

Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen (Performance Analysis of Support Vector Regression (SVR) Methods in Predicting the Consumer Price Index)

Rokhmad Eko Cahyono^{[1]*}, Judi Prajetno Sugiono^[2], Suhatati Tjandra^[2]

^[1] ^[2] ^[3] Program Studi S2 Teknologi Informasi
Sekolah Tinggi Teknik Surabaya – STTS, Surabaya, Indonesia

E-mail: ^[1] ekomjkt28@gmail.com

E-mail: ^[2] jpsugiono@stts.edu

E-mail: ^[3] tati@stts.edu

KEYWORDS:

Kernel Function, Consumer Price Index (CPI), Mean Absolute Percentage Error (MAPE), Support Vector Regression (SVR), Input Variables, Output Variables

ABSTRACT

The stability of commodity prices for food is very influential on the economy of a region because stable prices have a direct impact on the level of people's purchasing power. The need to maintain the stability of food commodity prices is the background of this research and this can be anticipated by forecasting the Consumer Price Index (CPI). The CPI is an index number that calculates the average change in prices of goods and services consumed by households and society. The purpose of this study is to predict the CPI of the Foodstuff Group using the Support Vector Regression (SVR) method by utilizing Linear, Polynomial, Gaussian Radial Basis Function (RBF) and SPLine Kernel Functions. Selection of this SVR method, because SVR is able to map input vectors into higher dimensions and can produce a function with a trend of bumpy data following the data path formed, resulting in more accurate predictive values. Research is limited to the city of Surabaya, the period of time the study begins January 1, 2016 until December 31, 2018. The data source used is the Surabaya Food Basic Price data as an input variable with 34 input attributes and the CPI data for Surabaya city Foodstuffs group as output variables. The results of this study are expected to be able to predict CPI with an error rate below 10 %, which is indicated by MAPE (Mean Absolute Percentage Error) < 10. The result of the trial obtained the best value of the MAPE= 0.1716, using the Kernel Gaussian RBF

KATA KUNCI:

Fungsi Kernel, Indeks Harga Konsumen (IHK), Mean Absolute Percentage Error (MAPE), Support Vector Regression (SVR), Variabel Input, Variabel Output

ABSTRAK

Kestabilan harga komoditas bahan makanan sangat berpengaruh terhadap perekonomian suatu daerah karena harga yang stabil berdampak langsung pada tingkat daya beli masyarakat. Perlunya untuk menjaga kestabilan harga komoditas bahan makanan merupakan latar belakang dari penelitian ini dan hal tersebut dapat diantisipasi dengan melakukan peramalan pada Indeks Harga Konsumen (IHK). Indeks Harga Konsumen adalah angka indeks yang menghitung rata-rata perubahan harga barang dan jasa yang dikonsumsi oleh rumah tangga dan masyarakat. Tujuan penelitian ini adalah untuk memprediksi IHK Kelompok Bahan Makanan menggunakan metode Support Vector Regression (SVR) dengan memanfaatkan fungsi Kernel Linier, Polynomial, Gaussian Radial Basis Function (RBF) dan SPLine. Pemilihan metode SVR ini, karena SVR mampu memetakan vektor input ke dalam dimensi yang lebih tinggi dan dapat menghasilkan suatu fungsi dengan trend data yang bergelombang mengikuti jalur data yang terbentuk, sehingga menghasilkan nilai prediksi lebih akurat. Penelitian dibatasi hanya pada kota Surabaya, periode waktu penelitian mulai 1 Januari 2016 sampai dengan 31 Desember 2018. Sumber data yang digunakan adalah data Harga Bahan Makanan Pokok kota Surabaya sebagai variabel input dengan 34 atribut input dan data IHK kota Surabaya kelompok Bahan Makanan sebagai variabel output. Hasil penelitian ini diharapkan mampu memprediksi IHK dengan tingkat kesalahan (error) dibawah 10 %, yang ditunjukkan dengan MAPE (Mean Absolute Percentage Error) < 10. Hasil uji coba diperoleh nilai terbaik yaitu MAPE=0.1716, dengan menggunakan Kernel Gaussian RBF

I. PENDAHULUAN

Kestabilan harga komoditas bahan makanan merupakan salah satu indikator kemajuan bidang perekonomian di suatu daerah. Menjaga tingkat kestabilan harga komoditas bahan makanan di masyarakat pada periode sekarang dan yang akan datang merupakan inti permasalahan yang akan diselesaikan dalam penelitian ini.

Kestabilan harga bahan makanan dapat dijaga dengan mengetahui lebih awal nilai dari Indeks Harga Konsumen (IHK), dengan cara melakukan prediksi IHK menggunakan salah satu metode prediksi dalam kecerdasan buatan yaitu metode Support Vector Regression (SVR). IHK adalah angka indeks yang menghitung rata-rata perubahan harga barang dan jasa yang dikonsumsi oleh rumah tangga dan masyarakat. IHK terbagi dalam 7 kelompok pengeluaran dan beberapa sub-kelompok pendukung. Kelompok pengeluaran terdiri dari kelompok *Bahan Makanan*, kelompok *Makanan Jadi*, *Minuman*, *Rokok dan Tembakau*, kelompok *Perumahan*, *Air*, *Listrik*, *Gas dan Bahan Bakar*, kelompok *Sandang*, kelompok *Kesehatan*, kelompok *Pendidikan*, *Rekreasi dan Olahraga* dan kelompok *Transportasi*, *Komunikasi dan Jasa Keuangan*. Kelompok *Bahan Makanan* terdiri dari beberapa sub-kelompok pendukung yaitu sub-kelompok *Padi-Padian*, *Umbi-Umbian*, *Daging*, *Ikan Segar*, *Ikan Diawetkan*, *Telur dan Susu*, *Sayur-Sayuran*, *Kacang-Kacangan*, *Buah-Buahan*, *Bumbu-Bumbuan*, *Lemak dan Minyak*, dan *Bahan Makanan lainnya* [1]. Di Jawa Timur penghitungan IHK dilakukan oleh Badan Pusat Statistik (BPS) dengan melakukan survey di 8 daerah yang telah dipilih yaitu Surabaya, Madiun, Kediri, Probolinggo, Malang, Jember, Banyuwangi dan Sumenep. Sumber data input BPS adalah harga barang yang dihitung berdasarkan harga transaksi di tingkat pedagang eceran, harga transaksi yang dikumpulkan dari beberapa pasar tradisional dan pasar modern yang mewakili harga di daerah tersebut [2]. IHK yang diperoleh akan disajikan secara online setiap bulan pada website BPS <https://www.bps.go.id> dan dapat diakses oleh masyarakat umum.

Penelitian ini dibatasi untuk prediksi IHK pada kelompok *Bahan Makanan*, yang merupakan tujuan penelitian ini yaitu memprediksi IHK kelompok *bahan makanan* menggunakan metode

SVR, sedangkan daerah penelitian dibatasi di kota Surabaya. Diharapkan dengan melakukan perhitungan dan mengetahui nilai IHK lebih awal, dapat merencanakan langkah-langkah solutif agar kestabilan harga komoditas bahan makanan tetap terjaga.

Dalam penelitian ini menggunakan sumber data yang berbeda dengan sumber data yang digunakan oleh BPS, yaitu jika BPS menggunakan sumber data dengan melakukan survey harga di tingkat pedagang eceran, pasar tradisional dan pasar modern, sedangkan penelitian ini menggunakan sumber data online dari Dinas Perindustrian dan Perdagangan provinsi Jawa Timur, yang disajikan dalam website www.siskaperbapo.com [3]. Pembagian kelompok data dalam website tersebut hampir sama dengan website BPS, yaitu dalam website ini data terbagi dalam kelompok *Komponen Utama* dan *Sub-Komponen*. *Komponen utama* menyajikan data harga komoditas bahan makanan antara lain: *Beras*, *Gula Pasir*, *Minyak Goreng*, *Daging* dan lainnya, sedangkan komoditas tersebut juga memiliki beberapa sub-komponen, misalnya komoditas *Beras* memiliki 3 sub-komponen yaitu *Beras Bengawan*, *Beras Mentik* dan *Beras IR64*, komoditas *Daging* memiliki sub-komponen yaitu *Daging Sapi Murni*, *Daging Ayam Broiler* dan sebagainya. Data tersebut bersifat *time series* (runtun waktu) yaitu data historis yang tersusun berdasarkan urutan waktu yang mempunyai keterkaitan antara data masa lampau dan data saat ini. Data tersebut digunakan sebagai variabel input, dengan periode pengambilan data mulai 1 Januari 2016 sampai 31 Desember 2018, yang terdiri dari 34 atribut, sedangkan variabel output adalah IHK kota Surabaya periode waktu yang sama.

Metode prediksi menggunakan SVR dengan aplikasi 4 jenis fungsi Kernel, yaitu *Kernel Linier*, *Polynomial*, *Gaussian Radial Basis Function (RBF)* dan *SPLine*. Metode ini dipilih karena SVR mampu menemukan fungsi $f(x)$ sebagai *hyperplane*, untuk semua data input yang mempunyai deviasi paling besar ε dari target aktual y_i untuk semua data training dan dapat membuat error setipis mungkin.

Hipotesis dalam penelitian ini adalah sistem mampu memprediksi IHK kelompok *Bahan Makanan* dengan tingkat *error* (kesalahan) prediksi

dibawah 10 % atau MAPE < 10. Selain itu model prediksi IHK ini diharapkan dapat bermanfaat dan dikembangkan pada penelitian penelitian berikutnya dengan menggunakan metode prediksi yang lain seperti perceptron dan Back Propagation Neural Network (BP NN), sehingga hasil penelitian juga bermanfaat bagi masyarakat.

Beberapa penelitian sebelumnya yang mendukung penelitian ini antara lain penelitian dengan judul A New Application of The Support Vector Regression On The Construction of Financial Conditions Index to CPI Prediction, oleh Ye Wang, Bo Wang dan Xinyang Zhang. Penelitian ini menjelaskan model regresi menggunakan metode SVR dalam membangun Financial Conditions Index (FCI) dan mencari keterkaitan antara Indeks Komposit yang merupakan indikator keuangan dan inflasi di masa depan. Ada 8 atribut sebagai Variabel Input, yang merupakan data time series, digunakan dalam membangun FCI, sedangkan variabel output adalah Consumer Price Index (CPI) atau Indeks Harga Konsumen (IHK) [4].

Penelitian dengan judul Inflation Forecasting Using Support Vector Regression yang dilakukan oleh Linyun Zhang dan Jinchang Li. Penelitian ini menjelaskan cara melakukan pengembangan model SVR untuk memprediksi Consumer Price Indeks (CPI) atau IHK dengan memanfaatkan fungsi Kernel dan menjelaskan kelebihan SVR yang mempunyai RMSE paling kecil dalam melakukan prediksi dibandingkan dengan menggunakan BP Neural Network dan Regresi Linier [5].

Penelitian dengan judul A Support Vector Machine for Regression and Application to Financial Forecasting yang dilakukan oleh Theodore B. Trafalis dan Huseyin Ince. Penelitian ini memberikan penjelasan pada kasus prediksi harga saham dengan membandingkan penggunaan metode SVR dengan Back Propagation dalam prediksi harga saham tersebut. Dalam kasus ini obyek yang diteliti adalah harga saham IBM, saham Yahoo dan saham America Online (AOL). Data training diambil dari data sampel periode 7 Mei 1998 sampai 28 September 1998. Data testing diambil periode 29 September 1998 sampai 19 Oktober 1998. Untuk optimasi dan uji coba dengan menggunakan dua algoritma yaitu Standar-QP dan Primal-Dual [6].

Penelitian dengan judul Prediksi Indeks Harga Konsumen (IHK) Kelompok Perumahan, Air, Listrik, Gas Dan Bahan Bakar Menggunakan Metode Support Vector Regression. Penelitian ini dilakukan oleh Krishnanti Dewi, Putra Pandu Adikara dan Sigit Adinugroho. Penelitian ini menggunakan data IHK periode Januari 2011 sampai Desember 2016 [7].

A. Teori Dasar Support Vector Regression (SVR)

Support Vector Regression (SVR) ditemukan oleh Vladimir N. Vapnik tahun 1999. SVR merupakan penerapan dari metode *Support Vector Machine* (SVM), dalam kasus regresi. Perbedaan SVM dan SVR adalah SVM digunakan untuk mencari *hyperplane* (fungsi pemisah) yang terbaik diantara 2 obyek yang tidak terbatas jumlahnya dengan cara memaksimalkan jarak (margin) antara dua obyek yang berbeda sedangkan SVR digunakan untuk menemukan suatu fungsi yang memiliki deviasi paling besar ε dari target aktual y_i . Tujuan SVR adalah untuk menemukan fungsi $f(x)$ sebagai *hyperplane* yang berupa fungsi regresi, pada semua input data memiliki deviasi paling besar ε dari target aktual y_i untuk semua data training dan membuat error setipis mungkin [8]. Keuntungan menggunakan SVR yaitu SVR sangat cocok untuk data set berdimensi tinggi dan sangat cocok digunakan untuk kasus non linier dengan menggunakan fungsi Kernel [9].

Penggunaan metode SVR mempunyai kelebihan dibandingkan metode *Linier Regression* (regresi linier) yaitu jika regresi linier menghasilkan fungsi linier yang berupa garis lurus, pada algoritma SVR menghasilkan trend data yang bergelombang mengikuti jalur data yang terbentuk, sehingga prediksi data yang dihasilkan lebih akurat. Sebagai contoh terdapat l set data training (x_i, y_i) , $i = 1, 2, \dots, l$ yang mana x_i merupakan vektor input $x = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}^n$ dan y_i merupakan output $y = \{y_1, y_2, \dots, y_l\} \subseteq \mathbb{R}$ dan l adalah banyaknya data training. Metode SVR digunakan untuk mencari fungsi $f(x)$ yang mempunyai deviasi maksimal sebesar ε untuk mendapatkan target y_i dari semua data training.

Fungsi regresi $f(x)$ dengan batas deviasi (ε) sama dengan 0, merupakan fungsi regresi yang sempurna, sedangkan model SVR pada fungsi

regresi linier, mengacu pada Persamaan (1) di bawah ini:

$$f(x) = w^T \phi(x) + b \tag{1}$$

- $\phi(x)$: adalah titik di dalam feature space F , hasil dari pemetaan x di dalam input space
- w : adalah vektor bobot, yang mempunyai dimensi l
- b : adalah bias
- x : adalah input atau vektor input
- $f(x)$: adalah fungsi regresi

Dengan asumsi $x_i \in \mathbb{R}^n$ dan $y_i \in \mathbb{R}$ dimana $i = 1, 2, 3, \dots, l$

Sedangkan Koefisien w dan b mempunyai fungsi untuk meminimalkan fungsi resiko, seperti pada Persamaan (2) di bawah ini :

$$R = \min \frac{1}{2} \|w\|^2 + C \frac{1}{l} \left(\sum_{i=1}^l L_{\epsilon}(y_i, f(x_i)) \right) \tag{2}$$

yang memiliki batasan :

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

Dimana :

$$L_{\epsilon}(y_i, f(x_i)) = \begin{cases} 0, & \text{untuk } |y_i - f(x_i)| < \epsilon \\ |y_i - f(x_i)| - \epsilon, & \text{untuk } |y_i - f(x_i)| \geq \epsilon \end{cases}$$

- L_{ϵ} : merupakan Loss function yang bertipe ϵ -insentive loss function
- R : Risk Function (Fungsi Resiko)
- $\|w\|$: merupakan normalisasi w
- ϵ : merupakan Epsilon yaitu deviasi atau derajat toleransi terhadap error
- C : Nilai penalti jika deviasi lebih besar dari batas error

$L_{\epsilon}(y_i, f(x_i))$ menyatakan fungsi penalti untuk ϵ -insentive loss function, yang dikenai penalti jika error $|y_i - f(x_i)|$ lebih dari atau sama dengan ϵ . Faktor $\|w\|^2$ dinamakan regularisasi, meminimalkan $\|w\|^2$ akan didapatkan fungsi setipis mungkin, sehingga bisa mengontrol *function capacity* (kapasitas fungsi). Faktor lain untuk menunjang ketelitian fungsi adalah memperhitungkan *empirical error* (kesalahan

empirik) dengan menggunakan ϵ -insentive loss function. Koefisien ϵ ($+\epsilon$ dan $-\epsilon$) merupakan derajat toleransi terhadap error. Menurut Vapnik (1995), dalam kasus yang *Feasible*, dimana pasangan x_i, y_i masih dalam nilai ϵ , dapat menggunakan ide ϵ -insentive loss function, dengan meminimalkan norm w agar mendapatkan generalisasi yang baik untuk fungsi regresi $f(x)$. Problem Optimasi untuk kondisi *Feasible* (pembatas yang layak) dapat diselesaikan menggunakan persamaan sebagai berikut :

$$F = \text{Min} \frac{1}{2} \|w\|^2 \tag{3}$$

yang memiliki pembatas sebagai berikut :

$$\begin{aligned} y_i - w^T \phi(x_i) - b &\leq \epsilon \\ w^T \phi(x_i) - y_i + b &\leq \epsilon \\ \text{dimana } i &= 1, 2, 3, \dots, l \end{aligned}$$

Dalam Persamaan (3) di atas, diasumsikan semua titik berada dalam range atau dalam margin dengan kondisi *Feasible* yaitu $f(x) \pm \epsilon$.

Pada kondisi *Infeasible* (ketidaklayakan pembatas), problem Optimasi adalah ditunjukkan dengan adanya beberapa titik yang keluar dari range atau margin. Penyelesaian problem kondisi *Infeasible* ini adalah dengan menambahkan variabel *slack* (ξ, ξ^*), seperti Persamaan (4) sebagai berikut :

$$IF = \text{Min} \frac{1}{2} \|w\|^2 + C \sum_1^l (\xi_i + \xi_i^*) \tag{4}$$

yang memiliki pembatas sebagai berikut :

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

Konstanta $C > 0$, berfungsi untuk menentukan *trade off* (tawar menawar) antara ketipisan fungsi f dan batas atas deviasi yang lebih kecil dari ϵ , yang masih bisa di toleransi. C adalah parameter untuk menentukan berapa besar deviasi dari parameter epsilon yang masih bisa ditoleransi, yaitu semakin besar nilai C maka semakin besar nilai toleransi yang masih diijinkan. Parameter epsilon (ϵ) berfungsi untuk mengontrol lebar zona regresi dalam pemrosesan data. Semakin besar nilai epsilon

maka estimasi regresi semakin datar mendekati regresi linier.

Koefisien epsilon (ϵ) dalam algoritma SVR adalah ekuivalen dengan tingkat akurasi data training. Nilai ϵ yang kecil berhubungan dengan nilai yang tinggi pada variabel *slack* dengan akurasi tinggi. Sebaliknya nilai ϵ yang besar atau tinggi berhubungan dengan nilai variabel *slack* yang kecil dan akurasi yang rendah. Penambahan variabel *slack* ini untuk mengatasi masalah pembatas margin yang tidak layak dalam problem optimasi.

$$0 \leq \alpha_i^* \leq C, \quad i = 1, 2, 3, \dots, l$$

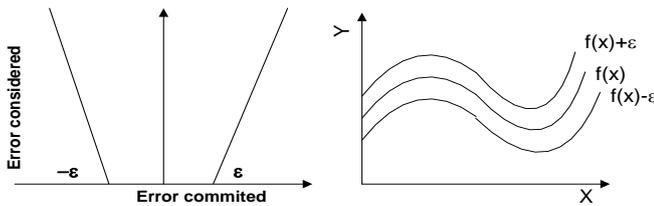
Pada koefisien C didefinisikan oleh user, $K(x_i, x_j)$ merupakan *dot product Kernel*, didefinisikan $K(x_i, x_j) = \phi^T(x_i) \phi(x_j)$

Dengan menggunakan Lagrange multiplier, maka kondisi optimalitas dari fungsi regresi sebagai berikut:

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i') K(x_i, x) + b \tag{6}$$

$K(x_i, x)$ adalah *Fungsi Kernel* antara data training x_i dan data testing x .

Performansi model SVR tergantung penggunaan hyperparameter yaitu C , epsilon (ϵ), degree, parameter kernel dan fungsi kernel yang digunakan [[9]]. Menurut Santosa, 2007, perbedaan pada *loss function* akan menghasilkan formulasi SVR yang berbeda dirumuskan dengan Formulasi ϵ -insensitive *loss function* sebagai berikut:



Gbr 1. Batas Error dalam SVR

Gambar 1, menjelaskan bagaimana error dalam SVR dihitung, sampai pada batas garis Error ϵ (epsilon), nilai error sama dengan 0, sedangkan di luar ϵ , nilai error dihitung sebagai *error-epsilon*. Sampai batas garis $(+\epsilon)$ dan $(-\epsilon)$, nilai error adalah 0 atau $\epsilon = 0$, sedangkan jika ada titik yang berada di luar margin atau diluar *tube* (tabung) harus ditambahkan variabel *slack* ξ , sehingga nilai error merupakan *error epsilon* dan dikenai penalti sebesar C . Sedangkan dari sebuah garis akan dibentuk sebuah tabung yang memiliki toleransi terhadap error. Fungsi regresi akan sempurna apabila batas deviasinya sama dengan 0. Pengertian *support vector* dalam SVR adalah data training yang terletak pada dan diluar batas ϵ . Jumlah *support vector* akan turun dengan naiknya ϵ .

$$L_{\epsilon}(y) = \begin{cases} |f(x) - y| - \epsilon, & \text{untuk yang lain} \\ 0, & \text{untuk } |f(x) - y| < \epsilon \end{cases} \tag{7}$$

Kernel Trick adalah metode yang dilakukan dengan cara perkalian antara input fitur data x yang kemudian diganti dengan fungsi kernel, sesuai persamaan di bawah ini :

$$(x_i, x_j) = \phi(x_i) \phi(x_j) = k(x_i, x_j) \tag{8}$$

Fungsi Kernel untuk mentransformasikan data pada *input space* yang non-linear ke dalam ruang fitur yang dimensinya tinggi (*feature space*), untuk menemukan *hyperplane* yang optimal. Permasalahan pada prediksi IHK ini merupakan permasalahan yang memiliki pola nonlinier sehingga untuk mengatasi masalah ketidaklinieran tersebut digunakan Fungsi Kernel.

Persamaan (5) di bawah ini merupakan formulasi dual untuk penyelesaian problem optimasi dari SVR, sebagai berikut :

$$\max -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*) \tag{5}$$

Subject to:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, 3, \dots, l$$

B. Metode Evaluasi Prediksi

Metode evaluasi untuk menentukan akurasi *performance* (kinerja) dengan menggunakan Mean Absolute Percentage Error (MAPE), yang merupakan prosentase pada kesalahan absolut rata rata yang dihitung dengan menemukan kesalahan absolute setiap periode, dan membagi dengan nilai

observasi pada periode tersebut, juga meratakan persentase absolute ini.

$$MAPE = \frac{\sum_{i=1}^n \frac{|x_i - F_i|}{x_i}}{n} \times 100\% \quad (9)$$

X_i = data nilai sebenarnya periode i
 F_i = data hasil ramalan periode i
 n = periode waktu ramalan

II. METODOLOGI

Metodologi penelitian ini menjelaskan tahapan langkah dalam prediksi IHK dan Perancangan Sistem untuk menghasilkan prediksi yang akurat, yaitu :

1. Melakukan identifikasi masalah

yaitu fokus pada penyelesaian permasalahan untuk menjaga kestabilan harga makanan yang merupakan kebutuhan dasar bagi masyarakat, dengan melakukan prediksi IHK kelompok bahan makanan, sehingga nilai IHK dapat diketahui lebih awal.

2. Studi Literatur

- a. Meliputi studi tentang konsep prediksi menggunakan SVR, analisa data time series, teori IHK dan pemanfaatan data online dari *website* (situs online)
- b. Studi Literasi untuk pendalaman materi yang berhubungan dengan penelitian

3. Pengumpulan Data

- a. Data yang digunakan merupakan data online harga konsumen bahan pokok kota Surabaya periode 1 Januari 2016 sampai dengan 31 Desember 2018 diambil dari website Dinas Perindustrian dan Perdagangan (www.siskaperbapo.com), yang merupakan variabel input dengan 34 atribut.
- b. Data online Indeks Harga Konsumen (IHK) periode tahun yang sama, diambil dari website Badan Pusat Statistik kota Surabaya (www.bps.go.id), sebagai variabel target

4. Analisa dan Perancangan Sistem

- a. Analisa dilakukan dengan membagi data menjadi dua bagian yaitu data training dan data testing

- b. Perancangan sistem dengan membuat *flowchart* (diagram alir) untuk menjelaskan urutan proses

5. Uji Coba Hasil Perancangan Sistem

- a. Melakukan uji coba menggunakan metode SVR, memanfaatkan 4 jenis fungsi Kernel dengan menggunakan software Matlab.

Fungsi Kernel tersebut sebagai berikut :

1. Kernel Linier

$$K(x_i, x) = x_i^T x \quad (10)$$

2. Kernel Polynomial

$$K(x_i, x) = (x_i^T x + 1)^d \quad (11)$$

3. Kernel Gaussian-RBF

$$K(x_i, x) = \exp(-\gamma \|x - x_i\|^2) \quad (12)$$

$$K(x_i, x) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (13)$$

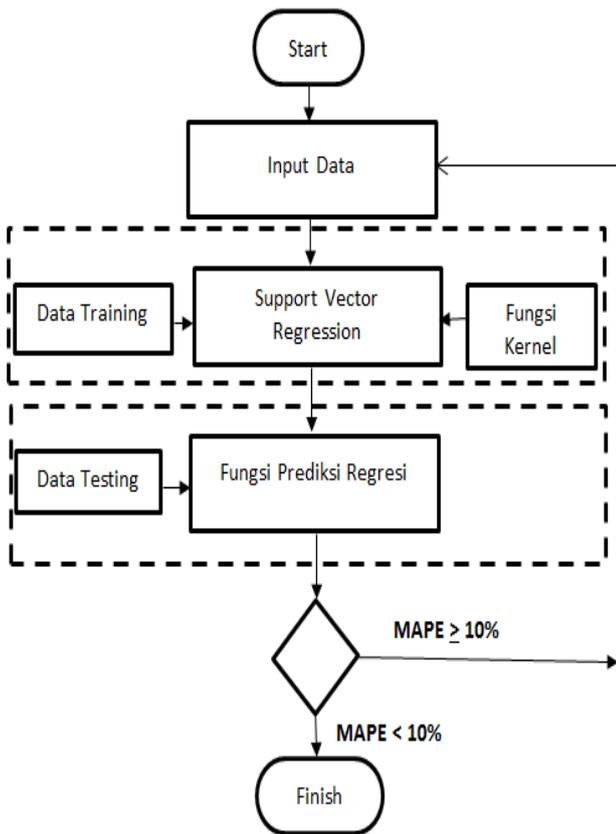
4. Kernel SPLines

$$K(x_i, x) = \prod_{m=1}^n K_m(x_m, X_{im}) \quad (14)$$

Proses pemilihan fungsi Kernel dengan menggunakan metode *Cross Validation* (CV).

- b. Pemerolehan nilai MAPE dan mengelompokkan dalam tabel dan grafik

Sedangkan perancangan sistem digambarkan dengan *Flow Chart* (Diagram Alir) seperti pada Gambar 2 dibawah ini :



Gbr 2. Diagram Alir Perancangan Sistem Prediksi IHK

Penjelasan Flow Chart proses Prediksi IHK pada Gambar 2 di atas, sebagai berikut :

- Input data : input data berasal dari data online Harga Konsumen Bahan Makanan kota Surabaya, sebagai variabel input, dengan 34 atribut, yang diambil dari website Dinas Perindustrian dan Perdagangan Provinsi Jawa Timur (www.siskaperbapo.com), mulai 1 Januari 2016 sampai 31 Desember 2018.
- Proses Training: tahapan proses training meliputi juga proses learning yaitu program diberikan sejumlah data input dan data output. Pada Fase training ini berusaha mencari hubungan antara variabel input yang menggunakan data Harga Makanan Kota Surabaya dengan variabel output berdasarkan data IHK. Tujuan tahapan ini adalah untuk menghasilkan fungsi prediksi regresi yang baru, dengan menggunakan metode SVR yang memanfaatkan 4 jenis Fungsi Kernel, yaitu Kernel Linier, Polynomial, Gaussian RBF dan SPLine, yang digunakan untuk proses *Non Linier Regression* (regresi non linier), dengan persamaan fungsi regresi sebagai berikut:

1. Sebagai contoh terdapat data set sebagai berikut :

$$D = \{(y_i, f(x_i))\}_{i=1}^N$$

x_i adalah vektor input, y_i hasil yang bersangkutan, N adalah ukuran data set

2. Persamaan regresi secara umum adalah :

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

Dimana $\varphi(x)$ adalah suatu titik di dalam feature space, w dan b adalah koefisien yang harus diprediksi dengan meminimalkan fungsi resiko yang didefinisikan pada point no.3

3. Melakukan proses Optimasi pada Fungsi regresi non-linier dengan fungsi resiko sebagai berikut :

Problem Optimasi

$$R = \min \frac{1}{2} \|w\|^2 + C \frac{1}{l} \left(\sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i)) \right)$$

4. Penambahan variabel slack pada regresi non linier, meminimalkan fungsi resiko

Meminimalkan Fungsi Resiko

$$R(w, \xi^*) = \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

5. Fungsi persamaan formulasi dual pada SVR sebagai berikut :

$$\max -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*)$$

Subject to:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, 3, \dots, l$$

$$0 \leq \alpha_i^* \leq C, \quad i = 1, 2, 3, \dots, l$$

α_i dan α_i^* adalah lagrange multiplier, dan $K(x_i, x_j)$ merupakan Fungsi Kernel yang dapat mengoptimalkan komputasi dalam input space.

Koefisien C didefinisikan oleh user, sedangkan $K(x_i, x_j)$ adalah dot product Kernel,

$$K(x_i, x_j) \leq \varphi^T(x_i), \varphi(x_j)$$

6. Setelah melalui beberapa tahapan proses di atas diperoleh Persamaan SVR yang baru sebagai berikut :

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i') K(x_i, x) + b$$

$K(x_i, x)$ adalah *Fungsi Kernel* antara data training x_i dan data testing x .

7. Proses pemrograman menggunakan software Matlab

- Proses Testing : merupakan tahapan program yang diberikan beberapa data baru sebagai data testing , untuk menguji tingkat akurasi dari metode yang telah didapatkan dan untuk menguji algoritma yang telah dihasilkan pada tahap training dan learning.
- Proses Evaluasi Kinerja : Performansi atau kinerja dalam metode SVR dilakukan evaluasi menggunakan parameter MAPE. Jika evaluasi kinerja mendapatkan $MAPE < 10$, artinya *error* (kesalahan) hasil prediksi IHK dibawah 10%, maka model prediksi IHK yang dihasilkan sudah sesuai dengan tujuan penelitian, dan apabila $MAPE \geq 10$ atau error di atas 10%, maka Diagram Alir Gambar 2 diatas, kembali mulai dari awal yaitu proses input data dan proses selanjutnya.

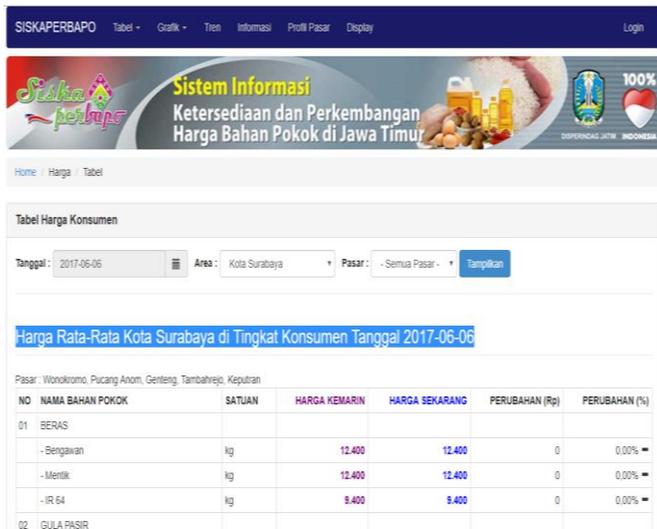
III. HASIL DAN PEMBAHASAN

Tahap ini adalah tahapan proses uji coba data training yang diolah dengan metode SVR untuk mendapatkan hasil prediksi IHK pada data testing dengan tingkat akurasi yang baik. Tahapan ini juga untuk menguji program SVR yang telah dibuat apakah sudah sesuai untuk aplikasi prediksi IHK. Jumlah data training Harga Bahan Makanan sebanyak 1224 data dan jumlah data testing IHK kelompok Bahan Makanan sebanyak 36 data. Pada Tabel 1, merupakan Contoh Harga Bahan Makanan kota Surabaya dari website Siskaperbapo, pada tanggal 6 Juni 2017. Variabel input dalam tabel tersebut terdiri dari 34 atribut, yaitu X1,X2,.....X34. Atribut X1 adalah Beras Bengawan sampai dengan X34 adalah Buncis, sedangkan kolom berikutnya menjelaskan Harga kemarin dan Harga sekarang. Pada Gambar 3,

adalah tampilan website www.siskaperbapo.com kota Surabaya pada 06 Juni 2017

TABEL I. CONTOH HARGA BAHAN MAKANAN POKOK KOTA SURABAYA PADA 6 JUNI 2017

Label	Variabel Input	Harga kemarin (Rupiah)	Harga sekarang (Rupiah)
X1	Beras Bengawan	12.400	12.400
X2	Beras Mentik	12.400	12.400
X3	Beras IR64	9.400	9.400
X4	Gula Pasir Dalam Negeri	10.995	11.037
X5	Minyak Goreng Bimoli Botol/ 2 ltr	12.400	12.400
X6	Minyak Goreng tanpa Merk / Minyak Curah	11.700	11.700
X7	Daging Sapi Murni	111.867	111.867
X8	Daging Ayam Broiler	29.800	29.800
X9	Daging Ayam Kampung	69.00	69.000
X10	Telur Ayam Ras/Petelur	19.600	20.000
X11	Telur Ayam Kampung	28.520	28.520
X12	Susu Kental Manis Merk Bendera	9.700	9.700
X13	Susu Kental Manis Merk Indomilk	9.900	9.900
X14	Susu Bubuk Merk Bendera (Instant)	38.660	38.660
X15	Susu Bubuk Merk Indomilk (Instant)	38.980	38.980
X16	Jagung Pipilan Kering	7.700	7.650
X17	Garam beryodium jenis Bata	620	620
X18	Garam beryodium jenis Halus	4000	4000
X19	Tepung Terigu Segitiga Biru (Kw Medium)	8.500	8.500
X20	Kacang Kedelai Eks Impor	9.300	9.300
X21	Mie Instant Indomie Rasa Kari Ayam	2.220	2.220
X22	Cabe Biasa	27.600	27.600
X23	Cabe Rawit	41.200	37.600
X24	Bawang Merah	26.400	28.600
X25	Bawang Putih	51.000	54.000
X26	Ikan Asin Teri	70.000	70.000
X27	Kacang Hijau	20.600	20.600
X28	Kacang Tanah	26.600	26.600
X29	Ketela Pohon	4.600	4.600
X30	Kol / Kubis	7.000	7.200
X31	Kentang	14.600	14.400
X32	Tomat	10.400	10.400
X33	Wortel	15.400	15.400
X34	Buncis	10.400	9.600



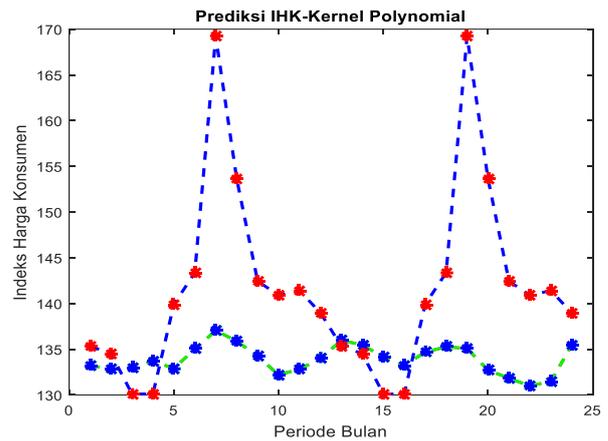
Gbr 3. Contoh Tampilan Website siskaperbapo pada 6 Juni 2017

A. Hasil Uji Coba

Uji coba dilakukan untuk menguji program SVR prediksi yang telah dibuat sudah sesuai, yang ditunjukkan dengan nilai MAPE yang kecil. Sebagai contoh dengan sample dataset nilai IHK selama 2 tahun, yaitu proses training dengan menggunakan data set mulai 1 Januari 2016 sampai dengan 31 Desember 2017 dan data testing menggunakan data 1 Januari 2017 sampai dengan 31 Desember 2018.

B. Uji Coba dengan Kernel Polynomial

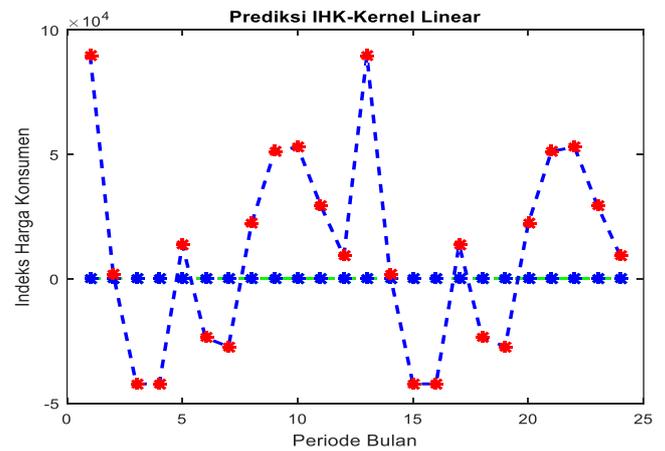
Dari grafik pada Gambar 4 terlihat bahwa nilai prediksi IHK (*titik merah*) dibandingkan dengan nilai aktual IHK (*titik biru*), mempunyai pola trending *hunting* (tidak stabil). Uji coba ini, diperoleh nilai MAPE = 22.2239, yang artinya tingkat *error* (kesalahan Prediksi) adalah 22,2239%. Dengan demikian model Kernel Polynomial ini kurang bagus jika digunakan memprediksi IHK, karena memiliki *error* > 10%



Gbr 4. Grafik Prediksi IHK periode 2016-2017 Menggunakan Kernel Polynomial

B. Uji Coba dengan Kernel Linier

Dari grafik pada Gambar 5 terlihat bahwa nilai prediksi IHK (*titik merah*) dibandingkan dengan nilai aktual IHK (*titik biru*), mempunyai pola trending yang tidak bagus, pola trending *hunting* (naik turun). Uji coba ini diperoleh nilai MAPE = 317.54, yang artinya tingkat kesalahan Prediksi IHK ini mempunyai *error* yang sangat besar, di atas 100%, sehingga model prediksi dengan menggunakan Kernel Linier ini tidak dapat digunakan untuk prediksi IHK

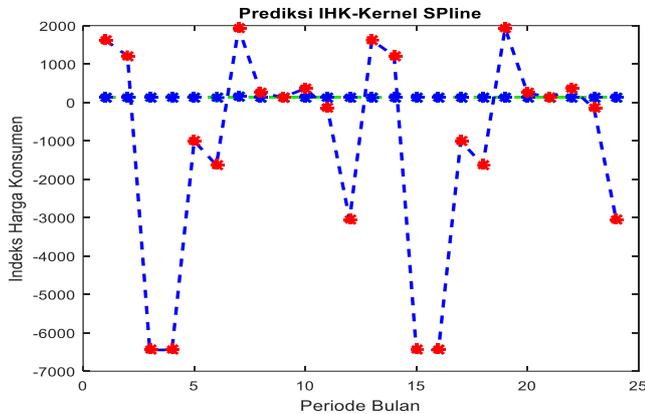


Gbr 5. Grafik Prediksi IHK periode 2016-2017 Menggunakan Kernel Linier

C. Uji Coba dengan Kernel SPLine

Dari grafik pada Gambar 6 terlihat bahwa nilai prediksi IHK (*titik merah*) dibandingkan dengan nilai aktual IHK (*titik biru*) mempunyai

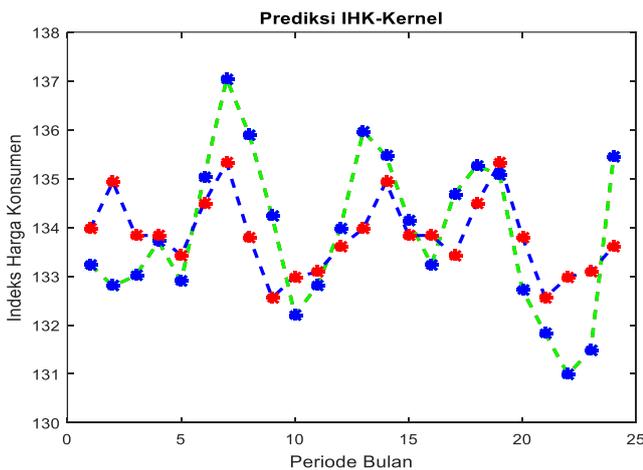
pola trending yang selalu *hunting* (tidak stabil), sehingga hasil uji coba dengan Kernel SPLine ini tidak dapat digunakan untuk Prediksi, karena mempunyai nilai MAPE= 3529.7



Gbr 6. Grafik Prediksi IHK periode 2016-2017 Menggunakan Kernel SPLine

D. Uji Coba dengan Kernel Gaussian-RBF

Dari grafik pada Gambar 7 terlihat bahwa nilai prediksi IHK (*titik merah*) dibandingkan dengan nilai aktual IHK (*titik biru*) mempunyai pola trending yang sangat bagus, yaitu nilai prediksi mengikuti pola trending dari nilai aktual. Dengan demikian penggunaan Kernel Gaussian RBF ini sangat bagus, dengan nilai MAPE= 0.1716 atau tingkat *error* (kesalahan) prediksi 0.1716 %



Gambar. 7 Grafik Prediksi IHK periode 2016-2017 Menggunakan Kernel Gaussian RBF

IV. PENUTUP

Dari hasil uji coba prediksi IHK dengan menggunakan 4 jenis fungsi Kernel yaitu Kernel Polynomial, Kernel Linier, Kernel SPLine dan Kernel Gaussian RBF, didapatkan bahwa nilai MAPE hasil prediksi yang paling rendah adalah dengan menggunakan Kernel Gaussian RBF, yaitu MAPE= 0.1716 atau tingkat *error* (kesalahan) prediksi hanya 0.1716 persen dan sesuai dengan tujuan penelitian yaitu MAPE < 10. Tabel 2 di bawah ini adalah rangkuman Evaluasi *Performance* (kinerja) metode SVR dalam memprediksi IHK kelompok Bahan Makanan kota Surabaya, yang ditunjukkan dengan nilai MAPE dari hasil uji coba.

TABEL III. EVALUASI KINERJA METODE SVR

Fungsi Kernel	MAPE
Gaussian RBF	0.1716
Polynomial	22.2239
Spline	3529.7
Linear	1.7185

V. UCAPAN TERIMA KASIH

Kami mengucapkan banyak terima kasih kepada Tim editorial Jurnal Teknologi Informasi dan Multimedia (JTIM) dan Tim Reviewer sehingga jurnal ini bisa terbit.

REFERENSI

- [1] <http://www.bi.go.id/>, “Indeks Harga Konsumen (IHK),” *META DATA*.
- [2] <https://bps.go.id/>, [2014-2018] *bps*. [Online], no. Indeks Harga Konsumen. .
- [3] <http://siskaperbapo.com/>, “[2014-2018] siskaperbapo. [Online].”
- [4] Y. Wang, B. Wang, and X. Zhang, “A new application of the support vector regression on the construction of financial conditions index to CPI prediction,” *Procedia - Procedia Comput. Sci.*, vol. 9, pp. 1263–1272, 2012.
- [5] Linyun Zhang and Jinchang Li, “Inflation Forecasting Using Support Vector Regression,” *Fourth Int. Symp. Inf. Sci. Eng.*, 2012.
- [6] Theodore B. Trafalis and Huseyin Ince, “A support Vector Machine for Regression and Applications to Financial Forecasting.”
- [7] K. Dewi, P. P. Adikara, and S. Adinugroho, “Prediksi Indeks Harga Konsumen (IHK) Kelompok Perumahan , Air , Listrik , Gas Dan Bahan Bakar Menggunakan Metode Support Vector Regression,” vol. 2, no. 10, pp. 3856–3862, 2018.

- [8] A. J. Smola and B. SCHOLKOPF, "A tutorial on support vector regression * ;," pp. 199–222, 2004.
- [9] R. E. Caraka, "Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis," vol. 7, no. 1, pp. 43–57, 2017.