

Pengaruh Preprocessing Data pada Metode SVR dalam Memprediksi Permintaan Obat

Bakhtiyar Hadi Prakoso¹⁾

¹⁾Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121
Email : ¹⁾bahtiyar.hp@unmuhjember.ac.id

ABSTRAK

Stock out merupakan permasalahan yang sering muncul pada Instalasi Farmasi Rumah Sakit. Kondisi ini disebabkan karena permintaan obat lebih banyak dari stok obat yang ada. Upaya untuk mengatasi *stock out* dengan mengelola persediaan obat dengan benar salah satunya adalah dengan cara memprediksi permintaan obat. Pada penelitian ini akan digunakan metode SVR. Dalam perhitungan SVR melewati proses sebuah *preprocessing data* yang berfungsi untuk meningkatkan akurasi hasil. Penelitian ini akan membandingkan metode *preprocessing linear scaling* dengan *z normalization*. Hasil *MAPE* menunjukkan *preprocessing* dengan *linear scaling* menghasilkan nilai yang lebih baik dibandingkan dengan *z-normalization*.

Kata kunci : Persediaan, *Stok out*, Prediksi

1. PENDAHULUAN

Pengelolaan obat pada Rumah Sakit merupakan hal yang sangat riskan. Masalah yang sering muncul yaitu kondisi *stock out*. Kondisi ini disebabkan karena permintaan obat lebih banyak dari stok obat yang ada. Jika kondisi ini tidak dapat teratasi dengan baik dapat berdampak pada pendapatan rumah sakit (Uthayakumar & Priyan, 2013). Menurut Kumar dalam satu tahun rumah sakit mengeluarkan sekitar 33% dari total anggaran pengeluaran seluruhnya (Kumar & Chakravarty, 2015).

Upaya yang dapat dilakukan dalam mengatur persediaan adalah dengan memprediksi permintaan obat kedepannya (Fruggiero et al, 2012). Terkait metode prediksi, beberapa dekade terakhir ini telah berkembang metode SVR dimana metode tersebut telah populer penggunaannya untuk menyelesaikan permasalahan regresi yang dapat digunakan untuk memprediksi kebutuhan masa depan (Lin & Weng,

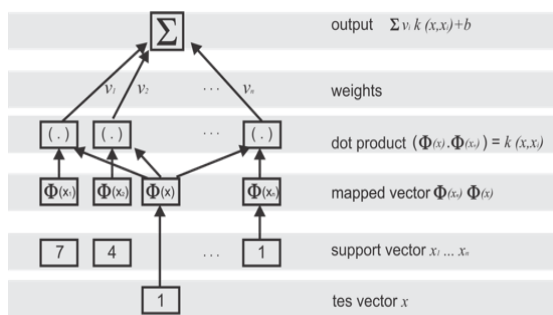
2004). Keuntungan dari metode SVR adalah tingkat akurasi yang lebih tinggi dibandingkan metode yang lainnya (Levis & Papageorgiou, 2005), dapat menganggulangi masalah *over fitting* yang tidak dapat diselesaikan oleh metode (*Artificial Neural Network*).

Perhitungan SVR akan melewati proses sebuah *preprocessing data*. Penggunaan *preprocessing data* ini bertujuan untuk meningkatkan tingkat akurasi pada perhitungan prediksi yang dilakukan. Terdapat beberapa metode yang digunakan untuk proses *preprocessing data*, yaitu *linear scaling* dan *z normaliation*. Pada penelitian ini akan dilakukan perbandingan proses *preprocessing data* antara *linear scaling* dan *z normalization* dengan menggunakan data permintaan obat pada rumah sakit. Dari uji coba tersebut diharapkan dapat menemukan metode *preprocessing data* yang tepat untuk membantu memprediksi kebutuhan obat.

2. TINJAUAN PUSTAKA

2.1 Support Vector Regression

Metode Support Vector Regression pertama kali dikenalkan oleh Vapnik (1999) ini merupakan hasil modifikasi metode Support Vector Machine (SVM) yang dibuat untuk menyelesaikan permasalahan regresi, sehingga output dari SVR adalah berupa bilangan riil dan kontinyu (Smola & Schölkopf, 2004). Cara kerja dari metode SVR dijelaskan pada Gambar 1 dibawah ini.



Gambar 1. Arsitektur Support Vector Regression (SVR)
(Sumber : Smola & Schölkopf, 2004)

Misalkan terdapat training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)\} \subset \mathcal{X} \times \mathbb{R}$ di mana \mathcal{X} merupakan input vector dari \mathbb{R}^d . Dalam konsep ϵ -SV Regression, Vapnik menjelaskan bahwa tujuan dari ϵ -SV Regression ini adalah mencari fungsi $f(x)$ yang mempunyai deviasi maksimal sebesar ϵ untuk mendapatkan nilai target y_i dari semua training data. Nilai kesalahan akan diterima apabila kurang dari nilai ϵ . Sebaliknya, nilai kesalahan tidak diterima apabila nilainya melebihi ϵ .

Fungsi linear secara umum dapat dituliskan sebagai fungsi f pada rumus 1.

$$f(x) = \langle w, x \rangle + b \quad w \in \mathcal{X}, b \in \mathbb{R} \quad (1)$$

Di mana $\langle \cdot, \cdot \rangle$ merupakan hasil dari dot product di \mathcal{X} . Fungsi 1 merupakan fungsi flatness sehingga tujuannya adalah

mencari nilai terkecil dari w . Cara untuk mencari nilai tersebut adalah meminimalkan bentuk Euclidean $\|w\|^2$. Permasalahan tersebut dapat diselesaikan dengan convex optimization yang dijelaskan pada rumus 2.

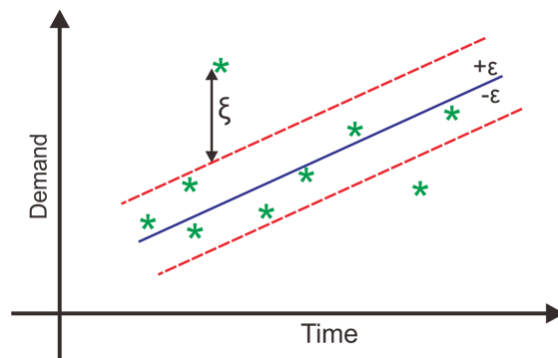
$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|w\|^2 \\ \text{Subject to :} \quad & y_i - \langle w, x \rangle - b \leq \epsilon \\ & \langle w, x \rangle + b - y_i \leq \epsilon \end{aligned} \quad (2)$$

Ada kondisi di mana nilai kesalahan diperbolehkan melebihi batas ambang nilai ϵ . Pada kondisi ini dibutuhkan soft margin atau bisa disebut dengan variable slack ξ_i, ξ_i^* . Sehingga persamaannya berubah seperti pada rumus 3.

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=0}^p (\xi_i + \xi_i^*) \\ \text{Subject to :} \quad & y_i - \langle w, x \rangle - b \leq \epsilon + \xi_i \\ & \langle w, x \rangle + b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \quad (3)$$

Konstanta $C > 0$ menentukan trade of antara flatness dari fungsi f dan seberapa besar tingkat deviasi kesalahan dari batas ϵ yang dapat ditolerasi. Formulasi tersebut terdapat pada rumus 4 yang biasa disebut dengan ϵ -insensitive loss function $|\xi_i|_\epsilon$.

$$|\xi_i|_\epsilon := \begin{cases} 0 & \text{if } |\xi_i| \leq \epsilon \\ |\xi_i| - \epsilon & \text{otherwise} \end{cases} \quad (4)$$



Gambar 2. Ilustrasi Konsep Support Vector Regression (SVR)
(Sumber: Levis & Papageorgiou, 2005)

Gambar 2 menjelaskan ilustrasi konsep SVR. Permasalahan SVR dapat dipandang sebagai *Convex Linear Programming NLP Optimization Problem* di mana tujuannya meminimalkan fungsi kuadrat untuk diubah menjadi sebuah batasan. Batasan-batasan tersebut merupakan *dual problem formulation* yang dapat dipecahkan dengan persamaan *Lagrange Multiplier* seperti pada rumus 5.

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} (\xi_i + \xi_i^*) - \sum_{i=1}^{\ell} \alpha_i \cdot (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^n \alpha_i^* \cdot (\varepsilon + \xi_i + y_i - \langle w, x_i \rangle - b) - \sum_{i=1}^{\ell} \mu_i \cdot \xi_i - \sum_{i=1}^{\ell} \mu_i^* \cdot \xi_i^* \quad (5)$$

Pada saat optimasi, *Karush-Kuhn-Tucker (KKT)* menjadikan turunan dari variabel L terhadap (w, b, ξ_i, ξ_i^*) sama dengan nol.

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\Rightarrow w - \sum_{i=1}^N ((\alpha_i - \alpha_i^*) \cdot x_i) = 0 \\ \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^N ((\alpha_i - \alpha_i^*) \cdot x_i) = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 &\Rightarrow C - \alpha_i - \mu_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} = 0 &\Rightarrow C - \alpha_i^* - \mu_i^* = 0 \end{aligned} \quad (6)$$

Dengan mensubstitusi dua persamaan di atas, maka didapatkan:

$$\max \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^{\ell} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell} y_i (\alpha_i - \alpha_i^*)$$

Dengan :

$$\begin{aligned} \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) &= 0 \\ 0 \leq \alpha_i &\leq C \quad \forall i = 1, \dots, N \\ 0 \leq \alpha_i^* &\leq C \quad \forall i = 1, \dots, N \end{aligned} \quad (7)$$

Dari proses penurunan rumus 6 dan 7 didapatkan dua variabel utama yaitu α_i dan α_i^* . Solusi permasalahan *dual problem optimization* diturunkan dari vektor w kemudian disubstitusikan ke dalam fungsi $f(x)$.

$$w = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot x_i \quad (8)$$

Jika rumus 1 dan 8 disubstitusikan maka didapatkan persamaan baru sehingga persamaannya seperti dijelaskan pada pada rumus 9.

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b \quad (9)$$

Di mana x_i merupakan *support vector* dan x merupakan *test vector*. Fungsi di atas adalah fungsi yang digunakan untuk menghitung permasalahan *linear*. Sedangkan untuk permasalahan *non linear* nilai x_i dan x terlebih dahulu ditransformasikan ke dalam *space feature* yang berdimensi tinggi dengan cara memetakan vektor x_i dan x ke dalam fungsi Φ , yang dijelaskan pada rumus 10.

$$\Phi : R^Z \rightarrow R^{Z'} \quad (10)$$

Hasil pemetaan tersebut akan menghasilkan rumus 11.

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot \Phi(x_i) \cdot \Phi(x) + b \quad (11)$$

Karena *input vector* x_t' dan x_t sudah ditransformasikan oleh fungsi Φ dan berada di dalam *feature space*, maka fungsi transformasi Φ dapat direpresentasikan dengan fungsi kernel K sebagai berikut:

$$K(x_i, x) = \Phi(x_i) \cdot \Phi(x) \quad (12)$$

Sehingga persamaan akhirnya menjadi :

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \quad (13)$$

Fungsi $K(x_t', x_t)$ merupakan *kernel trick*. Menunjukkan *kernel trick* yang sering dipakai dalam metode SVM/SVM.

2.2 Preprocessing Data

Menurut Crone tahapan *preprocessing data* yang biasa diimplementasikan kedalam metode *Support Vector Regression* terdapat dua macam yaitu sebagai berikut (Crone, et.al, 2006).

Linear Scaling

$$Z_i = lb + \frac{X_i - MIN(X)}{MAX(X) - MIN(X)} * (ub - lb) \quad (14)$$

Keterangan

- lb* : lower bound
- ub* : upper bound
- MAX(X)* : nilai maksimal dari dari sekumpulan time series X
- MIN(X)* : nilai minimal dari dari sekumpulan time series X
- X_i* : nilai pada time series X pada ke *i*

Normalization

$$Z_i = \frac{X_i - mean(X)}{SD(X)} \quad (15)$$

Keterangan

- mean(X)* : nilai mean dari dari sekumpulan time series X
- SD(X)* : nilai Standart Deviasi dari dari sekumpulan time series X
- X_i* : nilai pada time series X pada ke *i*

2.3 MAPE (Mean Absolute Presentage Error)

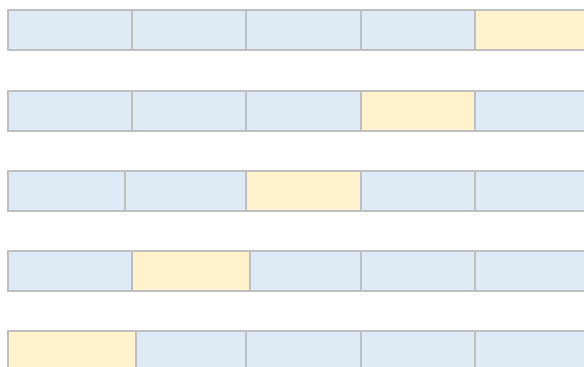
Tingkat akurasi dari prediksi dapat diketahui dengan menghitung *Mean Absolute Presentage Error (MAPE)*. Adapun rumus untuk menghitung *MAPE* dijelaskan pada rumus 16 (Hyndman & Koehler, 2006) :

$$MAPE = \frac{1}{n} \sum_{i=0}^n \frac{|x_i - \bar{x}_i|}{x_i} \times 100\% \quad (16)$$

Di mana $x = (x_1, x_2, x_3, \dots, x_n)$ merupakan data aktual dan $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_2, \dots, \bar{x}_n)$ merupakan data prediksi.

2.4 MAPE (Mean Absolute Presentage Error)

K-Fold Cross Validation membagi data menjadi dua bagian yaitu data *testing* dan data *training*. Metode ini membagi data sebanyak *k* blok data di mana masing-masing blok mempunyai himpunan anggota yang sama besar. Satu blok data akan digunakan untuk *data testing* sedangkan *k-1* blok lainnya dijadikan untuk *data training*. Nilai uji kesalahan dihitung dengan mencari rata-rata dari *k* macam uji tersebut (Bergmeir, Costantini, & Bennitez, 2014).

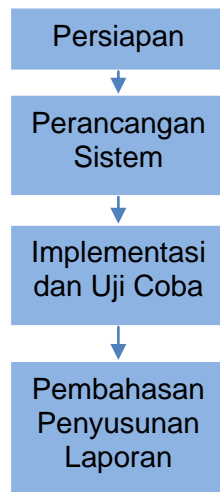


Gambar 3. Skema 5-fold cross validation

Gambar 3 merupakan skema *5-fold cross validation*. Blok warna kuning adalah bagian data yang digunakan untuk *testing*, sedangkan blok warna biru bagian data yang digunakan untuk proses *training*.

3. METODE PENELITIAN

Tahapan-tahapan penelitian yang dikerjakan dalam penelitian ini dijelaskan pada Gambar 4.



Gambar 4. Alur Penelitian

Persiapan

Tahap ini merupakan tahap awal untuk melakukan penelitian. Tahap-tahapnya adalah dengan studi literatur dan pengambilan data. Studi literatur yang dipelajari berkaitan dengan konsep dari Peramalan, Java, dan R Programing. Sementara untuk pengambilan data dilakukan pada rumah sakit swasta yang ada di Kabupaten Jember.

Perancangan Sistem

Perancangan sistem ini meliputi perancangan arsitektur dari aplikasi yang dibangun. Merancang tatap muka yang digunakan, serta merancang code program yang digunakan.

Implementasi dan Uji coba

Dari perancangan yang telah dilakukan maka kemudian dilakukan pembangunan terhadap sistem. Setelah sistem sudah selesai dibuat langkah selanjutnya adalah melakukan uji coba terhadap sistem yang dibuat. Uji coba ini digunakan untuk menilai apakah sistem yang dibuat masih menimbulkan kesalahan atau tidak.

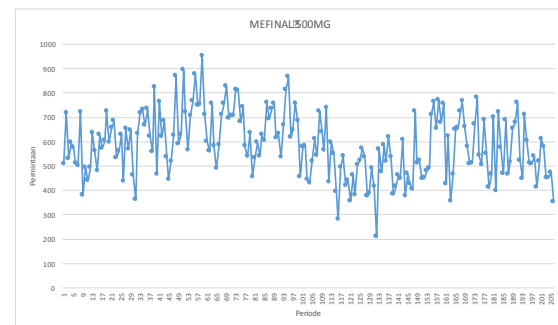
Pembahasan dan Penyusunan dokumen laporan penelitian

Pada tahap ini dilakukan pembahasan mengenai hasil pengamatan

selama uji coba, selain itu juga dilakukan pengamatan terhadap kekurangan yang ada serta kemungkinan pengembangan yang dapat dilakukan pada penelitian selanjutnya. Hasilnya pembahasana ini kemudian akan di susun dalam sebuah laporan penelitian.

4. HASIL DAN PEMBAHASAN

Data uji diambil di salah satu Rumah Sakit Swasta di Jember. Data obat yang digunakan adalah MEFINAL 500MG dalam bentuk rekap mingguan dengan total 200 periode data. Obat ini diambil karena mempunyai permintaan yang cukup banyak dibandingkan obat-obat yang lain. Gambar 5 dibawah ini merupakan representasi visual dari permintaan obat MEFINAL 500MG. Representasi gambar dijelaskan dalam bentuk grafik.



Gambar 5. Grafik Permintaan Obat Mefinal 500MG

Terlihat pada Gambar 5 Permintaan obat Mefinal 500MG menunjukkan bahwa pola data dibentuk cenderung fluktuatif pada nilai *mean*. Dari segi variansi, data cenderung lebih stabil.

Tabel 1. Deskripsi Statistic Mefinal 500MG

Min.	1st Qu.	Median	Mean	3rd Qu	Max.
208.0	490.2	579.0	584.7	684.0	951.0

Tabel 1 menunjukkan bahwa data *time series* Mefinal 500MG mempunyai nilai terendah 208.0 dan nilai tertinggi sebesar 951.0 Nilai Quarter pertama sebesar 490.2 dan Quarter ketiga sebesar 684.0. Nilai rata-rata atau mean sebesar 584,7 dan median sebesar 579.0.

Uji dilakukan dengan metode *K-Fold Cross Validation* dengan nilai $K=5$. Sehingga data dibagi menjadi 5 blok yang mempunyai panjang data yang sama. Data uji menggunakan data permintaan obat Mefinal 500MG yang terbagi menjadi dua jenis (berdasarkan perlakuan *preprocessing*) yaitu: data *linear scalling* dengan skala $[-1,+1]$ dan data *Z-Normalization*. Adapun hasil uji dari kedua jenis data adalah sebagai berikut:

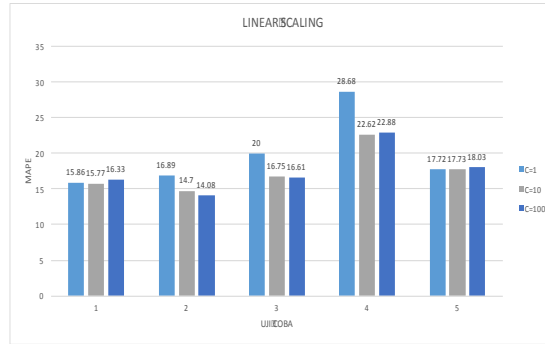
Data Linear Scalling

Skala yang digunakan pada proses *linear scalling* adalah skala $[-1,+1]$, skala ini umum digunakan untuk pengujian dengan metode SVR. Adapun hasil pengujian prediksi terhadap data *linear scalling* dijelaskan pada Tabel 2 dibawah.

Tabel 2. Hasil Pengujian (Data *linear scalling*)

UJI KE	MAPE (%)		
	C =1 dan $\epsilon = 0.1$	C =10 dan $\epsilon = 0.1$	C =100 dan $\epsilon = 0.1$
1	15,86	15,77	16,33
2	16,89	14,7	14,08
3	20,0	16,75	16,61
4	28,68	22,62	22,88
5	17,72	17,73	18,03
Rata – Rata	19,83	17,51	17,59

Dari Tabel 2 dapat disimpulkan bahwa nilai *MAPE* terendah adalah pada kondisi $C=10$ dan $\epsilon = 0.1$ yaitu sebesar 17,51%.



Gambar 6. Perbandingan Data Uji Mefinal 500MG (Data *Linear Scalling*)

Gambar 6 merupakan hasil visualisasi dari perbandingan data uji *linear scalling* dari berbagai kondisi parameter yang berbeda.

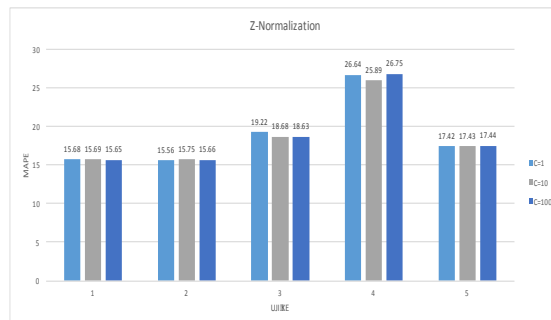
Data Z-Normalization

Data selanjutnya yang digunakan untuk uji adalah data *Z-Normalization*. Adapun hasil pengujian dengan *Z-Normalization* dapat di jelaskan pada tabel berikut ini.

Tabel 3. Hasil Pengujian (Data *linear scalling*)

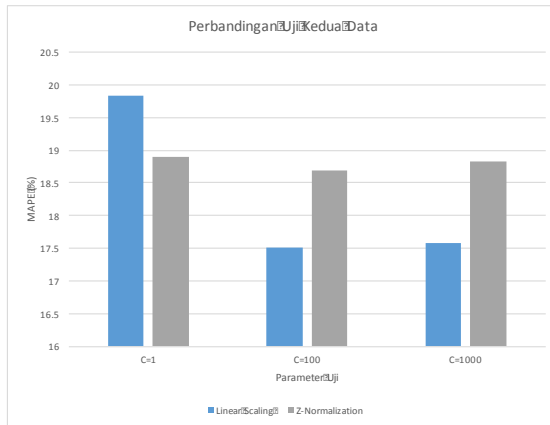
UJI KE	MAPE (%)		
	C =1 dan $\epsilon = 0.1$	C =10 dan $\epsilon = 0.1$	C =100 dan $\epsilon = 0.1$
1	15,68	15,69	15,65
2	15,56	15,75	15,66
3	19,22	18,68	18,63
4	26,64	25,89	26,75
5	17,42	17,43	17,44
Rata-Rata	18,09	18,69	18,83

Dari Tabel 3 dapat disimpulkan bahwa Nilai *MAPE* terendah adalah pada kondisi $C=10$ dan $\epsilon = 0.1$ yaitu sebesar 18,09%.



Gambar 7. Perbandingan Data Uji Mefinal 500MG (Data *Z-Normalization*)

Gambar 7 merupakan hasil visualisasi dari perbandingan data uji *Z-Normalization* dari berbagai kondisi parameter yang berbeda.



Gambar 8. Perbandingan Hasil Uji Ke Dua Data

Gambar 8 menunjukkan hasil perbandingan uji dari kedua data yang digunakan. Nilai diperoleh dari nilai *MAPE* tertinggi dari masing-masing parameter yang digunakan yaitu pada $C=1$, $C=10$, dan $C=1000$. Pada uji tersebut terlihat bahwa pada kondisi $C=1$ nilai data dengan *Z-Normalization* menghasilkan nilai yang lebih bagus dari pada data dengan transformasi *linear*. Pada kondisi $C=10$, dan $C=100$ nilai *MAPE* terbaik dihasilkan dari data dengan *Z-Normalization*.

5. KESIMPULAN DAN SARAN

Adapun kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut : (1) Hasil *MAPE* terendah untuk data Mefinal 500MG yang diproses melalui *linear scaling* sebesar 17,51% dengan parameter uji $C=10$ dan $\epsilon = 0.1$ dan (2) Hasil *MAPE* terendah untuk data Mefinal 500MG yang diproses melalui *Z-Normalization* sebesar 18,09% dengan parameter uji sebesar $C=1$ dan $\epsilon = 0.1$ Berdasarkan hasil pengujian dengan dua macam data yaitu *Linear Scaling* dan *Z-*

Normalization, didapatkan bahwa hasil *preprocessing* data yang menghasilkan *MAPE* yang paling baik adalah dengan menggunakan *Linear Scaling*. Sebagai masukan untuk penelitian selanjutnya untuk meningkatkan akurasi dimungkinkan penggunaan variasi kernel yang lain, serta penggunaan metode *hybrid SVR* juga dapat menjadi alternatif pada penelitian selanjutnya.

DAFTAR PUSTAKA

- Bergmeir, C., Costantini, M., & Bennitez, J. M. (2014). On the usefulness of cross-validation for directional forecast evaluation. *Computational Statistics and Data Analysis*, 76(2009), 132–143. doi:10.1016/j.csda.2014.02.001
- Crone, S. F., Guajardo, J., & Weber, R. (2006). The impact of preprocessing on support vector regression and neural networks in time series prediction. *Proceedings of the 2006 International Conference on Data Mining, DMIN*, 37–44. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.6266&rep=rep1&type=pdf>
- Fruggiero, F., Iannone, R., Martino, G., Miranda, S., & Riemma, S. (2012). A forecast model for pharmaceutical requirements based on an artificial neural network. *Service Operations and Logistics, and Informatics (SOLI), 2012 IEEE International Conference on*, 263–268.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. doi:10.1016/j.ijforecast.2006.03.001
- Kumar, S., & Chakravarty, A. (2015). ABC–VED analysis of expendable medical stores at a tertiary care hospital. *Medical Journal Armed*

- Forces India*, 71(1), 24–27.
doi:10.1016/j.mjafi.2014.07.002
- Levis, A., & Papageorgiou, L. (2005). Customer demand forecasting via support vector regression analysis. *Cemical Engineering Reaserch and Design*, (August). doi:10.1205/cherd.04246
- Lin, C., & Weng, R. C. (2004). Simple Probabilistic Predictions for Support Vector Regression. *Statistics*, 1–16. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8313&rep=rep1&type=pdf>
- Smola, A., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222.
- doi:10.1023/B:STCO.0000035301.49549.88
- Uthayakumar, R., & Priyan, S. (2013). Pharmaceutical supply chain and inventory management strategies for optimization: A study on pharmaceutical company and hospital. *Operations Research for Health Care*. doi:10.1016/j.orhc.2013.08.001
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, 10(5), 988–999. doi:10.1109/72.788640