Penerapan Metode *Ridge Regression* dan *Support Vector Regression* (SVR) untuk Prediksi Indeks Batubara di PT XYZ

Rizky Amalia Putri, Wiwiek Setya Winahju, dan Muhammad Mashuri
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail:m_mashuri@statistika.its.ac.id

***Abstrak*—Semen merupakan salah satu bahan baku yang amat penting dalam pembangunan infrastruktur. Salah satu perusahaan yang bergerak di bidang produksi semen adalah PT XYZ. Tahapan terpenting dalam proses pembuatan semen adalah pada tahap pembakaran batu kapur dan tanah liat (*clinker*). Dalam proses pembakaran *clinker* membutuhkan ba-han bakar utama yaitu batubara. Semakin banyak jumlah produksi *clinker* yang dihasilkan dan semakin sedikit batubara yang digunakan dalam proses pembakaran, maka semakin efektif dan efisien proses produksi tersebut. Dalam penelitian ini akan dilakukan analisis untuk memprediksi indeks batu-bara dengan beberapa variabel yang diduga mempengaruhi ya-itu kualitas batubara, bahan baku, dan operasional yang kemu-dian akan dilakukan estimasi terhadap indeks batubara. Me-tode yang digunakan untuk mengestimasi indeks batubara ada-lah metode Regresi Ridge dan metode *Support Vector Regres-sion* (SVR). Model yang terbentuk dengan metode SVR akan dibandingkan dengan metode regresi *ridge* yang kemudian akan dipilih model terbaiknya diantara kedua model yang ter-bentuk menggunakan nilai RMSE. Hasil analisis didapatkan metode terbaik dengan nilai RMSE terkecil yaitu *Support Vec-tor Regression* (SVR) dan menggunakan kernel-polynomial yang menghasilkan parameter sigma bernilai 0,100 dan nilai c sebesar 1 dengan nilai RMSE sebesar 0,619.**

***Kata Kunci*—Batubara, Clinker, Indeks, Ridge, Support Vector Regression**

I. PENDAHULUAN

BIDANG pembangunan selalu berkembang sejalan de-ngan semakin banyaknya penduduk yang berada di Indonesia. Kebutuhan pembangunan infrastruktur yang paling besar adalah semen. Semen merupakan salah sa-tu bahan baku komoditas strategis yang penting dalam kehi-dupan pembangunan manusia modern. Pembangunan manu-sia modern identik dengan pembangunan infrastruk-tur[1].

Salah satu perusahaan yang bergerak di bidang produksi semen adalah PT XYZ. PT XYZ merupakan perusahaan manufaktur yang bergerak dibidang produksi semen, karena semen merupakan suatu produk yang sangat dibutuhkan dalam pembangunan infrastruktur, oleh karena itu dibutuhkan se-men yang berkualitas baik dengan proses produksi yang efektif dan efisien.

Proses pembutan semen di PT XYZ terdiri dari lima tahapan, yaitu penyediaan bahan mentah, penggilingan bahan mentah, pembakaran *clinker*, penggilingan akhir, lalu yang terakhir adalah pengantongan sak semen/pengemasan. Tahap pembakaran batu kapur dan tanah liat (*clinker*) merupakan salah satu tahapan penting dalam proses produksi semen. Proses pembakaran *clinker* membutuhkan bahan bakar utama yaitu batubara. Batubara adalah salah satu bahan bakar fosil yang berasal dari batuan sedimen yang dapat terbakar dan terbentuk dari endapan organik, utamanya adalah sisa-sisa tumbuhan dan terbentuk melalui proses pembatubaraan. Ke-menterian Energi dan Sumber Daya Mineral (ESDM) merilis data cada-ngan batubara Indonesia, yang kini mencapai 39,89 miliar ton. Angka ini naik dibanding sebelumnya yang sebesar 37 miliar ton pada awal 2018[2]. Indeks batubara pa-da PT XYZ didapatkan dari jumlah produksi *clinker* dibagi dengan jumlah batubara dengan satuan ton, dimana semakin tinggi indeks batubara maka semakin baik proses pemba-karan *clinker* atau dengan arti lain semakin banyak jumlah produksi *clinker* yang dihasilkan dan semakin sedikit batu-bara yang digunakan dalam proses pembakaran maka sema-kin efektif dan efisien penggunaan batubara pada proses pro-duksi tersebut.

Berdasarkan penelitian-penelitian sebelumnya belum pernah dilakukan penelitian terkait prediksi indeks batubara berdasarkan faktor-faktor yang diduga berpengaruh dalam proses produksi semen. Oleh karena itu pada penelitian ini akan dilakukan analisis untuk memprediksi indeks batubara dengan variabel yang digunakan adalah variabel indeks batu-bara dan beberapa variabel yang diduga mempengaruhi yang kemudian akan dilakukan estimasi terhadap indeks batubara. Metode yang digunakan untuk mengestimasi indeks batubara adalah metode *Ridge Regression* dikarenakan adanya kasus multikolinieritas dan *Support Vector Regression* (SVR) me-miliki tujuan memetakan vektor input ke dalam dimensi yang lebih tinggi. SVR juga digunakan karena beberapa proses produksi pada semen memiliki indikasi yang berhubungan pada variabel prediktor yang satu dengan yang lain, namun pada metode SVR tidak mempermasalahkan hal tersebut. Kemudian, model yang terbentuk dengan metode SVR akan dibandingkan dengan metode regresi *ridge* yang akan dipilih model terbaiknya diantara kedua model yang terbentuk.

II. TINJAUAN PUSTAKA

A. Regresi Ridge

Regresi ridge adalah regresi linier berganda engan kasus multikolinieritas. Penggunaan regresi ridge didahului oleh penggunaan regresi linier berdanda untuk mendeteksi multikolinieritas. Model dari regresi linier berganda adalah:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i, i = 1,2,3,\ldots,n \quad (1)$$

dengan

y_i = nilai variabel dependen pada observasi ke-i.

x_i = nilai varibael independen pada observasi ke-i.

p = banyaknya variabel independen x yang berpengaruh variabel dependen y.

ε_i = komponen galat yang diasumsikan berdistribusi normal dengan mean 0 dan memiliki variansi .

$\alpha, \beta_1, \ldots \beta_p$ = koefisien regresi.

Estimasi parameter regresi *linier* dinyatakan dalam bentuk sebagai berikut:

$$\hat{\beta} = (X^t X)^{-1} X^T Y \quad (2)$$

Sedangkan, estimasi parameter regresi *ridge* dinyatakan dalam bentuk sebagai berikut:

$$\hat{\beta} = (X^t X + cI)^{-1} X^T Y \quad (3)$$

Salah satu metode untuk mendeteksi multikolinieritas dengan menggunakan *ridge trace* (jarak *ridge*). Salah satu kesulitan dalam menggunakan *ridge trace* adalah membentuk nilai c yang tepat[3]. Metode ini bertujuan untuk mengatasi kondisi yang tidak diinginkan yang disebabkan oleh korelasi atau hubungan yang tinggi antara beberapa variabel singular atau tunggal sehingga menghasilkan nilai dugaan parameter model regresi yang tidak stabil[4]. Suatu acuan yang sering digunakan un-tuk melihat besarnya c adalah dengan melihat VIF dan melihat kecenderungan plot estimator *ridge trace.* Bila kenaikan VIF yang mendekati nilai satu menunjukkan bahwa variabel bebas tidak saling berkorelasi dengan variabel bebas lainnya.

Untuk mengetahui adanya hubungan linier antara varia-bel bebas dengan variabel terikat, dilakukan pengujian signifikansi parameter menggunakan hipotesis :

$H_0: \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$ (tidak ada hubungan linier antara variabel bebas dengan variabel terikat)

$H_1: \exists \beta_j \neq 0,\ j = 1,2,3,\ldots,k$ (ada hubungan linier antara variabel bebas dengan variabel terikat)

dengan tingkat signifikasi (α) tertentu pada uji statistik :

$$F_{hitung} = \frac{JKR/k}{JKR/(n-k-1)} \quad (4)$$

maka keputusan hipotesis dapat dilakukan penolakan H_0 jika $F_{hitung} > F_{tabel}$ atau gagal tolak H_0 jika $F_{hitung} \leq F_{tabel}$. Untuk mengetahui koefisien yang diperoleh berarti atau tidak, dilakukan pengujian signifikansi parameter menggunakan hipotesis :

$H_0 : \beta_i = 0$ (koefisien regresi tidak signifikan)

$H_1 : \exists \beta_i \neq 0$, i=1,2,3,...,k (koefisien regresi signifikan)

dengan tingkat signifikasi (α) tertentu pada uji statistik

$$t_{hitung} = \frac{b_i}{sb_i}, \quad (5)$$

maka keputusan hipotesis dapat dilakukan penolakan H_0 jika

$$|t_{hitung}| > t_{(\frac{\alpha}{2}, n-k-1)}$$

atau gagal tolak H_0 jika

$$|t_{hitung}| \leq t_{(\frac{\alpha}{2}, n-k-1)}$$

B. Support Vector Regression (SVR)

Tujuan dari Support Vector Regression (SVR) adalah un-tuk menemukan sebuah fungsi *f(x)* sebagai suatu hyperplane (garis pemisah) berupa fungsi regresi mana sesuai dengan semua input data dengan sebuah error dan membuat setipis mungkin[5]. Sebelum melakukan pemo-delan dengan SVR perlu dilakukan pre-processing data terlebih dahulu untuk mengidentifikasi adanya ketidak-lengkapan data yang terjadi karena pada saat pengum-pulan data terdapat instrumen yang rusak karena ke-salahan manusia (human error) ataupun kesalahan komputer[6].

Fungsi bernilai kontinyu yang sedang didekati dapat dituliskan dalam persamaan (7). Untuk data multidimensi, perlu menambah *x* satu per satu dan memasukkan *b* ke dalam vektor *w* untuk notasi secara matematis.

$$y = f(x) = < w, x > + b = \sum_{j=1}^{M} w_j x_j + b, y, b \in \mathbb{R}, x, w \in \mathbb{R}^M \quad (6)$$

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = \mathbf{w}^T \mathbf{x} + b \quad x, w \in \square^{M+1} \quad (7)$$

$D(x,y)=\pm\varepsilon$ adalah jarak terjauh *support vector* dari *hyperplane* disebut margin. Memaksimalkan margin akan meningkatkan probabilitas data ke dalam radius $\pm\varepsilon$. Jarak dari *hyperplane* $D(x,y)=0$ ke data (x,y) adalah $|D(x,y)|/\|\boldsymbol{w}\|$, dimana :

$$\boldsymbol{w} = (1 - \boldsymbol{w}^T)^T \quad (8)$$

Diasumsikan bahwa jarak maksimum data terhadap *hyperplane* adalah δ, maka estimasi yang ideal akan terpenuhi dengan :

$$|D(x,y)| \leq \delta \|\boldsymbol{w}\|$$
$$\frac{|D(x,y)|}{\|\boldsymbol{w}\|} \leq \delta$$
$$\delta \|\boldsymbol{w}\| = \varepsilon \quad (9)$$

Oleh karena itu untuk memaksimalkan margin δ, diperlukan $\| w \|$ yang minimum. Optimasi penyelesaian masalah dengan bentuk *Quadratic Programming* :

$$min \frac{1}{2} \|\boldsymbol{w}\|^2 \quad (10)$$

dengan syarat

$$y_i - \boldsymbol{w}^T \varphi(x_i) - b \leq \varepsilon \text{ untuk } i = 1, \ldots, k$$
$$\boldsymbol{w}^T \varphi(x_i) - y_i + b \leq \varepsilon \text{ untuk } i = 1, \ldots, k$$

Meminimalkan $\|\boldsymbol{w}\|^2$ akan membuat suatu fungsi setipis (*flat*) mungkin, sehingga bisa mengontrol kapasitas fungsi (*function* capacity). Diasumsikan bahwa semua titik ada dalam rentang $f(x)\pm\varepsilon$ (*feasible*), jika terdapat ketidak layakan (*infiesibility*) atau ada beberapa titik yang keluar dari rentang $f(x)\pm\varepsilon$, maka ditambahkan variabel slack ξ dan ξ^* untuk mengatasi masalah pembatasan yang tidak layak (*infeasible contraints*) dalam problem optimasi. Gambar 1 menjelaskan

bahwa semua titik diluar margin akan dikenai pinalti. Selanjutnya problem optimisasi di atas bisa diformulasikan sebagai berikut :

$$min\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}(\xi + \xi_i^*) \quad (11)$$

dengan syarat:

$$y_i - \boldsymbol{w}^T\varphi(x_i) - b - \xi_i \le \varepsilon, i = 1,\dots,l$$
$$\boldsymbol{w}^T\varphi(x_i) - y_i + b - \xi_i^* \le \varepsilon, i = 1,\dots,l$$
$$\xi_i, \xi_i^* \ge 0$$

Loss Function adalah fungsi yang menunjukan hubungan antara *error* dengan bagaimana error ini dikenai pinalti berupa ξ. Perbedaan *loss function* akan menghasilkan formula SVR yang berbeda.

$$L_\varepsilon(y, f(x,\mathbf{w})) = \begin{cases} 0 & |y-f(x,\mathbf{w})| \le \varepsilon; \\ |y-f(x,\mathbf{w})|-\varepsilon & \text{lainnya,} \end{cases} \quad (12)$$

$$L_\varepsilon(y, f(x,\mathbf{w})) = \begin{cases} 0 & |y-f(x,\mathbf{w})| \le \varepsilon; \\ (y-f(x,\mathbf{w})-\varepsilon)^2 & \text{lainnya,} \end{cases} \quad (13)$$

$$L(y, f(x,\mathbf{w})) = \begin{cases} c|y-f(x,\mathbf{w})|-\frac{c^2}{2} & |y-f(x,\mathbf{w})| > c; \\ |y-f(x,\mathbf{w})|-\varepsilon & |y-f(x,\mathbf{w})| \le c \end{cases} \quad (14)$$

Dengan konstanta *c>0* menentukan (*trade off*) antara ketipisan fungsi *(flatness of function) f(x)* dan batas atas deviasi yang lebih besar dari *ε* masih ditoleransi. Semua deviasi yang lebih besar dari *ε* akan dikenai pinalti sebesar *c*.

Pada pendekatan *soft margin* dapat ditambahkan variabel *slack x,x** untuk menjaga terhadap adanya *outlier*. Masalah optimisasi yang diperoleh *c* adalah regularisasi, sehingga parameter yang dapat disempurnakan yaitu yang memberikan bobot lebih untuk meminimalkan kerataan atau kesalahan untuk masalah optimisasi multi-tujuan ini.

Masalah optimisasi kuadratik bersyarat ini dapat diselesaikan dengan *Lagrangian*. Pengganda *Lagrange* de-ngan variabel ganda adalah $\lambda, \lambda^*, \alpha, \alpha^*$ dan bilangan *real* non-negatif. Fungsi yang dioptimalkan menjadi:

$$min\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi + \xi_i^*$$

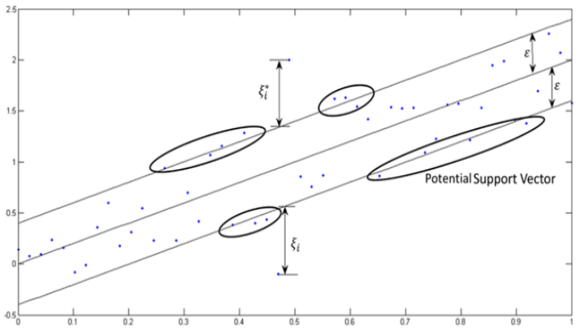
dengan

$$y_i - \boldsymbol{w}^T x_i \le \varepsilon + \xi_i^*, i = 1,\dots,N$$
$$\boldsymbol{w}^T x_i - y_i \le \varepsilon + \xi_i^*, i = 1,\dots,N$$
$$\xi_i, \xi_i^* \ge 0 \quad i = 1,\dots,N$$

Solusi optimasi adalah berdasarkan fungsi Lagrange berikut :

$$L(\mathbf{w},\xi^*,\xi,\lambda,\lambda^*,\alpha,\alpha^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i + \xi_i^* - \sum_{i=1}^{N}\alpha_i^*(y_i - \mathbf{w}^T x_i - \varepsilon - \xi_i^*)$$
$$+\sum_{i=1}^{N}\alpha_i(-y_i + \mathbf{w}^T x_i - \varepsilon - \xi_i) - \sum_{i=1}^{N}\lambda_i\xi_i + \lambda_i^*\xi_i^*$$

Notasi $L(\boldsymbol{w}, \xi^*, \xi, \lambda, \lambda^*, \alpha, \alpha^*)$ dinamakan *Lagrangian*. Untuk mendapatkan solusi yang optimal, maka dilakukan turunan parsial dari *L* terhadap $w, \xi, \xi^*, \lambda, \lambda^*, \alpha, \alpha^*$. Persamaan (16) didapatkan dengan mengambil turunan parsial yang disama dengankan nol. Pengganda *Lagrange* yang sama dengan nol sesuai dengan data di dalam tabung, sedangkan vektor dukungan memiliki pengganda *Lagrange* bernilai nol.



Gambar 1. Ilustrasi *Support Vector Regression*

$$\boldsymbol{w} = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)x_i$$

$$f(x) = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)x_i^T x, \alpha_i, \alpha_i^* \in [0, C]$$

$$\max_{\alpha,\alpha^*} -\varepsilon\sum_{i=1}^{N_{SV}}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{j=1}^{N_{SV}}\sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)x_i^T x_j,$$

dengan

$$\sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i) = 0, \alpha_i, \alpha_i^* \in [0, C] \quad (17)$$

Pada bagian ini, vektor bobot ditambah dengan skalar *b*, dan derivasi dari formulasi matematis SVR dilakukan, dengan mengabaikan perhitungan eksplisit dari *b*. Data training yang berada di luar batas tabung akan me-miliki nilai bukan nol α_i atau α_i^* ; keduanya tidak boleh nol. Selanjutnya, karena titik tersebut tidak berada di luar tabung $\xi_i=0$, sehingga akan mengarah ke hasil dalam Persamaan (19) bila $\alpha \in (0, C)$.

$$y_i - \boldsymbol{w}^T x_i - b - \varepsilon - \xi_i = 0 \quad (18)$$
$$y_i - \boldsymbol{w}^T x_i - b - \varepsilon = 0 \quad (19)$$
$$b = y_i - \boldsymbol{w}^T x_i - \varepsilon \quad (20)$$
$$-y_i + \boldsymbol{w}^T x_i - b - \varepsilon = 0 \quad (21)$$
$$b = -y_i + \boldsymbol{w}^T x_i - \varepsilon \quad (22)$$

Banyak teknik *machine learning* yang dikembangkan dengan asumsi kelinieran, sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Untuk fungsi non linier, data dapat dipetakan ke ruang dimensi yang lebih tinggi melalui φ, yang disebut ruang kernel untuk mencapai akurasi yang lebih tinggi[5].

$$min\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi + \xi_i^*, \quad (23)$$

dengan

$$y_i - \boldsymbol{w}^T\varphi(x_i) \le \varepsilon + \xi_i^*, i = 1,\dots,N$$
$$\boldsymbol{w}^T\varphi(x_i) - y_i \le \varepsilon + \xi_i, i = 1,\dots,N$$
$$\xi_i, \xi_i^* \ge 0, \quad i = 1,\dots,N$$
$$\boldsymbol{w} = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)\varphi(x_i) \quad (24)$$

$$\max_{\alpha,\alpha^*} -\varepsilon\sum_{i=1}^{N_{SV}}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{j=1}^{N_{SV}}\sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i - x_j) \quad (25)$$

$$\alpha_i, \alpha_i^* \in [0, C], i = 1,2,3,\dots,N_{SV}, \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i) = 0$$

$$f(x) = \sum_{i=1}^{N_{SV}}(\alpha_i^* - \alpha_i)k(x_i, x) \quad (26)$$

Tabel 1.
Variabel Penelitian

Simbol	Variabel	Satuan	Skala Data
Y	*Indeks* Batubara	ton/ton	Rasio
X_1	*Speed Kiln*	Rpm	Rasio
X_2	ILC *Exit Temperature*	°C	Rasio
X_3	SLC *Exit Temperature*	°C	Rasio
X_4	*Temperature Stage* 4 ILC	°C	Rasio
X_5	*Temperature Stage* 4 SLC	°C	Rasio
X_6	ILC *Coal*	Ton	Rasio
X_7	SLC *Coal*	Ton	Rasio
X_8	*Kiln Coal*	Ton	Rasio
X_9	LSF (*Lime Saturation Factor*)	-	Rasio
X_{10}	SIM (*Silica Modulus*)	-	Rasio
X_{11}	ALM *(Alumina Modulus)*	-	Rasio
X_{12}	AC *(Ash Content)*	%	Rasio
X_{13}	VM *(Volatile Matter)*	%	Rasio
X_{14}	FC *(Fixed Carbon)*	%	Rasio
X_{15}	TS *(Total Sulfur)*	%	Rasio
X_{16}	TM *(Total Moisture)*	%	Rasio
X_{17}	GHV *(Gross Heating Value)*	cal/g	Rasio

Nilai $K(x_i,x_j)$ merupakan fungsi kernel yang menunjukkan pemetaan linier pada *feature space* yang tidak selalu bisa di-ekspresikan secara eksplisit sebagai kombinasi antara α, y dan $\varphi(x)$. Fungsi kernel yang digunakan dalam penelitian ini adalah :

a. Kernel Linier (15)

$$\varphi(x) = K(x, x') = x^T x \quad (27)$$

b. Kernel Polynomial

$$\varphi(x) = K(x, x') = (\gamma(x^T x) + 1)^d \quad (28)$$

c. *Radial Basis Function* (RBF)

$$\varphi(x) = K(x, x') = exp(-\gamma||x - x_i||^2) \quad (29)$$

C. Kebaikan Model

RMSE merupakan alat untuk mengukur kebaikan suatu model berdasarkan pada *error* hasil estimasi. *Error* yang ada menunjukkan seberapa besar perbedaan hasil estimasi dengan nilai yang akan diestimasi dan nilai *mean square error* di-gunakan untuk menghitung tingkat error dari dua buah hasil percobaan model. Pengukuran RMSE dilakukan dengan ru-mus seperti berikut ini:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{n}} \quad (30)$$

RMSE : *Root Mean Square Error*
n : Jumlah Sampel
yi : Nilai Aktual
Yi : Nilai Prediksi Sesuai Model

D. Proses Produksi Semen

Proses pembutan semen di PT XYZ terdiri dari lima tahapan, yaitu penyediaan bahan mentah, penggilingan bahan mentah (batu kapur dan tanah liat), pembakaran clinker (campuran dari batu kapur dan tanah liat yang telah digiling), penggilingan akhir, lalu yang terakhir adalah pengantongan sak semen/pengemasan. Unit pembakaran merupakan bagian terpenting karena terjadi pembentukan komponen utama semen. Tujuan dari proses pembakaran ini ialah untuk menghasilkan *clinker* bermutu baik dengan pemakaian energi serendah mungkin dan operasi pembakaran berlangsung stabil dalam waktu yang lama. Salah satu faktor utama untuk mendapatkan hasil pembakaran yang baik ialah rancangan kiln feed (*raw mix design*) yaitu menentukan komposisi kimia dan ukuran partikel atau kehalusan dari *raw mix*. *Raw mix* dirancang untuk menghasil-kan clinker bermutu baik (mempunyai senyawa alite C3S, *belite* C2S, *aluminate* C3A, *ferrite* C4AF dalam jumlah cukup dan mudah digiling) yang dapat diukur dengan perhitungan modulus LSF, SIM, dan ALM. Selain itu, batubara juga me-miliki kualitas senyawa kimia yang terkandung didalamnya seperti AC (*Ash Content*), VM (*Volatile Matter*), FC (*Fixed Carbon*), TS (*Total Sulfur*), TM (*Total Moisture*), dan GHV (*Gross Heating Value*). Proses pada tahap ini meliputi pema-nasan awal umpan baku di *preheater* (pengeringan, dehidrasi dan dekomposisi), pembakaran di kiln (klinkerisasi) dan pendinginan di *grate cooler* (*quenching*). Selanjutnya klinker yang dihasilkan disimpan di *clinker* silo. Terdapat empat zona proses pemanasan yang terdiri dari pemanasan material yang baru masuk ke dalam *kiln* (*calsino zone)*, kemudian pemanasan material yang lebih tinggi (*tran-sisi zone*), pemanasan material secara penuh (*burning zone*), dan pendinginan secara cepat (*cooling zone*).

E. Indeks Batubara

Indeks adalah indikator ataupun ukuran atas sesuatu. Indeks batubara adalah suatu ukuran untuk mengukur kinerja proses produksi (clinker) yang dihasilkan dari tahap pem-bakaran batubara. Semakin banyak produksi clinker yang di--hasilkan dan semakin sedikit batubara yang digunakan dalam proses pembakaran maka semakin efektif dan efisien proses produksi tersebut. Indeks batubara didapatkan melalui rumus sebagai berikut:

$$\text{Indeks batubara} = \frac{\text{clinker}}{\text{raw coal}} \quad (31)$$

III. METODOLOGI PENELITIAN

A. Sumber Data

Pada penelitian ini sumber data yang akan digunakan merupakan data sekunder yang diperoleh pada produksi clinker dan pemakaian batubara bulan Januari 2019 - Juli 2019. Data diperoleh dari PT XYZ. Data yang digunakan sebanyak 213 observasi.

B. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini terdapat pada Tabel 1.

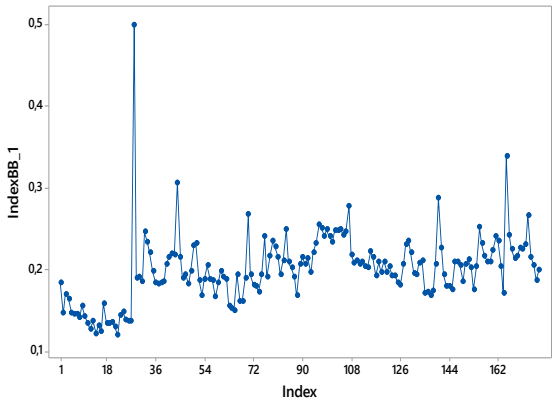
C. Langkah Analisis

Langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut.

1. Mengumpulkan data sekunder *indeks* batubara di PT XYZ.
2. Melakukan *pre-processing* pada data *indeks* batubara.
3. Melakukan eksplorasi data *indeks* batubara.

Tabel 2.
Hasil Uji Serentak (ANOVA)

Model	*Sum of Squares*	Df	*Mean Square*	F_h	*P-Value*
Regression	78,631	17	4,625	236,12	0,000
Error	2,233	114	0,019		
Total	80,865	131			



Gambar 2. Pola Data Indeks Batubara PT XYZ

4. Melakukan pembagian data t*raining* sebesar 90% dan data *testing* sebesar 10%.
5. Pemodelan regresi linear menggunakan persamaan (1) dan pemodelan regresi *Ridge*; (a)Pemodelan dengan regresi linear didahului dengan pemeriksaan hubungan antara semua variabel, jika terdapat hubungan antara variabel prediktor dengan respon dan telah diindikasikan terdapat kasus multi-kolinearitas maka pemodelan dilakukan mengguna-kan regresi *Ridge*; (b)Melakukan pemodelan regresi *Ridge*; (c)Pemeriksaan asumsi identik, independen, distribusi normal, dan asumsi non-multikolinearitas. Jika ter-dapat pelanggaran asumsi, maka dilakukan penang-gulangan menggunakan transformasi.; (d)Menghitung kebaikan model menggunakan nilai RMSE pada persamaan (32).
6. Melakukan pemodelan SVR dengan fungsi kernel di persamaan (28); (a)Melakukan pemodelan dengan SVR; (b)Melakukan *tuning* parameter untuk mendapatkan parameter optimum dari ketiga kernel; (c)Menentukan range nilai parameter C, ε dan γ untuk optimasi hyperplane pada data *training*; (d)Melakukan pemodelan dengan SVR berdasarkan range nilai parameter; (e)Mendapatkan model dan menghitung nilai RMSE dengan persamaan (32).
7. Pemilihan model terbaik menggunakan kriteria nilai RMSE di persamaan (32).
8. Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

Analisis dan pembahasan pada penelitian ini mencakup beberapa tahap yaitu eksplorasi data, pembagian data *training* dan *testing*, analisis regresi linier berganda serta identifikasi asumsinya, kemudian penggunaan metode *Support Vector Regression* (SVR) serta pemilihan model terbaik. Tahapan prediksi akan dilakukan setelah mendapatkan model terbaik.

A. Eksplorasi Data Indeks Batubara dengan Variabel Prediktor

Eksplorasi secara visiual dapat ditunjukkan dengan time series plot dan deskriptif. Berikut adalah pola indeks batubara dari bulan Januari 2019 hingga Juli 2019 di PT XYZ, dimana indeks batubara didapatkan dari jumlah produksi *clinker* dibagi dengan jumlah batubara dengan satuan ton.

Hasil plot pada Gambar 2 menunjukkan pola Indeks batubara PT XYZ pada bulan Januari 2019 hingga Juli 2019 mengalami fluktuasi. Pada akhir bulan Januari indeks batu-bara di PT XYZ mengalami peningkatan yang sangat tinggi yaitu sebesar 0,5, hal ini disebabkan adanya *trouble* pada alat perekap data. Data outlier selanjutnya dibuang untuk meng-atasi ketidaknormalan residual. Sebanyak 213 observasi di-lakukan pembuangan outlier, dan jumlah observasi setelah di-lakukan pembuangan outlier adalah sebanyak 177 observasi.

B. Pemodelan Menggunakan Ridge Regression

Pada penelitian ini dilakukan pemodelan *indeks* batubara dengan variabel prediktor yang diduga mempengaruhi meng-gunakan metode regresi linier berganda. Metode tersebut digunakan untuk mengetahui faktor-faktor yang memengaruhi indeks batubara di PT XYZ. Selanjutnya akan dihitung nilai RMSE yang kemudian akan dibandingkan dengan metode *Support Vector Regression* (SVR).

1) Tahapan Pemodelan Indeks Batubara Menggunakan OLS

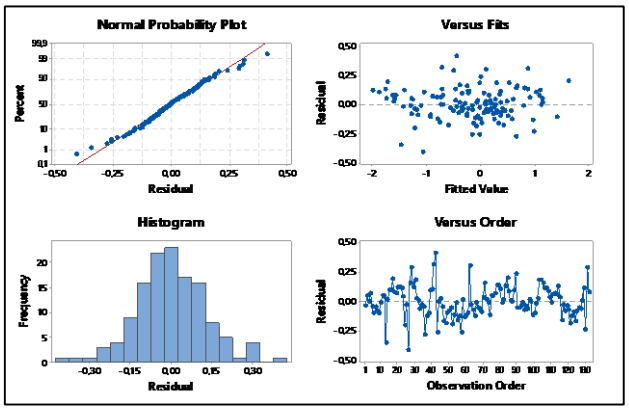
Pengujian signifikansi model regresi secara serentak dilakukan untuk menguji variabel prediktor secara bersama-sama berpengaruh terhadap model. Hipotesis yang digunakan adalah:

H_0: $\beta_1 = \beta_2 = \beta_3 = ... = \beta_{17} = 0$ (tidak ada hubungan linier antara variabel-variabel bebas dengan indeks batubara)

H_1: $\exists\beta_j \neq 0$, j = 1, 2, 3,..., 17 (minimal terdapat satu variabel yang berhubungan linier dengan indeks batubara). Dengan tingkat signifikansi $\alpha = 0{,}05$, pengujian serentak model regresi dapat dilihat pada Tabel 2 yaitu dari nilai deviansi yang terbentuk. Berdasarkan hasil analisis diperoleh nilai $D(\hat{\beta})$ sebesar 236,12 dan nilai $\chi^2_{(0,05;17)}$= 27,587.Keputusan pengujian ini tolak H_0 karena nilai $D(\hat{\boldsymbol{\beta}}) > \chi^2_{(0,05;17)}$ dan *p-value* kurang dari 0,05 yang berarti bahwa seluruh parameter secara bersama-sama mempunyai pengaruh dalam model dengan nilai R^2 sebesar 97,24%

Dalam analisis regresi, adanya kasus multikolinieritas dapat menimbulkan koefisien parsial regresi tidak terukur secara presisi sehingga nilai standar *error* besar. Oleh sebab itu, sebelum melakukan analisis lebih lanjut dengan regresi linier berganda, terlebih dahulu akan dilakukan pemeriksaan terhadap kasus multikolinieritas antar variabel prediktor. Salah satu cara untuk memeriksa adanya kasus multikolinieritas adalah melihat nilai VIF (*Variance Inflation Factors*), dimana nilai VIF yang lebih dari 10 merupakan indikasi adanya multikolinieritas. Berikut hasil pemeriksaan multikolinieritas disajikan pada Tabel 3.

Dapat dilihat pada Tabel 3 terdapat variabel prediktor yang memiliki nilai VIF lebih dari 10 yaitu pada variable X_4, X_5, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{16}, dan X_{17}, maka dapat disimpul-kan



Gambar 3. Asumsi IIDN Data Indeks Batubara

bahwa terjadi kasus multikolinieritas. Artinya terdapat ke-terkaitan antara variabel prediktor yang akan digunakan un-tuk memodelkan *indeks* batubara PT XYZ. Kesimpulan mo-del regresi OLS yang didapatkan kurang baik. Selanjutnya dilakukan identifikasi asumsi klasik, yaitu residual berdistri-busi normal, heteroskedastisitas, dan autokorelasi dengan ha-sil plot yang ada pada Gambar 3.

Dapat dilihat secara visual dari Gambar 3 bahwa pada pendeteksian asumsi kenormalan dapat dilihat dari gambar *normal probability* plot yang terlihat bahwa plot mengikuti garis kenormalan (warna merah), sehingga secara visual telah mengikuti asumsi distribusi normal, hal ini didukung oleh gambar histogram yang membentuk distribusi normal. Gambar versus fits menjelaskan bahwa data tidak mem-bentuk suatu pola dan tersebar merata sehingga secara visual dapat diputuskan bahwa data bersifat identik, sedangkan pa-da gambar versus order yang memperlihatkan secara visual bahwa data tersebar merata di atas maupun di bawah angka nol, sehingga secara visual dapat diputuskan bahwa telah memenuhi asumsi independen.

Adanya asumsi multikolinieritas yang tidak terpenuhi yang akan memberikan berbagai dampak yaitu penaksir OLS bersifat BLUE, tetapi mempunyai variansi dan kovariansi yang besar sehingga sulit mendapatkan taksiran (estimasi) yang tepat, maka kasus multikolinieritas perlu diatasi dengan menggunakan regresi *ridge*. Nilai VIF untuk masing-masing nilai c ($0 \le c \le 1$) dari data indeks batubara PT XYZ pada bulan Januari 2019 hingga Juli 2019 untuk setiap prediktor didapatkan nilai VIF semakin kecil jika ditambahkan tetapan bias (c), karena VIF yang optimal adalah bila nilai VIF<10 maka didapatkan nilai c=0,02 karena memiliki semua nilai VIF<10. Grafik VIF dengan berbagai nilai c ditunjukan oleh gambar berikut : Untuk mendapatkan nilai β^* yang sesuai dengan nilai c=0,02 yaitu dengan meggunakan persamaan regresi *ridge* atau dapat melihat tabel berikut :

$$Y^* = -0{,}3604\ Z_1 + 0{,}0897\ Z_2 + 0{,}0606\ Z_3 - 0{,}0553\ Z_4 - 0{,}1244\ Z_5 + 0{,}5498\ Z_6 + 0{,}5411\ Z_7 + 0{,}2113\ Z_8 - 0{,}0436\ Z_9 - 0{,}0606\ Z_{10} + 0{,}0687\ Z_{11} - 0{,}0347\ Z_{12} + 0{,}0105\ Z_{13} + 0{,}0471\ Z_{14} - 0{,}0121\ Z_{15} - 0{,}0438\ Z_{16} - 0{,}0302\ Z_{17}$$

Setelah didapatkan model Y^*, selanjutnya dilakukan transformasi ke bentuk awal dengan persamaan sebagai berikut.

Tabel 3.
Tuning Parameter Kernel-Polynomial

Degree	Sigma	C	RMSE
1	0,001	0,25	0,994
1	0,001	0,50	0,969
1	0,001	1,00	0,923
1	0,010	0,25	0,823
1	0,010	0,50	0,725
1	0,010	1,00	0,674
1	0,100	0,25	0,642
1	0,100	0,50	0,624
1	0,100	1,00	0,619
2	0,001	0,25	0,968
2	0,001	0,50	0,923
2	0,001	1,00	0,852
2	0,010	0,25	0,745
2	0,010	0,50	0,711
2	0,010	1,00	0,699
2	0,100	0,25	1,010
2	0,100	0,50	1,077
2	0,100	1,00	1,099
3	0,001	0,25	0,945
3	0,001	0,50	0,886
3	0,001	1,00	0,801
3	0,010	0,25	0,751
3	0,010	0,50	0,733
3	0,010	1,00	0,746
3	0,100	0,25	1,734
3	0,100	0,50	1,950
3	0,100	1,00	1,958

Sehingga model regresi *ridge* yang diperoleh adalah

$$Y_{RT} = -0{,}0773 - 0{,}5527X_1 + 0{,}1338X_2 + 0{,}1142X_3 - 0{,}1623X_4 - 0{,}2686X_5 + 0{,}4279X_6 + 0{,}4634X_7 + 0{,}3011X_8 - 0{,}0741X_9 - 0{,}0990X_{10} + 0{,}1076X_{11} - 0{,}0279X_{12} + 0{,}0141X_{13} + 0{,}0688X_{14} - 0{,}0098X_{15} - 0{,}0599X_{16} - 0{,}0408X_{17}$$

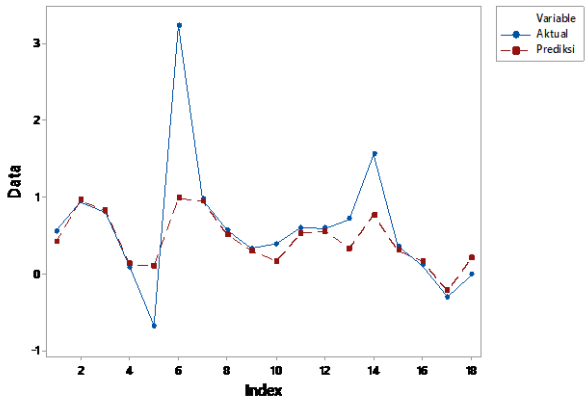
Selanjutnya, dilakukan perhitungan nilai RMSE dari model *Ridge Regression* dengan *c* sebesar 0,02, yang akan diguna-kan untuk dibandingkan dengan metode *Support Vector Reg-ression*. Hasil perhitungan RMSE menggunakan metode *Rid-ge Regression* dengan nilai *c*=0,02 adalah sebesar 2,088. Ni-lai RMSE tersebut akan dibandingkan dengan nilai RMSE metode *Support Vector Regression* (SVR).

C. Support Vector Regression (SVR)

Selanjutnya, dilakukan pemodelan menggunakan *Sup-port Vector Regression* (SVR) menggunakan tiga kernel, ya-itu kernel linier, polynomial, dan RBF dengan 17 variabel prediktor dan 1 variabel respon yaitu indeks batubara. Hasil yang didapatkan akan dibandingkan menggunakan nilai RMSE untuk mendapatkan model dengan parameter yang optimum.

Pembentukan model SVR disertai *tuning* parameter untuk mendapatkan prediksi parameter yang paling opti-mum. Untuk pemodelan kali ini menggunakan semua data indeks batubara dari bulan Januari 2019 hingga Juli 2019 de-ngan 17 variabel, pemodelan meliputi Kernel-Linier, Kernel-Polyno-mial dan Kernel-*Radial Basis Function* (RBF).

Kernel Linier merupakan sebuah metode yang ketika di-gunakan, data akan terpisah oleh sebuah garis linier yang di-sebut *hyperplane*. Hasil penggunaan kernel-linier dengan da-ta indeks batubara mendapatkan nilai parameter *cost* (C) ya-itu senilai 1, dengan nilai kriteria RMSE yaitu sebesar 0,6892864.



Gambar 4. Prediksi *Indeks* Batubara

Hal ini menyatakan bahwa dengan metode SVR menggunakan kernel-linier, model memiliki nilai akurasi prediksi sebesar 0,6892864, yang kemudian nilai ini akan di-bandingkan dengan metode lainnya.Kernel-polynomial merupakan kernel yang bersifat non-linier. Kernel ini memetakan suatu data ke dimensi yang di-namakan *feature space*. Kernel-polynomial memiliki fungsi khusus untuk memetakan ke *feature space* yang biasanya berbentuk kurva parabola. Berikut adalah hasil *tuning* parameter untuk kernel-polynomial.

Hasil *tuning* parameter pada Tabel 3 untuk kernel-polynomial, didapatkan hasil bahwa terdapat dua parameter yang digunakan yaitu sigma dan *cost* (C) dengan kriteria terbaik yaitu RMSE sebesar 0,619, didapatkan parameter yang paling optimal adalah 0,100 (sigma) dan 1 (*cost*) dengan nilai *degree* sebsar 1. Akurasi prediksi kernel-polynomial menggunakan model dan parameternya memiliki nilai prediksi sebesar 0,619 yang kemudian dilakukan perbandingan terhadap semua model yang telah didapatkan. Selanjutnya, dilakukan pemodelan SVR menggunakan kernel RBF (*Radial Basis Function*)

Kernel-RBF (*Radial Basis Function*) merupakan salah satu dari beberapa kernel untuk kasus non-linier. Kernel ini memetakan suatu data ke dimensi yang lebih tinggi dan membentuk kurva yang fleksibel sehingga dapat mengikuti pola data yang digunakan. Berikut ini adalah hasil dari *tu-ning* parameter dengan metode kernel-RBF.

Tabel 4 menunjukkan hasil *tuning* parameter untuk SVR dengan kernel-RBF dengan keseluruhan data *indeks* ba-tubara. Terdapat dua parameter untuk kernel-RBF yaitu sig-ma dan *cost* (C). Dengan kriteria RMSE sebesar 0,704 maka parameter yang terpilih dan paling optimal untuk data indeks batubara ini adalah sigma bernilai 0,01127152 dan nilai cost sebesar 1. Dengan model kernel-RBF ini, maka akurasi pre-diksi menggunakan model dan parameter terpilihnya memi-liki nilai sebesar 0,704, yang kemudian akan dibandingkan.

D.Pemilihan Model Indeks Batubara Terbaik

Pemilihan model terbaik dalam penelitian ini mengguna-kan kriteria RMSE. Model-model yang terbentuk akan di-bandingkan dengan kriteria tersebut. Model yang akan di-bandingkan adalah Regresi *Ridge* dan *Support Vector Reg-ression* (SVR). SVR terbagi menjadi tiga kernel yaitu kernel-linier, kernel-polynomial, dan kernel-RBF. Berikut adalah hasilnya.

Tabel 4.
Tuning Parameter Kernel-RBF

Sigma	C	RMSE
0,01127152	0,25	0,775
0,01127152	0,50	0,717
0,01127152	1,00	0,704
0,08480589	0,25	0,750
0,08480589	0,50	0,725
0,08480589	1,00	0,714
0,15834027	0,25	0,781
0,15834027	0,50	0,752
0,15834027	1,00	0,737

Tabel 5.
Pemilihan Model Terbaik

No.	Metode		C	RMSE
1.	Regresi Ridge		0,020	2,088
2.	*Support Vector Regression*	Linier	1	0,689
		Polynomial	1	0,619
		RBF	1	0,704

Tabel 5 memberikan informasi bahwa, dari beberapa model yang terbentuk, kriteria model terbaik adalah pada metode *Support Vector Regression* (SVR) dengan Kernel-Polynomial yang memiliki nilai kriteria RMSE bernilai 0,619, dengan parameter sigma bernilai 0,100 dan nilai *cost* sebesar 1.Dengan model kernel-polynomial yang terpilih sebagai metode terbaik dalam penelitian ini disertai parameter pada Tabel 5, akan dilakukan prediksi indeks batubara, berikut adalah hasilnya.

Gambar 4 memberikan informasi bahwa, terdapat 18 data indeks batubara. Secara visual nilai prediksi dengan mo-del SVR kernel-polynomial telah mengikuti data aktual, na-mun pada observasi keenam dan kesepuluh cukup memiliki perbedaan yang jauh pada besaran indeks batubara, namun untuk observasi yang lain, telah mengikuti data aktual, de-ngan prediksi ini didapatkan nilai kesalahan RMSE adalah sebesar 0,605. Nilai kesalahan RMSE pada metode SVR kernel polynomial ini lebih rendah dibandingkan dengan pemodelan *Ridge Reg-ression* dan SVR (kernel linier dan RBF).

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilaku-kan pada bab 4, maka diperoleh kesimpulan sebagai berikut; (1)Pola indeks batubara PT XYZ pada bulan Januari 2019 hingga Juli 2019 mengalami fluktuasi. Pada akhir bulan Januari indeks batubara di PT XYZ mengalami pening-katan yang sangat tinggi yaitu sebesar 0,5, hal ini di-sebabkan adanya *trouble* pada alat perekap data. Rata-rata indeks batubara PT XYZ pada bulan Januari 2019 hingga Juli 2019 adalah sebesar 0,199 dengan indeks batubara tertinggi sebesar 0,499 ton dan indeks

batubara terendah yaitu sebesar 0,119; (2)Pada hasil analisis regresi linier teridentifikasi adanya multikolinieritas, sehingga perlu ditangani menggunakan metode *Ridge Regression*. Kemudian dilakukan pemodel-an menggunakan metode *Support Vector Regression* (SVR) dengan fungsi kernel menghasilkan bahwa kernel polynomial menghasilkan nilai RMSE terkecil dibanding-kan dengan kernel linier, RBF, dan metode *Ridge Regres-sion*. Sehingga, metode yang terpilih adalah *Support Vec-tor Regression* (SVR) dengan kernel-polynomial. Hasil prediksi menunjukkan bahwa dengan model kernel-poly-nomial telah memprediksi indeks batubara di bulan Juli 2019 dengan baik, terlihat secara visual bahwa secara me-nyeluruh telah memprediksi dengan baik, walaupun ter-dapat beberapa yang berbeda jauh antara nilai prediksi dan nilai aktual. Hasil prediksi menggunakan SVR de-ngan kernel-polynomial mendapatkan nilai RMSE se-besar 0,619 dengan parameter sigma bernilai 0,100 dan nilai *c* sebesar 1. Sedangkan, nilai RMSE menggunakan metode lainnya menghasilkan RMSE yang lebih tinggi, yaitu metode *ridge regression* menghasilkan RMSE se-besar 2,088, metode SVR menggunakan kernel-linier se-besar 0,689, metode SVR menggunakan kernel-RBF. Se-hingga, metode terbaik dalam memprediksi indeks batu-bara PT XYZ adalah SVR dengan kernel-polynomial.

B. Saran

Berdasarkan kesimpulan yang diperoleh, dapat dirumus-kan saran sebagai pertimbangan penelitian selanjutnya ada-lah sebagai berikut; (1)Untuk meningkatkan indeks batubara PT XYZ, perlu dilakukan optimasi terhadap *speed kiln*, ILC *Exit tempera-ture*, SLC *Exit Temperature*, *Temperature Stage* 4 SLC, ILC *Coal*, SLC *Coal*, dan *Kiln Coal;* (2)Pada pola *time series* plot pada akhir bulan Januari indeks batubara di PT XYZ mengalami peningkatan yang sangat tinggi yaitu sebesar 0,5, hal ini disebabkan adanya *trou-ble* pada alat perekap data, sehingga PT XYZ disarankan untuk lebih mengontrol kinerja dari alat perekap data un-tuk menghindari ada-nya *missing* dalam pencatatan ob-servasi; (3)Dapat ditambahkan variabel yang diduga berpengaruh terhadap indeks batubara PT XYZ supaya informasi yang didapatkan lebih lengkap, misalnya faktor cuaca yang mempengaruhi kadar air dalam batubara.

DAFTAR PUSTAKA

[1] A. A. A. Hidayat;, *Metode Penelitian Keperawatan Dan Teknik Analisis Data*. Jakarta: Salemba Medika, 2009.

[2] A. Arvirianty, "Cadangan Batu Bara RI Naik Jadi 39,89 Miliar Ton," *cnbcindonesia.com*, 24-Jun-2019. [Online]. Available: https://www.cnbcindonesia.com/news/20190624120905-4-80160/cadangan-batu-bara-ri-naik-jadi-3989-miliar-ton. [Accessed: 15-Jun-2020].

[3] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Applications to Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970, doi: 10.1080/00401706.1970.10488635.

[4] N. R. Draper and H. Smith, *Applied Regression Analysis, 3rd Edition*, 3a ed. New York: John Wiley & Sons, 1998.

[5] N. Cristianini and B. Schölkopf, "Support vector machines and kernel methods: The new generation of learning machines," *AI Magazine*, vol. 23, no. 3, pp. 31–41, 2002.

[6] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. USA: Elsevier, 2012.