ANALISIS AKURASI SUPPORT VECTOR MACHINE DENGAN FUNGSI KERNEL GAUSSIAN RBF UNTUK PRAKIRAAN BEBAN LISTRIK HARIAN SEKTOR INDUSTRI

Luqman Assaffat[1]*
[1] Jurusan Teknik Elektro, Fakultas Teknik, Universitas Muhammadiyah Semarang
Jl. Kasipah no. 12, Semarang 50254
* assaffat@unimus.ac.id

Abstrak

Keakuratan prakiraan beban listrik harian di sektor industri memegang peranan dalam penghematan energi listrik. Penghematan energi listrik dapat dilakukan dengan pengaturan operasional industri berdasarkan laporan prakiraan beban listrik listrik tersbut. Salah satu metode yang berhasil di dalam prediksi beban listrik adalah Suppor Vector Machine dengan berbagai macam fungsi Kernel yang mendukungnya. Penelitian ini bertujuan menganalisis akurasi sistem peramalan beban listrik harian yang diterapkan pada sektor industri menggunakan Support Vector Machine (SVM) dengan fungsi Kernel Gaussian RBF. Data penelitian ini merupakan data beban listrik harian pada salah satu industri farmasi terkemuka di Indonesia, yaitu PT. Phapros Indonesia selama tahun 2014. Untuk mendukung keakuratan penelitian ini, parameter data latih SVM tidak hanya berasal dari data times series beban listrik, tetapi juga berasal dari data kapasitas produksi dan jenis hari kerja.

***Kata kunci**: support vector machine, fungsi kernel, gaussian RBF, peramalan beban listrik harian, nilai error prediksi, MAPE*

PENDAHULUAN

Ditinjau dari sudut pandang ekonomi, energi listrik merupakan salah satu faktor terbesar yang berpengaruh dalam pertumbuhan ekonomi. Untuk mendukung pertumbuhan ekonomi yang berkelanjutan di masa mendatang, prakiraan beban listrik menjadi suatu hal yang harus diperhatikan (Hsu dkk, 2003). Suatu prakiraan beban listrik yang akurat merupakan alat yang sangat penting dalam hal pendukung keputusan di bidang energi listrik (Hahn dkk, 2009). Keakuratan prakiraan beban listrik menjadi kunci dalam perencanaan sistem tenaga listrik. Hal ini digunakan sebagai alat untuk menentukan jadwal waktu perawatan sistem tenaga listrik, pembangunan fasilitas pembangkitan baru, perencanaan pembelian mesin-mesin pembangkit baru, dan pengembangan sistem transmisi dan distribusi tenaga listrik (Al-Hamadi, 2005).

Terdapat beberapa pendekatan yang dapat digunakan di dalam prakiraan beban listrik, antara lain yaitu metode regresi dan *time series* klasik dan metode kecerdasan buatan. Salah satu metode kecerdasan buatan yang sering digunakan dan terkenal yaitu metode jaringan syaraf tiruan (ANN), sedangkan metode kecerdasan buatan yang relatif baru yaitu metode *support vector machine* (SVM) (Hahn, 2009).

Support vector machine (SVM) merupakan salah satu algoritma *Machine learning* yang paling populer untuk klasifikasi dan regresi (Maali dkk, 2013). Selama dekade terakhir, SVM menjadi metode yang kuat untuk pola klasifikasi dan regresi, memiliki tingkat keberhasilan yang tinggi saat diterapkan di berbagai bidang (Qi dkk, 2013). Sehingga banyak dari kalangan komunitas *Machine learning* berminat untuk mempelajari dan mengembangkan SVM karena kinerjanya yang sangat baik dalam berbagai masalah pembelajaran (Ceperic dkk, 2012).

METODOLOGI

Konsep SVM

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space* (Vapnik, 1999). *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. Margin merupakan jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing kelas. Subset data training set yang paling dekat ini disebut sebagai *support vector*. Upaya mencari lokasi *hyperplane* optimal ini merupakan inti dari proses pembelajaran pada SVM (Cristiani, 2000).

Diasumsikan kedua class –1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d, yang didefinisikan :

$$\vec{w} \cdot \vec{x} + \quad (1)$$

Sebuah pola yang termasuk kelas –1 (sampel negatif) dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x}_i + b \quad (2)$$

Sebuah pola yang termasuk kelas +1 (sampel positif) dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x}_i + b \quad (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) problem, yaitu mencari titik minimal persamaan (4), dengan memperhatikan *constraint* persamaan (5).

$$\min \tau(w) = \frac{1}{2} \quad (4)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq \quad (5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya *Lagrange multiplier* sebagaimana ditunjukkan pada persamaan (6).

$$L(\vec{w}, b, \alpha) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{l} \alpha_i \left(y_i\left((\vec{x}_i \cdot \vec{w}) + b\right) - 1\right) \quad (6)$$

Fungsi Kernel

Pada umumnya masalah dalam domain dunia nyata (*real world problem*) jarang yang bersifat *linear separable* dan kebanyakan bersifat *non linear*. Untuk menyelesaikan permasalahan *non linear*, SVM dimodifikasi dengan memasukkan fungsi Kernel. Dalam *non linear* SVM, pertama-tama data dipetakan oleh fungsi ke ruang vektor yang berdimensi lebih tinggi. Pada ruang vektor yang baru ini, *hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan. Hal ini sejalan dengan teori Cover yang menyatakan "Jika suatu transformasi bersifat *non linear* dan dimensi dari *feature space* cukup tinggi, maka data pada *input space* dapat dipetakan ke *feature space* yang baru, di mana pola-pola tersebut pada probabilitas tinggi dapat dipisahkan secara linear" (Smola,2004).

Kernel Trick, yang dirumuskan :

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\quad (7)$$

Kernel trick memberikan berbagai kemudahan, karena dalam proses pembelajaran SVM, untuk menentukan *support vector*, dan hanya cukup mengetahui fungsi Kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi *non linear*. Berbagai jenis fungsi Kernel dikenal, sebagaimana dirangkumkan pada Tabel 1.

Selanjutnya hasil klasifikasi dari data diperoleh dari persamaan berikut :

$$f(\Phi(\vec{x})) = \vec{w}.\Phi(\quad (8)$$

$$f(\Phi(\vec{x})) = \sum_{i=1, \vec{x}_i \in SV}^{n} \alpha_i y_i \Phi(\vec{x}).\Phi(\vec{x}_i) + b \quad (9)$$

$$f(\Phi(\vec{x})) = \sum_{i=1, \vec{x}_i \in SV}^{n} \alpha_i y_i K(\vec{x}, \vec{x} \quad (10)$$

Persamaan di atas dimaksudkan dengan subset dari *training set* yang terpilih sebagai *support vector*, dengan kata lain data yang berkorespondensi.

Tabel 1. Jenis Fungsi Kernel

Jenis Kernel	Definisi
Polynomial	$K(\bar{x}_i, \bar{x}_j) = K(\bar{x}_i, \bar{x}_j + 1)^p$
Gausian	$K(\bar{x}_i, \bar{x}_j) = exp\left(-\frac{\|(\bar{x}_i, \bar{x}_j)\|^2}{2\sigma^2}\right)$
Sigmoid	$K(\bar{x}_i, \bar{x}_j) = tanh(\alpha \bar{x}_i, \bar{x}_j \beta)$

Koefisien Korelasi Pearson's (r)

Rumus koefisien korelasi Pearson's (r) digunakan pada analisis korelasi sederhana untuk variabel interval/rasio dengan variabel/rasio. Koefisien Korelasi Pearson's (r) dirumuskan sebagai berikut (Misbahuddin dkk, 2013).

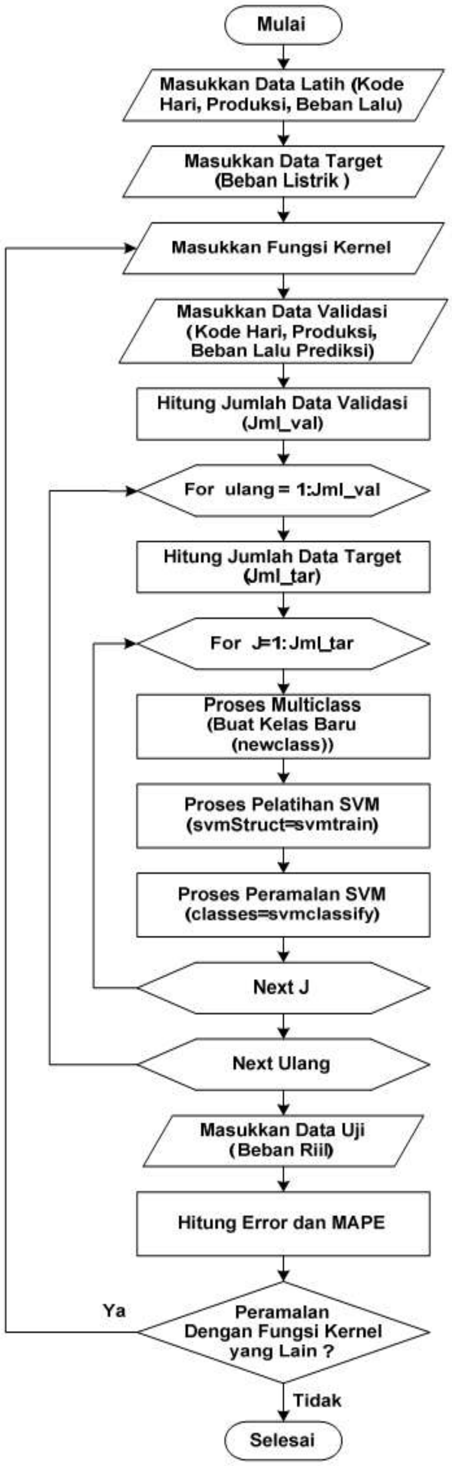
$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 -}} \quad (11)$$

Keterangan :

r = koefisien korelasi Pearson's
n = jumlah variabel
X = variabel bebas
Y = variabel terikat

Implementasi Program

Tahapan implementasi program atau *coding* merupakan langkah menterjemahkan logika diagram alir ke dalam program Matlab R2012. Diagram alir peramalaan beban listrik dengan *Support Vector Machine* diperlihatkan pada Gambar 1.



Gambar 1. Diagram Alir

Data Penelitian

Data penelitian analisis akurasi *support vector machine* dengan fungsi Kernel Gaussian RBF untuk prakiraan beban listrik pada sektor industri ini menggunakan data yang bersumber dari PT. Phapros Indonesia, yang beralamat di Jl. Simongan No. 131 Semarang. Perusahaan ini merupakan sebuah industri farmasi yang menghasilkan obat-obatan dalam bentuk padatan dan cair. Variabel data yang digunakan sebagai input vektor *x*, adalah data time series dari :

1. Beban listrik harian (MWh)
2. Kapasitas produksi industri (KG)
3. Hari kerja dan hari libur

HASIL DAN PEMBAHASAN

Jumlah waktu pelatihan SVM ini dibuat secara variasi, dari satu bulan sampai sebelas bulan. Sedangkan waktu prakiraan dibuat dengan data kelompok selama satu bulan waktu prakiraan. Rentang waktu antara data pelatihan dan data validasi mempunyai jarak satu minggu untuk hari yang sama.

Hasil prediksi beban listrik pada sektor industri menggunakan SVM berdasarkan fungsi Kernel Gaussian RBF diperoleh nilai MAPE terkecil 2,63% dengan menggunakan waktu pelatihan SVM selama 8 bulan.

Pengaruh Rentang Waktu Pelatihan Terhadap *Error*

Pengaruh rentang waktu pelatihan terhadap *error* ditentukan dengan menghitung nilai koefisien korelasi Pearson's (r), seperti pada Tabel 2

Tabel 2. Perhitungan Koefisien Korelasi Pearson's Pengaruh Rentang Waktu Pelatihan

No	X	Y	X^2	Y^2	XY
1	31	4,75	961	22,56	147,25
2	59	5,66	3.481	32,04	333,94
3	90	6,56	8.100	43,03	590,40
4	120	5,93	14.400	35,16	711,60
5	151	5,92	22.801	35,05	893,92
6	181	3,89	32.761	15,13	704,09
7	212	5,06	44.944	25,60	1.072,72
8	243	2,63	59.049	6,92	639,09
Jumlah	**1.087**	**40,40**	**186.497**	**215,50**	**5.093,01**

Uji korelasi dilakukan terhadap variabel X yang merupakan lama waktu pelatihan (jumlah hari) dan variabel Y yang merupakan nilai

MAPE (%) berdasarkan prediksi dengan metode SVM berdasarkan fungsi Kernel Gaussian RBF. Perhitungan nilai koefisien Pearson's (r) dari hasil prakiraan dengan fungsi Kernel Gaussian RBF diperlihatkan pada Tabel 2. Berdasarkan tabel 2, nilai koefisien Pearson di hitung sebagai berikut :

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(8)(5.093{,}01) - (1.087)(40{,}40)}{\sqrt{[(8)(186.497) - (1.087)^2][(8)(215{,}50) - (40{,}40)^2]}} = \frac{-3.170{,}72}{5.338{,}24}$$

$$r = -0{,}593$$

Nilai koefisien korelasi Pearson's r = -0,593 memberikan arti bahwa antara lama waktu pelatihan dengan nilai MAPE yang didapatkan dalam prakiraan SVM terdapat hubungan negatif dengan kekuatan hubungan yang sedang. Artinya, jika waktu pelatihan semakin lama, maka nilai MAPE akan semakin kecil.

Pengaruh Rentang Waktu Prakiraan Ke Depan Terhadap *Error*

Pengaruh rentang waktu prakiraan ke depan terhadap *error* ditentukan dengan menghitung nilai koefisien korelasi Pearson's. Uji korelasi dilakukan terhadap variabel X yaitu lama waktu prakiraan ke depan (jumlah hari) dan variabel Y yang merupakan nilai MAPE (%) seeprti yang ditunjukkan pada tabel tiga 3. Berdasarkan tabel 3, nilai koefisien Pearson di hitung sebagai berikut :

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(30)(1352{,}46) - (465)(79{,}90)}{\sqrt{[(30)(9455) - (465)^2][(30)(415{,}51) - (78{,}90)^2]}} = \frac{3884{,}61}{20511{,}80}$$

$$r = 0{,}1894$$

Tabel 3. Perhitungn Koefisien Korelasi Pearson's Pengaruh Rentang Waktu Prakiraan Ke Depan

No	X	Y	X^2	Y^2	XY
1	1	0,00	1	0,00	0,00
2	2	0,00	4	0,00	0,00
3	3	3,68	9	13,51	11,03
4	4	0,00	16	0,00	0,00
5	5	0,56	25	0,31	2,79
6	6	1,61	36	2,60	9,68
7	7	9,36	49	87,55	65,50
8	8	0,67	64	0,45	5,37
9	9	4,70	81	22,11	42,32
10	10	1,56	100	2,45	15,64
11	11	0,66	121	0,44	7,28
12	12	0,68	144	0,46	8,14
13	13	0,25	169	0,06	3,29
14	14	8,55	196	73,18	119,76
15	15	1,25	225	1,57	18,81
16	16	1,46	256	2,14	23,40
17	17	2,73	289	7,45	46,40
18	18	5,10	324	26,00	91,78
19	19	3,74	361	13,98	71,05
20	20	7,14	400	51,02	142,86
21	21	2,70	441	7,30	56,76
22	22	1,93	484	3,71	42,36
23	23	2,23	529	4,98	51,34
24	24	0,77	576	0,59	18,46
25	25	1,43	625	2,03	35,64
26	26	0,11	676	0,01	2,89
27	27	0,25	729	0,06	6,72
28	28	7,32	784	53,54	204,88
29	29	5,29	841	28,00	153,46
30	30	3,16	900	10,00	94,87
Jumlah	**465**	**78,90**	**9455**	**415,51**	**1352,46**

Tabel 3. merupakan Tabel Perhitungan nilai koefisien korelasi Pearson's yang dilakukan untuk nilai MAPE terkecil dari hasil penelitian, yaitu hasil prakiraan yang menggunakan fungsi Kernel Gausian RBF dengan waktu pelatihan 8 bulan. Nilai koefisien korelasi Pearson's r = 0,1894 memberikan arti bahwa antara lama waktu prakiraan ke depan dengan nilai *error* yang didapatkan dalam prakiraan SVM terdapat hubungan positif yang rendah atau lemah. Artinya, jika waktu prakiraan semakin lama, maka nilai *error* akan semakin besar.

KESIMPULAN

Penelitian ini menghasilkan kesimpulan sebagai berikut:

1. Pada prakiraan yang menggunakan metode SVM dengan fungsi Kernel Gaussian RBF, rentang waktu pelatihan mempengaruhi *error* hasil prakiraan, semakin lama waktu pelatihan SVM maka nilai *error* prakiraan semakin kecil.
2. Rentang waktu prakiraan ke depan pada prakiraan menggunakan metode SVM dengan fungsi Kernel Gaussian RBF

mempengaruhi *error* hasil prakiraan, semakin lama waktu prakiraan ke depan maka nilai *error* prakiraan semakin besar

DAFTAR PUSTAKA

Al-Hamadi, H.M., Soliman, S.A., (2005), Long-Term/Mid-Term Electric Load Forecasting Based on Short-Term Correlation and Annual Growth, Electric Power Systems Research 74 (2005) 353–361

Cristianini N., Taylor J.S., (2000), An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge Press University, 2000

Ceperic, V., Gielen, G., Baric, A., (2012), Expert Systems with Applications 39, 10933–10942

Hahn, H., Nieberg, S.M., Pickl, S., (2009) , Electric Load Forecasting Methods: Tools for Decision Making, European Journal of Operational Research 199, 902–907

Hsu, C.C., Chen, C.Y., (2003), Regional Load Forecasting in Taiwan-Applications of Artificial Neural Networks, Energy Conversion and Management 44 1941–1949

Maali, Y., Al-Jumaily, A., (2013), Self-Advising Support Vector Machine, Knowledge-Based Systems 52, 214–222.

Misbahuddin, Hasan, I., (2013), Analisis Data Penelitian dengan Statistik, Bumi Aksara, Jakarta

Qi, Z., Tian, Y., Shi, Y., (2013), Robust Twin Support Vector Machine for Pattern Classification, Pattern Recognition 46, 305–316

Smola, A.J., Scholkopf, B., (2004), A Tutorial On Support Vector Regression, Statistic and Computing 14, 199-222

Vapnik, V.N., (1999), The Nature of Statistical Learning Theory 2nd edition, Springer-Verlag, New York Berlin Heidelberg