari semua lokasi mengalami penurunan produksi hingga mencapai 10%. Sedangkan untuk nilai kenaikan hanya mencapai 8% terdapat pada 5 lokasi alternatif penghasil kelapa sawit. Dengan demikian dapat dikatakan untuk setiap periode selanjutnya produksi kelapa sawit di Riau rata-rata akan mengalami kenaikan yang sangat kecil. Jika dilihat dari data percobaan yang digunakan, keadaan tersebut memiliki kesamaan pada periode 2006 ke 2007 yang hanya mengalami kenaikan produksi yang sangat minim untuk keseluruhan wilayah.

Kesimpulan

Dari percobaan yang dilakukan diperoleh model terbaik SVR yaitu pada kernel RBF dengan koefisien determinasi (R^2) sebesar 95% dengan nilai *error* galat (MSE) sebesar 6% terdapat pada *fold* 1 dengan rentang $\gamma=2^0$ dan $C 2^3$. Penentuan optimasi dari parameter C pada setiap kernel yaitu semakin tinggi atau besar nilai C maka semakin besar pula nilai MSE yaitu pada C pada rentang 2^4 dan 2^5 dan γ pada rentang 2^{-5} dan 2^4 . Berdasarkan hasil perbandingan diperoleh model terbaik pada RBF yaitu pada *fold* 1 dengan nilai MSE sebesar 6% dan R^2 sebesar 95% dengan rentang nilai $\gamma=2^0$ dan $C 2^3$. Dalam percobaan ini juga diperoleh nilai MSE terkecil 1,4% pada *fold* 2 akan tetapi pada rentang terbaik hanya memiliki nilai koefisien kolerasi sebesar 67%. Oleh karena itu model diambil berdasarkan nilai koefisien diterminasi tertinggi yaitu pada *fold* 1.

Dari skaterplot dapat dilihat bahwa kernel *Linear* dan *Polynomial* lebih cenderung menjauhi data aktualnya dibandingkan dengan kernel RBF. Secara konsep dan matematis prediksi akan diambil berdasarkan model terbaik dengan pola data yang tidak saling berjauhan secara signifikan. Hasil prediksi menunjukkan penurunan persentase produktifitas kelapa sawit untuk periode berikutnya. Nilai selisih data aktual dengan prediksi untuk rata-rata seluruh lokasi mencapai 10% dengan nilai penurunan hingga 52% dan 5 lokasi yang mengalami kenaikan sebesar 8% untuk periode selanjutnya.

Daftar Pustaka

- [1] Adhani, G., Buono, A., Faqih, A., Support Vector Regression modelling for rainfall prediction in dry season based on Southern Oscillation Index and NINO3.4, *International Conference on Advanced Computer Science and Information Systems*, 2013, pp:315-320.
- [2] Apriani, I., Pemanfaatan Limbah Cair Pabrik Minyak Kelapa Sawit Sebagai Energi Alternatif Terbaharukan (Biogas), Tesis, Program Studi Teknologi Mesin Bioinformatika, Institut Pertanian Bogor, Bogor, 2009.
- [3] Buono, A., Agmalaro, M. A., Faqih, M., Statistical Downscaling Model Based-on Support Vector Regression to Predict Monthly Rainfall: A Case Study in Indramayu District, *Convergence Asian Federation for Information Technology in Agriculture*, 2013.
- [4] Badan Pusat Statistik Provinsi Riau, *Riau Dalam Angka 2006-2013 dan Kabupaten Dalam Angka 2006-2013*. 2013.
- [5] Elinur, Analisis Konsumsi dan Penyediaan Energi dalam Perekonomian Indonesia, Tesis, Program Studi Teknologi Industri Pertanian, Institut Pertanian Bogor, Bogor, 2011.
- [6] Gunawan, H., Seleksi Hyperspectral Band Menggunakan Recursive Feature Elimination untuk Prediksi Produksi Padi dengan Support Vector Regression, Tesis, Program Studi Ilmu Komputer, Institut Pertanian Bogor, Bogor, 2012.
- [7] Hermantoro, R. W. P., Prediksi Produksi Kelapa Sawit Berdasarkan Kualitas Lahan Menggunakan Model Artificial Neural Network (ANN), *Jurnal Agroteknose*, 4(2), 2009.
- [8] Hidayat, R., Sistem Prediksi Status Gizi Balita dengan Menggunakan Support Vector Regression, Skripsi, Program Studi Ilmu Komputer, Institut Pertanian Bogor, Bogor, 2013.
- [9] Ibrahim, N., Wibowo, A., Support Vector Regression with Missing Data Treatment Based Variables Selection for Water Level Prediction of Galas River in Kelantan Malaysia, *Wseas Transactions on Mathematics*, 13(1), 2014, E-ISSN: 2224-2880.
- [10] Jain, Y. K., Bhandare, S. K., Min Max Normalization Based Data Perturbation Method for Privacy Protection, *International Journal of Computer & communication Technology*, 2(8), 2011.
- [11] Kusdiana, D., Kondisi Rill Kebutuhan Energi di Indonesia dan Sumber-sumber Energi Alternatif Terbaharukan, Direktorat Jendral Listrik dan Pemanfaatan Energi Departemen Energi dan Sumber Daya Mineral, 2008.
- [12] Kusuma, I. P., Studi Pemanfaatan Biomassa Limbah Kelapa Sawit Sebagai Bahan Bakar Pembangkit Listrik Tenaga Uap di Kalimantan Selatan (Studi Kasus Kabupaten Tanah Laut). *Procciding Seminar Nasional Teknologi Industri*. 2011.
- [13] Mahajoeno, E., Pengembangan Energi Terbaharukan Dari Limbah Cair Pabrik Minyak Kelapa Sawit, Tesis, Program Studi Teknologi Industri Pertanian, Institut Pertanian Bogor, Bogor, 2008.
- [14] Patel, V. R., Mehta, R. G., Impact of Outlier Removal and Normalization Approach in

- Modified K-Means Clustering Algorithm, *IJCSI International Journal of Computer Science Issues*. 8(5), 2011.
- [15] Paz, J. F., A Support Vector Regression Approach to Predict Carbon Dioxide Exchange, A.P. de Leon F. de Carvalho et al. (Eds.): *Distrib. Computing & Artif. Intell.*, AISC 79, 2010, pp. 157–164.
- [16] Piantari, E., Feature Selection Data Hiperspektral Untuk Prediksi Produktivitas Padi dengan Algoritme Genetika Support Vector Regression, Skripsi, Program Studi Teknologi Ilmu Komputer, Institut Pertanian Bogor, Bogor, 2011.
- [17] Pintoasari, A. P., Penentuan Alokasi Check-In Counter Terminal Bandara untuk 10 Tahun kedepan dengan Metode Support Vector Regression, Tesis, Program Studi Ilmu Komputer, Universitas Indonesia, Depok, Indonesia. 2010.
- [18] Pudyantoro A. R., Dampak Kebijakan Fiskal dan Sektor Hulu Migas Terhadap Perekonomian Provinsi Riau, Tesis, Program Studi Teknologi Industri Pertanian, Institut Pertanian Bogor, Bogor, 2013.
- [19] Rao, S. S., *Engineering Optimization: Theory and Practice*, 2009, John Wiley and Sons, Newyork.
- [20] Saepudin, A., Energi Terbarukan (Biogas) dari Limbah Kelapa Sawit, Pusat Penelitian Tenaga Listrik dan Mekatronik, Lembaga Ilmu Pengetahuan Indonesia (LIPI), 2010.
- [21] Santosa, B., *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. 2007. Graha Ilmu. Yogyakarta.
- [22] Smola, A., Schölkopf, B., “A Tutorial on Support Vector Regression”, NeuroCOLT, Technical Report NC-TR-98-030, *Royal Holloway College, University of London UK*, 2003.