

# **PREDIKSI HARGA KEDELAI LOKAL DAN KEDELAI IMPOR DENGAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *FORWARD SELECTION***

**Fatkuroji<sup>1</sup>, Stefanus Santosa<sup>2</sup>, Ricardus Anggi Pramunendar<sup>3</sup>**

<sup>1,3</sup>Program Pascasarjana Teknik Informatika, Universitas Dian Nuswantoro

<sup>2</sup>Politeknik Negeri Semarang

## ***Abstrak***

*Besarnya permintaan harga kedelai yang tinggi untuk kebutuhan makanan baik untuk olahan atau bahan jadi menjadikan harga kedelai sangat fluktuatif seiring impor kedelai yang terus meningkat. Pola harga kedelai yang sangat fluktuatif memicu gejolak ekonomi yang memicu terjadinya inflasi di salah satu daerah. Untuk mengatasi hal tersebut perlu adanya suatu prediksi harga kedelai agar pemerintah dapat mengantisipasinya. belum ada model prediksi terhadap harga komoditas kedelai baik lokal maupun impor, model prediksi yang ada saat ini tentang komoditi sembako. Penelitian ini mengusulkan model prediksi harga kedelai dengan menerapkan algoritma Support Vector Machine (SVM) dengan optimasi menggunakan Forward Selection. Untuk prediksi kedelai lokal dengan menggunakan parameter inputan data 4 (empat) hari sebelumnya, K-fold=10, nilai C= 0,1 diperoleh nilai RMSE terkecil sebesar 154.025 +/- 114.993. Setelah dilakukan seleksi atribut menggunakan Forward Selection diperoleh nilai RMSE sebesar 79.749 +/- 16.051, terdapat peningkatan RMSE sebesar= 74.276. Untuk prediksi kedelai lokal dengan menggunakan parameter inputan data 5 (lima) hari sebelumnya, K-fold=15, nilai C= 0,1 diperoleh nilai RMSE terkecil sebesar 126.008 +/- 78.371, setelah dilakukan optimasi menggunakan Forward Selection diperoleh nilai RMSE sebesar 122.270 +/- 56.049, terdapat peningkatan RMSE sebesar= 3.738.*

*Kata kunci: Prediksi Harga Kedelai, Support Vector, Forward Selection*

## **1. PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

Kedelai merupakan komoditas strategis yang unik tapi kontradiktif dalam sistem usaha tani di Indonesia. Luas pertanaman kedelai kurang dari lima persen dari seluruh luas areal tanaman pangan, namun komoditas ini memegang posisi sentral dalam seluruh kebijaksanaan pangan nasional karena peranannya sangat penting dalam menu pangan penduduk. Kedelai telah dikenal sejak awal sebagai sumber protein nabati bagi penduduk Indonesia namun komoditas ini tidak pernah menjadi tanaman pangan utama seperti halnya padi [1].

Seperti halnya padi kebutuhan kedelai terus meningkat karena penduduk, juga meningkatnya konsumsi per kapita terutama dalam bentuk olahan dan tumbuhnya industri pakan ternak. Permintaan kedelai per kapita sejak periode 1970 sampai 1990 telah meningkat 160%. Sedangkan pada periode 1990-an sampai tahun 2010 diperkirakan tumbuh 2,92% per tahun. Peningkatan konsumsi kedelai yang begitu pesat dan tidak dapat diimbangi oleh peningkatan produksi kedelai dalam negeri, maka terjadi kesenjangan. Kesenjangan itu ditutup dengan kedelai impor yang banyak menyita. Sejak perdagangan kedelai lepas dari kontrol BULOG mulai tahun 1991 impor kedelai meningkat sangat pesat [2]

Pola pemasaran kedelai biasanya petani menjual kepada pedagang yang dapat memberi harga yang baik. Kebanyakan perdagangan dilakukan di pasar dan di desa. Petani menjual ke toko di ibu kota kabupaten jika kedelainya banyak; bila tidak, mereka menjual kepada tengkulak di desanya. Pembeli tak pernah membayar dengan cicilan; mereka selalu membayar tunai segera setelah kedelai diterima. Dari

besarnya permintaan harga kedelai yang tinggi untuk kebutuhan makanan baik untuk olahan atau bahan jadi menjadikan harga kedelai sangat fluktuatif seiring impor kedelai yang terus meningkat.

Pola harga kedelai yang sangat fluktuatif memicu gejolak ekonomi yang memicu terjadinya inflasi di salah satu daerah. Untuk mengatasi hal tersebut perlu adanya suatu prediksi harga kedelai agar pemerintah dapat mengantisipasinya.

## 1.2. Identifikasi Masalah Spesifik

Dari penelitian prediksi harga komoditas yang sudah ada sebelumnya, ada beberapa penelitian. Penelitian yang dilakukan oleh Varra Wuwung, Nelson Nainggola dan Marline Paendongan yang melakukan penelitian tentang prediksi harga beras membramo dan beras sultan di kota Manado pada tahun 2013. Penelitian ini menggunakan metode ARIMA. Metode ini merupakan gabungan dari model *Autoregressive* (AR) dan model *Moving Average* (MA). Pada penelitian ini data set yang digunakan adalah rata-rata bulanan, model ARIMA sendiri mempunyai kelemahan dalam mempresentasikan informasi waktu dan model ini kurang tepat digunakan untuk data yang sulit untuk distasionerkan. Model ARIMA baik untuk peramalan jangka pendek tapi kurang baik untuk peramalan jangka panjang. [3].

Selain penelitian dengan model ARIMA juga ada penelitian yang dilakukan yang dilakukan oleh Sarim yang prediksi harga komoditas bawang merah dan cabe merah di daerah Brebes. Metode yang digunakan adalah *Support Vector Machine*, parameter yang digunakan adalah pola *inputan*, *K-fold*, *kernel* tipe dalam algoritma SVM. Hasil yang didapatkan dalam percobaan ini menunjukkan hasil yang sangat bagus dengan nilai RMSE yang sangat kecil yaitu untuk bawang merah sebesar=2857.885 +/- 3619.079 dengan harga prediksi Rp= 24.714 sedangkan nilai real pada prediksi sebesar= Rp.24.476. terdapat selisih harga =Rp 238,-. Dengan nilai akurasi 0.96%. Sedangkan untuk harga cabe nilai RMSE=1140.213 +/- 1337.272. Dengan nilai prediksi Rp= 10.895, sedangkan nilai real pada harga prediksi sebesar = Rp 10.850, terdapat selisih harga=Rp 45, terdapat nilai selisih 0,42% lebih besar dari kenyataan [4].

## 1.3. Analisis Masalah

Dari penelitian yang dilakukan pada penelitian di atas pola prediksi yang digunakan hanya membandingkan antar tipe kernel saja yang ada pada model SVM, tidak membandingkan dengan model yang lain. Sedangkan penelitian yang dilakukan oleh Muis Nanja dan Purwanto dalam penelitian tersebut menunjukkan optimasi forward seleksi menunjukkan hasil akurasi yang tinggi untuk memprediksi harga lada penelitian ini menggunakan K-NN yang berbasis *Forward Selection*, penelitian ini berhasil mencari tingkat RMSE terendah dibandingkan dengan penelitian yang sebelumnya dengan menggunakan SVM, pada penelitian ini juga tingkat akurasinya lebih bagus dibandingkan SVM [6]. Pada penelitian yang lain, yang dilakukan oleh Imelda A.Muis, Muhammad Affandes, M.T, penelitian ini berhasil dengan mendapatkan nilai akurasi yang tinggi dalam melakukan klasifikasi, sehingga dapat diterapkan agar memberikan bantuan kepada pengguna dalam mengelola *tweet*, terutama *tweet* iklan. Keberhasilan dengan mendapatkan nilai akurasi tertinggi 97,54% untuk data yang belum dilakukan pemilihan *feature*, sedangkan untuk data yang sudah dilakukan pemilihan terhadap *feature* mencapai nilai akurasi tertinggi 99.12% pada penelitian ini terdapat pada titik  $c=2.5, \gamma=0.3$  dan  $c=3, \gamma=0.3$  pada data yang belum dilakukan pemilihan *feature* sedangkan nilai akurasi tertinggi pada data yang sudah dilakukan pemilihan *feature* terdapat pada titik  $c=20, \gamma=6$ . Pengujian pada penelitian ini terjadi kesetabilan nilai akurasi pada rentang nilai  $1= c = 3$  dan  $0.01 = \gamma = 10$  pada data yang belum dilakukan pemilihan *feature* sedangkan untuk data yang sudah dilakukan pemilihan *feature* terjadi kesetabilan nilai akurasi pada rentang nilai  $1= c = 300$  dan  $0.01 = \gamma = 10$ [5]

Dan juga penelitian yang dilakukan dilakukan oleh Neneng Suningsih tentang seleksi variabel dalam analisis regresi *multivariate* multiple, menerangkan bahwa proses perhitungan dan algoritma *Forward Selection* dapat dijadikan rujukan karena lebih cepat analisis regresi *multivariate* multiple.

Oleh sebab itu penelitian ini mencoba mengusulkan suatu model optimasi SVM dengan *Forward Selection* untuk memprediksi harga kedelai.

#### 1.4. Rumusan Masalah

Berdasarkan masalah yang telah dijelaskan, maka dapat diangkat suatu permasalahan yang bisa dijadikan bahan penelitian yaitu: belum ada model prediksi terhadap harga komoditas kedelai baik lokal maupun impor, model prediksi yang ada saat ini tentang komoditi sembako yaitu prediksi harga bawang merah dan cabe dengan nilai akurasi mencapai 91%.

#### 1.5. Tujuan Penelitian

Tujuan penelitian ini adalah membangun model prediksi harga kedelai dengan menerapkan algoritma *Support Vector Machine* (SVM) dengan optimasi menggunakan *Forward Selection* dengan tingkat akurasi yang lebih tinggi atau RMSE yang lebih rendah.

#### 1.6. Kontribusi Penelitian

- a. Bagi masyarakat  
Penelitian ini diharapkan memberikan informasi bagi upaya pembangunan sistem cerdas untuk prediksi harga kedelai lokal dan kedelai impor menggunakan SVM bagi yang membutuhkan baik masyarakat maupun pemerintah.
- b. Bagi Pengembangan Keilmuan  
Hasil penelitian ini merupakan sumbangan berupa model prediksi harga kedelai menggunakan *Support Vector Machine* dengan seleksi atribut *Forward Selection*.

## 2. TINJAUAN PUSTAKA

### 2.1. Konstelasi Penelitian Prediksi Harga Komoditas Pangan

Banyak penelitian dilakukan sebelumnya yang membahas tentang prediksi harga komoditas pangan namun ada kesulitan dalam menemukan prediksi harga komoditas kedelai. Varra Wuwung, Nelson Nainggolan, Marline Paendonga, melakukan penelitian yang berjudul 'Prediksi Harga beras Sultan dan Membramo di Manado dengan Menggunakan Metode ARIMA'. Penelitiannya menggabungkan metode *Autoregressive* (AR) dengan model *Moving Average* (MA). Model AR sendiri adalah model yang menyatakan bahwa nilai pengamatan pada waktu sebelumnya ditambah galat, sedangkan model MA nilainya tergantung pada nilai sekarang dan waktu sebelumnya. Data yang digunakan dalam penelitian ini menggunakan data sekunder yang berupa data kuantitatif harga bulanan beras sebanyak 63 data. Penelitian ini dilakukan pada tahun 2013[3].

Sarim melakukan penelitian tentang prediksi harga komoditas bawang merah dan cabe merah di daerah Brebes. Dalam penelitiannya metode yang digunakan adalah *Support Vector Machine*, parameter yang digunakan adalah pola *inputan*, *K-fold*, tipe kernel dalam algoritma SVM. Hasil yang didapatkan dalam percobaan ini menunjukkan hasil yang sangat bagus dengan nilai RMSE yang sangat kecil yaitu untuk bawang merah sebesar=2857.885 +/- 3619.079 dengan harga hari prediksi Rp= 24.714 sedangkan nilai real pada harga hari prediksi sebesar= Rp.24.476. terdapat selisih harga =Rp 238,-. Dengan nilai akurasi 0.96%. Sedangkan untuk harga cabe nilai RMSE=1140.213 +/- 1337.272. Dengan nilai prediksi Rp= 10.895, sedangkan nilai real harga pada hari prediksi sebesar = Rp 10.850, terdapat selisih harga=Rp 45, terdapat nilai selisih 0,42% lebih besar dari kenyataan [4].

Dewi Sinta, menggunakan metode *Ensemble K-Nearest Neighbor* untuk prediksi harga beras di Indonesia. Dalam penelitiannya digunakan metode regresi K-NN untuk memprediksi respon atau peubah *output*. Bedanya dengan penelitian ini adalah digunakannya modifikasi prediksi untuk data deret waktu dengan konsep pembobot. Penelitian ini menerapkan metode K-NN tunggal dan ensemble K-NN pada data harga beras di Indonesia agar diperoleh keakuratan prediksi. Penelitian ini bertujuan untuk mengetahui kinerja metode K-NN tunggal dan ensemble K-NN, kemudian memprediksi harga beras di Indonesia menggunakan metode terbaik. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Kementerian Pertanian (KEMENTAN) bagian distribusi dan cadangan pangan. Data dibagi menjadi dua kelompok, data dari Januari 2010 hingga Desember 2011 (t=1 hingga t=24) dijadikan

sebagai *data training* dan sisanya dari Januari - Desember 2012 ( $t=25$  hingga  $t=36$ ) sebagai data testing. Peubah yang digunakan terdiri dari peubah *output* ( ) yaitu data harga beras per bulan dari Januari 2010 hingga Desember 2012 sedangkan untuk peubah *input* yang digunakan adalah X1 luas panen padi (ha), X2 produktivitas (ku/ha), X3 total produksi padi (ton) dan X4 jumlah penduduk (ribu)[5].

Muhammad Asri melakukan penelitian dengan judul ‘Modifikasi Model Faktor Koreksi pada Metode Ensemble K-NN untuk Pola Data Trend yang Mengandung Musiman’, Program Studi Statistika FMIPA Universitas Hasanuddin. Studi kasus yang digunakan adalah pada gabah kering, pada penelitian ini akurasi hasil prediksi untuk ketiga pendekatan metode *ensemble* KNN menggunakan nilai MAPE, MAE dan RMSE. Aplikasi metode *ensemble* KNN pada data harga gabah kering giling menunjukkan bahwa nilai MAPE untuk metode *ensemble* KNN tanpa faktor koreksi sebesar 12,84%, dengan faktor koreksi *trend* sebesar 8,11% dan KNN-KSD sebesar 3,30%. Nilai MAE untuk metode *ensemble* KNN tanpa faktor koreksi sebesar 646,08, dengan faktor koreksi *trend* sebesar 414,77 dan KNN-KSD sebesar 168,15. Nilai RMSE untuk metode *ensemble* KNN tanpa faktor koreksi sebesar 683,40, dengan faktor koreksi *trend* sebesar 484,63 dan KNN-KSD sebesar 214,48. Hasil prediksi metode *ensemble* KNN-KSD pada pola data *trend* yang mengandung musiman lebih akurat dibandingkan metode *ensemble* KNN tanpa faktor koreksi maupun dengan faktor koreksi *trend*. Prediksi harga gabah kering giling di Indonesia menggunakan metode *ensemble* KNN-KSD berkisar antara Rp 4600 hingga Rp 5200 per kilogram. Kisaran ini hampir sama dengan harga gabah sebenarnya yaitu sekitar Rp 4600 hingga Rp 5500 per kilogram. Untuk prediksi pada data dengan pola *trend* musiman seperti data harga gabah, dianjurkan untuk menggunakan metode *ensemble* KNN-KSD.

## 2.2. Landasan teori

### 2.2.1 Data mining

Teori-teori tentang *data mining* sejak lama sudah sering dibahas dalam beberapa penelitian, adapun teori yang membah tentang *data mining* diantaranya adalah *Naïve-Bayes* dan *Nearest Neighbour*, *Decision Tree*, *k-Means* dan *Text Mining*. Seiring dengan perkembangan zaman yang begitu maju akhirnya menghadirkan beberapa algoritma-algoritma diantaranya: Jaringan Syaraf Tiruan, Algoritma Genetik, *Fuzzy C-Means*, dan juga SVM. Penggalan data (bahasa Inggris: *data mining*) adalah ekstraksi pola yang menarik dari data dalam jumlah besar. Suatu pola dikatakan menarik apabila pola tersebut tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. Penggalan data memiliki beberapa nama alternatif, meskipun definisi eksaknya berbeda, seperti KDD (*knowledge discovery in database*), analisis pola, arkeologi data, pemanenan informasi, dan intelegensia bisnis. Penggalan data diperlukan saat data yang tersedia terlalu banyak (misalnya data yang diperoleh dari sistem basis data perusahaan, *e-commerce*, data saham, dan data bioinformatika), tapi tidak tahu pola apa yang bisa didapatkan.

Dalam pemanfaatannya *data mining* sering digunakan untuk mendeteksi suatu kejadian yang ganjil seperti halnya deteksi penyakit tertentu, transaksi yang dianggap mencurigakan, bahkan bisa digunakan untuk mendeteksi penggunaan telepon yang dianggap berbahaya seperti penipuan via telepon dan juga penggunaan kartu kredit yang disalahgunakan, bahkan di bidang retailpun sudah menggunakan *data mining* untuk mengatur tata letak stand supaya sipembeli lebih mudah dalam mobilitasnya.

Banyak sekali vendor-vendor saat ini yang mulai mengkhususkan diri dalam bidang keilmuan *data mining* antara lain: Oracle, Microsoft, SPSS dan sebagainya. Produk yang dihasilkannya juga sangat banyak dan beragam diantaranya dalam bidang *Bussines Intelligent*, *Costumer Relation Management* (CRM) dan (BPM), begitupula dalam bidang kesehatan seperti halnya e-health, dan masih banyak lagi diterapkan di semua bidang keilmuan lainnya.

Ada 3 proses aktivitas *data mining* dalam pelaksanaannya yaitu:

- a. Eksplorasi data, dalam aktivitas ini terdiri dari pembersihan pada data itu sendiri, transformasi data, pemilihan ciri, dan juga pengurangan dimensi data

- b. Pembuatan model dan proses pengujian pada kebenaran data itu sendiri
- c. Penerapan model pada tahapan ini model akan diuji dengan data baru agar menghasilkan perkiraan dari sebuah kasus, untuk melihat model yang telah dibuat dapat memperoleh hasil yang bagus dalam menghadapi masalah.

### 2.2.2 *Time Series*

Pengertian analisis *Time Series* (trend) merupakan suatu metode analisis yang ditujukan untuk melakukan suatu estimasi atau peramalan pada masa yang akan datang. Untuk melakukan peramalan dengan baik maka dibutuhkan berbagai macam informasi (data) yang cukup banyak dan diamati dalam periode waktu yang relatif cukup panjang, sehingga dari hasil analisis tersebut dapat diketahui sampai berapa besar fluktuasi yang terjadi dan faktor faktor apa saja yang mempengaruhi terhadap perubahan tersebut.

Secara teoritis dalam analisis *Time Series* yang paling menentukan adalah kualitas data atau keakuratan dari informasi atau data yang diperoleh serta waktu atau periode dari data tersebut dikumpulkan. Jika data yang dikumpulkan tersebut semakin banyak maka semakin baik pula estimasi atau peramalan yang diperoleh, sebaliknya jika data yang dikumpulkan semakin sedikit, maka hasil estimasi atau peramalannya akan semakin jelek.

Analisis *Time Series* dapat digolongkan menjadi dua yaitu Analisis Jangka Pendek dan Analisis Jangka Panjang. Untuk Analisis Jangka Pendek terdapat kecenderungan model analisisnya dalam bentuk persamaan garis linier. Untuk jangka panjang model analisisnya cenderung mengalami fluktuasi sehingga model persamaannya jarang yang berbentuk garis linier (nonlinier). Contoh dalam jangka panjang faktor pertumbuhan penduduk sangat berpengaruh terhadap permintaan terhadap makanan, pakaian, perumahan, barang, jasa dan lain sebagainya. Demikian juga dengan adanya penemuan teknologi modern akan sangat mempengaruhi fluktuasi dalam perekonomian jangka panjang. Di samping hal tersebut yang dapat mempengaruhi fluktuasi perekonomian adalah faktor musim (iklim) yaitu perubahan iklim akan berpengaruh terhadap kebiasaan masyarakat dalam melakukan konsumsi.

### 2.2.3 *Support Vector Machine*

*Support Vector Machine* (SVM) adalah suatu metode tentang klasifikasi yang terpadu dikarenakan pada saat proses pelatihan, dibutuhkan target untuk pembelajaran, beda halnya dengan FCM yang dalam proses klasifikasi tidak dibutuhkan target pelatihan. SVM juga merupakan algoritma bekerja dengan pemetaan nonlinier guna mengubah data latih ke suatu dimensi yang lebih tinggi, dalam hal ini dimensi yang baru, yaitu dengan cara mencari *hyperplane* untuk memisahkan secara linier dan dengan pemetaan nonlinier ke dalam dimensi yang lebih tinggi ( *multivariate* ), SVM suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, yang sangat populer belakangan ini. SVM berada dalam satu kelas dengan ANN dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan. Keduanya masuk dalam kelas supervised learning. Baik para ilmuwan maupun praktisi telah banyak menerapkan teknik ini dalam menyelesaikan masalah-masalah nyata dalam kehidupan sehari-hari. Baik dalam masalah gene expression analysis, finansial, cuaca hingga di bidang kedokteran. Terbukti dalam banyak implementasi, SVM memberi hasil yang lebih baik dari ANN, terutama dalam hal solusi yang dicapai. ANN menemukan solusi berupa *local optimal* sedangkan SVM menemukan solusi yang *global optimal*. Tidak heran bila kita menjalankan ANN solusi dari setiap *training* selalu berbeda. Hal ini disebabkan solusi *local optimal* yang dicapai tidak selalu sama. SVM selalu mencapai solusi yang sama untuk setiap running. Dalam teknik ini, kita berusaha untuk menemukan fungsi pemisah (klasifier) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda.

SVM bekerja dengan prinsip *Structural Risk Minimization* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua *class* pada *input space* [7][8]. Usaha mencari lokasi *hyperplane* ini merupakan inti dari SVM. Tujuan utama dari SVM adalah agar meningkatkan kecepatan dalam pelatihan dan pengujian sehingga SVM dapat menjadi pilihan dalam memprediksi dengan data yang sangat besar

(misalnya, jutaan *support vector*). Masalah-masalah lain yang harus dikerjakan penelitiannya diantaranya menemukan kernel terbaik untuk suatu data set yang diujikan dan menemukan metode yang sangat efisien pada kasus kelas jamak.

### 2.2.4 K-fold Validation

Metode *K-fold cross validation* menggeneralisasikan pendekatan ini dengan segmentasi data ke dalam k partisi berukuran sama. Selama proses, salah satu dari partisi tersebut dipilih untuk diuji, sisanya digunakan untuk *training*. Prosedur ini diulangi k kali sedemikian sehingga setiap partisi digunakan untuk testing tepat satu kali. Total *error* ditentukan dengan menjumlahkan *error* untuk semua k proses tersebut.

Kasus khusus untuk metode *k-fold cross-validation* menetapkan  $K=N$ , ukuran dari data set. Metode ini dinamakan pendekatan *leave-one-out*, setiap *test set* hanya mengandung satu *record*. Pendekatan ini memiliki keuntungan dalam penggunaan sebanyak mungkin data untuk *training*. *Test set* bersifat *mutually exclusive* dan secara efektif mencakup keseluruhan *data set*. Kekurangan dari pendekatan ini adalah banyaknya komputasi untuk mengulangi prosedur sebanyak N kali. Penelitian ini menggunakan *k-fold 10* dan *k-fold 15*. *K-fold 10* merupakan *stratified 10-fold, cross-validation* artinya *k-fold 10* ini adalah metode evaluasi standar. Hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa *10-fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat.

### 2.2.5 Tipe Kernel

Penentuan model prediksi pada *Support Vector Machine*, fungsi kernel digunakan oleh *Support Vector Regression* (SVR) untuk *mapping* data dalam pola nonlinier ke dalam pola yang linier dengan memindahkan data ke dimensi yang lebih tinggi. *Hyperplane* dapat memisahkan kelas data. Fungsi kernel yang biasa digunakan yaitu: *Linier*, *Radial Basis Function (RBF)*, *Polynomial*, dan *Sigmoid*. Fitur fungsi kernel tersebut yang bisa dipilih oleh *user* untuk menganalisis data. Penelitian ini menggunakan keempat fungsi kernel tersebut untuk kemudian dibandingkan hasil prediksinya[9].

Dengan metode kernel suatu data  $x$  di-*input space* di-*mapping* ke *feature space*  $F$  dengan dimensi yang lebih tinggi melalui map  $\phi$  sebagai berikut  $\phi : X \rightarrow \phi(X)$ . Karena itu data  $x$  di *input space* menjadi  $\phi(x)$  di *feature space*.

Fungsi kernel yang biasa digunakan dalam beberapa penelitian adalah:

$$\text{Dot} : K(x,xi) = \phi(x)^T \phi(xi) \dots\dots\dots(9)$$

$$\text{Linier} : K(x,xi) = \phi(x)^T \phi(xi) \dots\dots\dots(10)$$

$$\text{Polynomial} : K(x,xi) = \phi(x)^T \phi(xi) \dots\dots\dots(11)$$

Fungsi kernel yang harus digunakan untuk substitusi *dot product* di *feature space* sangat tergantung pada data. Biasanya metode *cross-validation* digunakan untuk memilih fungsi kernel tersebut. Fungsi kernel ini yang akan menentukan *feature space* fungsi prediksi yang akan dicari [7].

### 2.2.6 Root Means Square Error (RMSE)

Dalam melihat hasil suatu prediksi dilakukan suatu evaluasi. Evaluasi tersebut digunakan untuk mengetahui keakuratan hasil peramalan yang telah dilakukan terhadap data yang sebenarnya. Beberapa metode dapat digunakan untuk melakukan perhitungan kesalahan peramalan. Beberapa metode yang digunakan diantaranya adalah *Root Mean Squared Error (RMSE)*. RMSE merupakan metode yang cukup sering digunakan dalam mengevaluasi kinerja prediksi. Dengan menggunakan RMSE, *error* yang ada menunjukkan seberapa besar perbedaan hasil estimasi dengan hasil yang akan diestimasi. Hal yang membuat berbeda karena adanya keacakan pada data atau karena tidak mengandung estimasi yang lebih akurat. RMSE merupakan akar nilai dari MSE yang sudah dicari sebelumnya. RMSE digunakan untuk mencari nilai keakuratan hasil peramalan dengan data history dengan menggunakan rumus yang

ditunjukkan dengan nilai kesalahan [10][11]. Semakin kecil nilai yang dihasilkan semakin bagus pula hasil peramalan yang dilakukan. Rumus RMSE dipresentasikan pada persamaan (3)

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}}$$

$y_t$  = Nilai aktual indeks

$\hat{y}_t$  = Nilai prediksi indeks

$n$  = Jumlah sampel

### 2.2.7 Forward Selection

*Forward Selection* adalah salah satu prosedur bertahap yang bertujuan untuk menambah variabel yang dikendalikan satu per satu ke dalam persamaan yang didasarkan pada Alpha tertentu untuk masukan. Alpha untuk masukan merupakan nilai yang menentukan salah satu prediktor yang saat ini tidak dalam model, harus ditambahkan ke model. Nilai P dari masing-masing prediktor dalam model ini tidak dibandingkan dengan tingkat ini, jika nilai P dari prediktor kurang dari tingkat, sehingga prediktor merupakan kandidat untuk dimasukkan ke dalam model. Korelasi sederhana dapat ditentukan dengan menggunakan matriks tabel korelasi sederhana. Prosedur ini akan berakhir ketika semua variabel yang masuk ke dalam model dan memiliki nilai P kurang dari Alpha tertentu untuk masukan[12]

- a. Menentukan variabel bebas pertama yg akan masuk ke dalam model dengan cara menghitung koefisien korelasi parsial antara y dengan setiap variabel bebas  $\square$   $ry_{x1}$ ,  $ry_{x2}$ ,  $ry_{x3}$ ,  $ry_{x4}$ . Selanjutnya pilih variabel bebas yang memiliki koefisien korelasi terbesar dg y. misal  $ry_{x2} > ry_{x1} > ry_{x3} > ry_{x4}$
- b. Regresikan y terhadap X2 ( $y = \beta_0 + \beta_2 X_2 + \varepsilon$ ), lakukan uji hipotesis  $H_0: \beta_2 = 0$  vs  $H_1: \beta_2 \neq 0$   $\square$  menghitung statistik uji t. Jika  $H_0$  ditolak maka X2 masuk ke dalam model, proses berlanjut dengan menentukan variabel bebas berikutnya yang akan masuk ke dalam model dengan menghitung koefisien korelasi parsial antara y dengan X1, X3, X4; X2 sudah ada di dalam model  $\square$  hitung  $ry_{x1|x2}$ ,  $ry_{x3|x2}$ ,  $ry_{x4|x2}$ . Variabel bebas berikut yang akan diuji adalah yang memiliki koefisien korelasi parsial terbesar. Misal  $ry_{x1|x2} > ry_{x3|x2} > ry_{x4|x2}$  sehingga X1 akan diuji signifikansinya di dalam model yang sudah mengandung X2.
- c. Regresikan y thdp X2 dan X1 ( $y = \beta_0 + \beta_2 X_2 + \beta_1 X_1 + \varepsilon$ ). Lakukan pengujian terhadap  $H_0: \beta_1 = 0$  vs  $H_1: \beta_1 \neq 0$ . Jika  $H_0$  ditolak maka X1 masuk ke dalam model, proses berlanjut dengan menentukan variabel bebas berikutnya yang akan masuk ke dalam model dg menghitung koefisien korelasi parsial antara y dengan X3, X4 dimana X1 dan X2 sudah ada di dalam model. Tapi jika X1 tidak signifikan ( $H_0$  tidak ditolak), maka proses berhenti dan model terbaik yg diperoleh dengan metode *Forward Selection* adalah  $y = \beta_0 + \beta_2 X_2 + \varepsilon$ .

## 2.3. Kerangka Pemikiran Prediksi Harga Kedelai Lokal dan Kedelai Impor Menggunakan SVM dengan *Forward Selection*

Kedelai yang banyak digunakan untuk konsumsi dan bahan olahan di masyarakat Indonesia yaitu kedelai lokal dan kedelai impor adapun varian dari kedelai yang banyak digunakan untuk bahan olahan dan konsumsi adalah lokal dan impor.

### 2.3.1 Kedelai Lokal

Tanaman kedelai (*Glicine max*) yang berasal dari Tiongkok dan kemudian dikembangkan ke berbagai negara, adalah tanaman semusim yang termasuk *Family Leguminosae*. Keunggulan teknis budidaya yang sederhana telah memungkinkan kedelai dapat dibudidayakan di daerah sub tropis dan tropis dengan skala masif. Kandungan gizi kedelai cukup tinggi, terutama proteinnya mencapai 34%, sehingga sangat

diminati sebagai sumber protein nabati yang relatif murah dibandingkan dengan protein hewani (Direktorat Jenderal Tanaman Pangan, 2004). Selain sebagai sumber protein nabati, kedelai juga sebagai pangan fungsional untuk mencegah timbulnya penyakit degeneratif, seperti jantung koroner dan hipertensi. Bahkan kandungan zat isoflavon pada kedelai ternyata berfungsi sebagai antioksidan. Perkembangan teknologi terakhir menunjukkan bahwa saat ini kedelai banyak digunakan sebagai sumber energi alternatif (*biofuel*).

Keanekaragaman manfaat kedelai telah mendorong tingginya permintaan kedelai di dalam negeri. Selain itu, manfaat kedelai sebagai salah satu sumber protein murah membuat kedelai semakin diminati. Semakin besarnya jumlah penduduk Indonesia berpotensi pada semakin meningkatnya permintaan kedelai.

Konsumsi kedelai yang terus meningkat pesat setiap tahunnya, juga sejalan dengan meningkatnya kesadaran masyarakat akan gizi yang ditandai oleh meningkatnya konsumsi per kapita kedelai sebesar 5,55%. Sebagian besar produksi kedelai diolah menjadi bahan pangan yang siap dikonsumsi oleh masyarakat, baik secara langsung maupun tidak langsung seperti tempe, tahu, kecap, dan kripik tempe. Sekitar 115.000 pengusaha tahu dan tempe anggota Koperasi Produsen Tempe dan Tahu Indonesia (KOPTI) adalah konsumen terbesar kedelai. Mereka membutuhkan 1,2 juta ton kedelai per tahun, atau lebih dari separuh dari total kebutuhan nasional sebanyak 2,2 juta ton per tahun. Pabrik kecap, perusahaan pakan ternak, dan industri makanan- minuman berada di urutan berikutnya sebagai konsumen kedelai.

### 2.3.2 Kedelai Impor

Saat ini, Indonesia menjadi salah satu negara pengimpor kedelai terbesar di dunia. Setiap tahunnya jumlah kedelai yang diekspor rata-rata di atas 1 juta ton dari total kebutuhan rata-rata di atas 2 juta ton. Dari jumlah itu, sekitar 88 persen digunakan sebagai bahan baku pembuatan tempe dan tahu, 10 persen untuk pangan olahan lainnya seperti industri tepung dan pati serta sisanya sebanyak 2 persen untuk benih. Sebagian besar kedelai diimpor berasal dari Amerika, Kanada, Argentina dan Brasil. Lingkup penelitian ini meliputi menelaah penawaran kedelai dunia dan permintaan impor kedelai Indonesia antara tahun 2005-2012, menganalisis kebijakan perkedelaaian Indonesia serta merumuskan alternatif strategi pengembangan agribisnis kedelai lokal di Indonesia. Amerika Serikat juga mengekspor lebih dari 30 juta ton kedelai setiap tahunnya atau sekitar 42,94 persen dari total ekspor dunia saat ini. Angka ini hanya sekitar 36 persen dari total produksi Amerika Serikat pada tahun 2009. Kebutuhan kedelai Indonesia rata-rata setiap tahunnya di atas angka 2 juta ton, 90 persen diantaranya digunakan sebagai bahan pangan, terutama pangan olahan yaitu sekitar 88 persen tahu dan tempe, 10 persen untuk pangan olahan lainnya seperti industri tepung dan pati serta sisanya sebanyak 2 persen untuk benih. Sayangnya, sekitar 63,41 persen dipasok oleh kedelai impor yang memiliki harga lebih murah dan kualitas lebih baik sedangkan sisanya 36,59 persen dipenuhi melalui produksi dalam negeri.

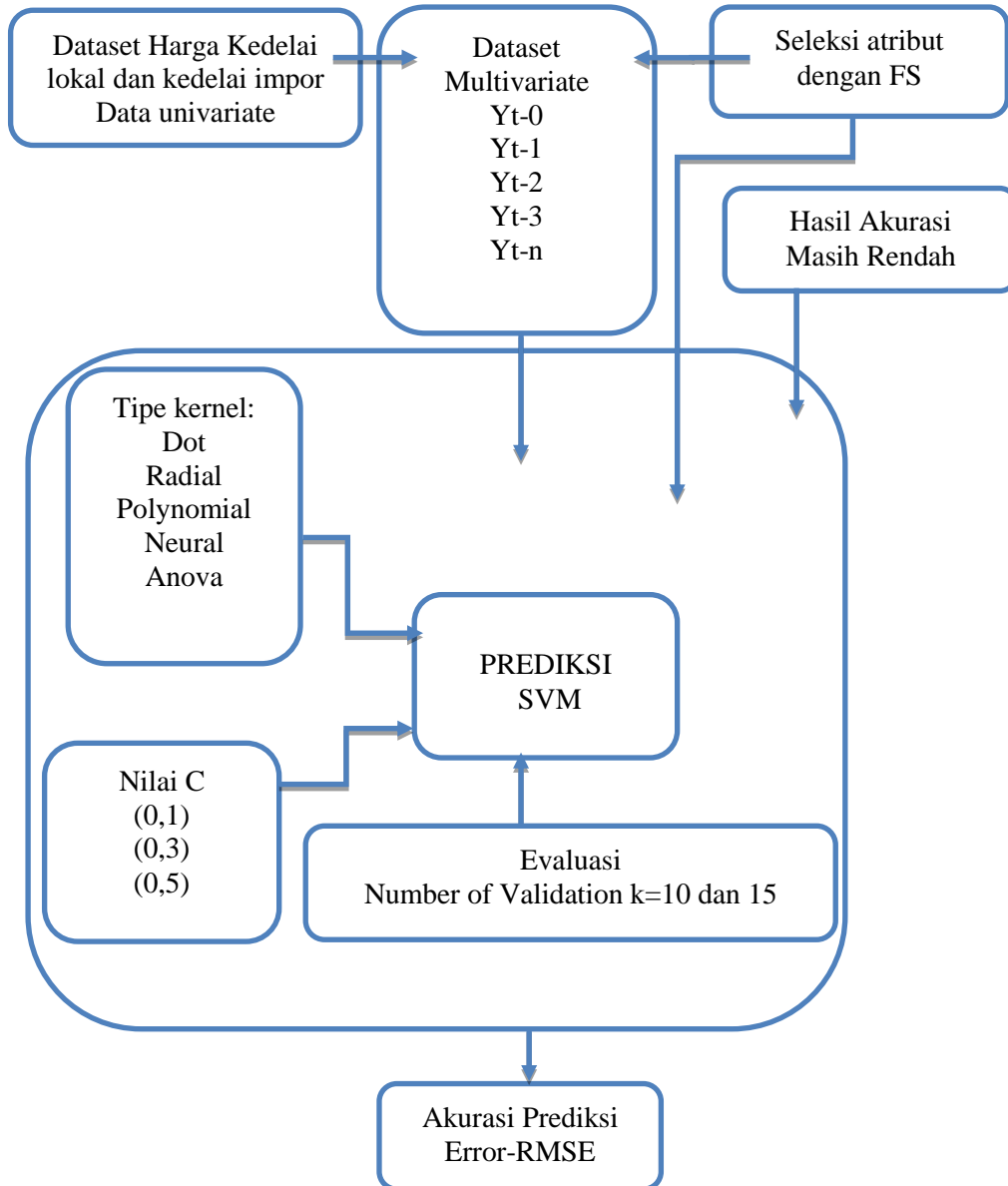
### 2.3.3 Model yang Diusulkan

Penelitian ini, menggunakan metode komputasi cerdas dan untuk mengolah data dan informasi yang terkait dengan penelitian ini. Beberapa masalah yang ada tentang komoditas pangan dalam hal ini kedelai adalah belum adanya suatu model prediksi harga kedelai baik lokal maupun impor. *Dataset* diambil dari *website* resmi kementerian perdagangan RI ([disperindag.go.id](http://disperindag.go.id)). Data kedelai lokal maupun kedelai impor yang masih berupa data *univariate* kemudian diubah menjadi *dataset* dalam bentuk *multivariate* ( $yt-0, yt-1, yt-2, yt-3 \dots yt-n$ ). Pada proses pertama pengujian menggunakan *inputan 2, inputan 3, inputan 4, inputan 5, inputan 6, inputan 7, inputan 8, inputan 9, inputan 10, inputan 15, inputan 20*. Penggunaan *inputan* ini mengacu pada penelitian sebelumnya yang dilakukan oleh Sarim [4]. Penelitian tersebut sudah menghasilkan nilai *error* yang bagus akan tetapi masih ada beberapa atribut yang perlu evaluasi. Proses pemilihan atributnya diseleksi dengan *Forward Selection*. Pada seleksi ini dimulai dari tidak ada variabel dalam model, data yang diujikan secara bertahap dari variabel 1 sampai dengan variabel yang lain. Pada proses ini lakukan pengujian terhadap  $H_0: \beta_1 = 0$  vs  $H_1: \beta_1 \neq 0$ . Jika  $H_0$  ditolak maka  $X_1$  masuk ke dalam model. Proses berlanjut dengan menentukan variabel bebas berikutnya yang akan masuk ke dalam model



dengan menghitung koefisien korelasi parsial antara  $y$  dengan  $X_3$ ,  $X_4$ .  $X_1$  dan  $X_2$  sudah ada di dalam model.

Model yang diusulkan adalah Model Prediksi Harga Kedelai Lokal dan Kedelai Impor Menggunakan SVM dengan parameter tipe kernel *Dot*, *Radial*, *Polynomial*, *Neural*, *Anova* dan juga nilai  $C$  yang di gunakan yaitu 0,1, 0,3 dan 0,5 dengan *Forward Selection*.



Gambar 1. Kerangka Pemikiran Harga Kedelai Lokal dan Kedelai Impor Menggunakan SVM dengan *Forward Selection*

### 3. METODE PENELITIAN

#### 3.1. Pengumpulan Data

*Dataset* dalam penelitian ini diperoleh dari data harga pasar *web site* resmi kementerian perdagangan republik Indonesia yang di ambil daftar harga harian dari 3 januari 2011 sampai dengan 10 maret 2015

#### 3.2. Pengolahan Data Awal

Dari *dataset* harga rata-rata pasar kedelai lokal dan kedelai impor untuk selanjutnya perlu dilakukan pengolahan data. *Dataset* yang ada merupakan data *univariate*, berikut contoh *dataset* yang masih berupa data *univariate*. Berikut contoh *dataset* yang masih berupa data *univariate*.

Tabel 1. *Dataset* Kedelai Lokal

NO	Tanggal	Harga Kedelai Lokal
1	03 Januari 2011	8,157
2	04 Januari 2011	8,884
3	05 Januari 2011	8,602

Untuk dapat diproses oleh algoritma SVM, maka data tersebut perlu dikonversi menjadi data *multivariate*. Hal ini dilakukan untuk menghitung prediksi harga yang akan datang / target (Yt), penentuan harga yang akan datang dipengaruhi oleh harga-harga pada periode-periode sebelumnya. Dalam penelitian ini *multivariate* yang peneliti lakukan dalam bentuk 2 *inputan*, 3 *inputan*, 4 *inputan* sampai dengan 7 *inputan*. Yang dimaksud dengan 2 *inputan* adalah melihat harga dari dua hari sebelumnya, sedangkan 3 *inputan* berarti melihat harga dari 3 hari sebelumnya dan begitu seterusnya.

Tabel 2. Tabel Hasil dari *Inputan* 5 Hari Sebelumnya

h-4	h-3	h-2	h-1	h-0
8157	8884	8602	8593	8594
8884	8602	8593	8594	8604
8602	8593	8594	8604	8665

Dan seterusnya sampai dengan mendapatkan hasil RMSE terendah dalam penelitian ini menggunakan *inputan* hari sebelumnya sampai dengan 20.

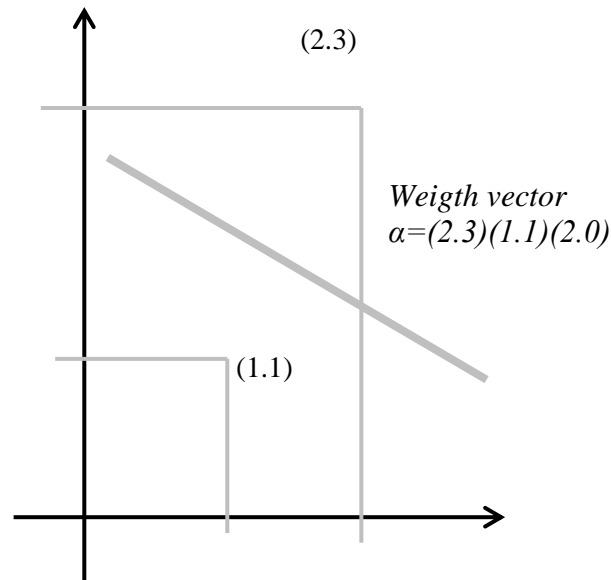
#### 3.3. Eksperimen dan Pengujian

Dalam tahapan ini metode yang pertama kali di ujikan adalah metode *support vectore machine* tanpa adanya optimasi. *Data set* awal kita setelah kita rubah *inputanya* kita ujiakan dengan beberapa *inputan* sampai dengan menemukan nilai RMSE terendah dari beberapa *inputan* dari 5 *inputan* sampai dengan 40 *inputan*, pada *X-Validation* juga dilakukan beberapa percobaan dari mulai *number of validation* 10 dan 15, hal yang sama juga kita lakukan percobaan dengan merubah nilai *c / cost*, pada penelitian ini nilai *cost* yang diteliti yaitu dari nilai *cost* =0,1, *cost*=0,3 *cost*=0,5. Pada tipe kernel pada SVM pada percobaan ini menggunakan 5 tipe kernel yaitu *dot*, *radial*, *polynomial*, *neural* dan *anova*, penggunaan 5 tipe kernel ini karena beberapa penelitian sebelumnya hanya menyertakan 5 tipe kernel dan mendapatkan nilai RMSE

terendah. Dari setiap tipe kernel yang ada nantinya akan dibandingkan dengan hasil penelitian dengan optimasi *Forward Selection*. Dalam tahapan ini di cari nilai RMSE terendah.

Selanjutnya melakukan pengujian optimasi svm dengan *Forward Selection* pada tahapan ini yang kita lakukan pengujian hanya pada nilai terendah dari tiap-tiap *input-an* saja. Dari mulai *dataset* yang sudah berupa *multivariate* dengan 5 *input-an* sampe dengan 20, number of validation 10 dan 15, nilai  $c = 0,1$  , 0,3 , 0,5 dan juta tipe kernel dari mulai *dot*, *radial*, *polynomial*, *neural* dan *anova*, dari hasil terendah tersebut dibandingkan dengan pengujian SVM tanpa optimasi.

Pola kerja dan perhitungan manual SVM untuk mencari vector dengan mengambil contoh 2 *inputan*, sebagai berikut.



$$\begin{aligned} \text{Weigth vector } \vec{\omega} &= (a, 2a) \\ \alpha + 2\alpha + \omega_0 &= -1 \text{ digunakan untuk poin (1.1)} \\ 2\alpha + 6\alpha + \omega_0 &= 1 \text{ digunakan untuk poin (2.3)} \\ \dots \\ \omega_0 = 1 - 8\alpha & \quad 3\alpha + 1 - 8\alpha = -1 \\ 5\alpha &= 2 \\ \alpha &= \frac{5}{2} \\ \omega_0 = 1 - 8\alpha &= 1 - 8 \frac{5}{2} = \frac{5}{5} - 16 = -\frac{11}{5} \\ \vec{\omega} &= \left( \frac{2}{5}, \frac{4}{5} \right) \\ g(\vec{x}) &= \frac{2}{5}x_1 + \frac{4}{5}x_2 - \frac{11}{5} \end{aligned}$$

### 3.4. Evaluasi dan Validasi Penelitian

Setelah mendapatkan pemodelan dengan performa terbaik dengan parameter diantaranya nilai *training*, tipe *kernel*, *Cost*, *K-fold* dan RMSE, selanjutnya diterapkan prediksi harga rata-rata hari berikutnya. Dari beberapa kesalahan peramalan dibandingkan dengan nilai sebenarnya maka didapatkan tingkat akurasi dari model prediksi SVM terhadap komoditas kedelai lokal dan kedelai impor. Dari hasil prediksi awal tanpa adanya optimasi nantinya akan dibandingkan dengan optimasi *Forward Selection* baik dari prediksi

kedelai lokal maupun kedelai impor. Setelah hasil prediksi dengan optimasi antara kedua komoditas tersebut bisa dianalisis kolerasi antara harga kedelai lokal dan kedelai impor.

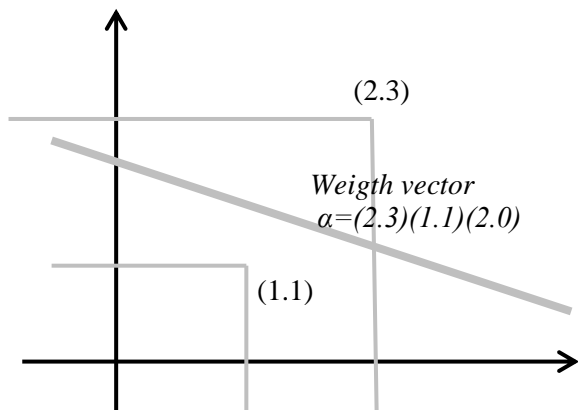
#### 4. HASIL DAN PEMBAHASAN

Data harga rata-rata komoditi kedelai lokal dan kedelai impor dalam *website* kemendag RI yang diambil dari 1 Januari 2011 sampai dengan September 2015 sejumlah 1098 record. Selanjutnya dari data set rata-rata harga yang ada tersebut harus dilakukan pengolahan data. Data yang ada pada web masih berupa data *univariate*. Data *univariate* adalah sekumpulan informasi yang masih dalam wujud data tunggal yang hanya memiliki satu atribut. Dalam hal ini data tunggal yaitu harga kedelai lokal dan kedelai impor yang masih harian. *Dataset* yang berupa *univariate* agar bisa di proses menggunakan algoritma *Support Vector Machine* harus terlebih dahulu dirubah menjadi data *multivariate*.

Pada penelitian kali ini *input*-an yang digunakan adalah 5 *input*-an, 10 *input*-an dan kelipatan 5 sampai dengan 40 *input*-an. Berikut adalah data *multivariate* yang digunakan untuk mengolah data menggunakan *Support Vector Machine*.

##### 1. Proses Percobaan

Proses pencarian *hyperplan* dengan peubah 2 variabel (dicontohkan dengan angka tunggal)



$$\begin{aligned} \text{Weigth vector } \vec{\omega} &= (a, 2a) \\ \alpha + 2\alpha + \omega_0 &= -1 \text{ digunakan untuk poin (1.1)} \\ 2\alpha + 6\alpha + \omega_0 &= 1 \text{ digunakan untuk poin (2.3)} \\ \omega_0 = 1 - 8\alpha & \quad 3\alpha + 1 - 8\alpha = -1 \\ 5\alpha &= 2 \\ \alpha &= \frac{2}{5} \\ \omega_0 = 1 - 8\alpha &= 1 - 8\left(\frac{2}{5}\right) = \frac{5 - 16}{5} = -\frac{11}{5} \\ \vec{\omega} &= \left(\frac{2}{5}, \frac{4}{5}\right) \\ g(\vec{x}) &= \frac{2}{5}x_1 + \frac{4}{5}x_2 - \frac{11}{5} \end{aligned}$$

Dalam penelitian ini parameter yang digunakan dalam algoritma SVM yang diujicobakan dalam penelitian ini adalah kernel *Dot*, *Radial*, *Polynomial*, *Neural*, *Anova*. Sedangkan untuk nilai *c* yang diujikan adalah 0.1, 0.3 dan 0.5 adapun *k-fold* yang digunakan adalah 10 dan 15. Untuk menguji masing-masing kelompok tersebut yang terdiri dari beberapa fungsi kernel dan memasukkan nilai *cost* serta range yang telah ditentukan, berikut akan disajikan hasil pengujiannya.

**4.1. Dataset dengan Input-an 2 Hari dengan *k-fold* =10 dan 15**

Prediksi harga kedelai lokal dengan *input-an* 2 hari sebelumnya dan untuk *k-fold validation* adalah 10, berikut adalah tabel hasil pengujian.

Tabel 3. Ringkasan Hasil Prediksi 2 *Input-an*

Tipe kernel	C	RMSE
Dot	0,1	156.996 +/- 107.663
Radial	0,1	878.511 +/- 45.035
Polynomial	0,1	14009.795 +/- 17004.741
Neural	0,1	872.636 +/- 36.144
Anova	0,1	877.445 +/- 45.593

Dapat disimpulkan untuk percobaan ini nilai yang terkecil RMSE adalah dengan tipe kernel = Dot, nilai C :0,1 dengan RMSE= 156.996 +/- 107.663 Prediksi harga kedelai lokal dengan *inputan* 2 hari sebelumnya dan untuk *k-fold validation* adalah 15, berikut adalah tabel hasil pengujiannya:

Dapat disimpulkan untuk percobaan ini nilai yang terkecil RMSE adalah dengan tipe kernel = Dot, nilai C :0,1 dengan RMSE= 155.455 +/- 108.908, dari percobaan yang dihasilkan dengan nilai *inputan* 2 dan *k-fold*= 10 dan 15 maka dapat diperoleh nilai RMSE terkecil dari percobaan adalah dengan 10 *inputan*, nilai C= 0.1, *k-fold*=15 dan menggunakan tipe kernel = Dot, nilai RMSE = 155.455 +/- 108.908.

**4.2. Dataset dengan Input-an 2 dengan *k-fold* =10 dan 15**

Prediksi harga kedelai impor dengan *input-an* 2 hari sebelumnya dan untuk *k-fold validation* adalah 10, berikut adalah tabel hasil pengujian.

Tabel 4. Ringkasan Hasil Prediksi 2 *Input-an*

Tipe kernel	C	RMSE
<i>Dot</i>	0,1	134.104 +/- 62.209
<i>Radial</i>	0,1	1207.437 +/- 73.236
<i>Polynomial</i>	0,1	5196201597.184 +/- 7112802096.211
<i>Neural</i>	0,1	1193.340 +/- 38.003
<i>Anova</i>	0,3	1217.719 +/- 77.879

Dapat disimpulkan untuk percobaan ini nilai yang terkecil RMSE adalah dengan tipe kernel = Dot, nilai C :0,1 dengan RMSE= 134.104 +/- 62.209.

Prediksi harga kedelai impor dengan *inputan* 2 hari sebelumnya dan untuk *k-fold validation* adalah 15, berikut adalah tabel hasil pengujian.

Dapat disimpulkan untuk percobaan ini nilai yang terkecil RMSE adalah dengan tipe kernel = Dot, nilai C :0,1 dengan RMSE= 129.375 +/- 81.811, dari percobaan yang dihasilkan dengan nilai *inputan* 5 dan *k-fold*= 10 dan 15 maka dapat diperoleh nilai RMSE terkecil dari percobaan adalah dengan nilai C= 0.1, *k-fold*=15 dan menggunakan tipe kernel = Dot, nilai RMSE = 129.375 +/- 81.811.

Setelah semua percobaan yang telah dilakukan dengan menggunakan *data training* dengan jumlah *inputan* dari mulai Dua (2) hingga Tiga puluh (30) dapat di ketahui nilai RMSE terkecil untuk kedelai lokal adalah dengan nilai *inputan* Empat (4) dengan nilai (C = 0,1 dan *K-fold* =10) dan tipe kernel menggunakan Dot sebesar = 154.025 +/- 114.993, sedangkan untuk kedelai impor dalah dengan lima *inputan* dengan nilai (C=0,1 dan *k-fold*=15 ) menggunakan tipe kernel Dot yaitu sebesar = 126.008 +/- 78.371. berikut ringkasan hasil pengujian.

Tabel 5. Ringkasan RMSE Terkecil

No	Komoditi	Nilai RMSE
1	Kedelai lokal	154.025 +/- 114.993
2	Kedelai Impor	126.008 +/- 78.371

Setelah didapat nilai RMSE terkecil selanjutnya dilakukan pengujian agar dapat menentukan nilai prediksi. Berikut hasil dari prediksi harga kedelai impor adalah

Tabel 6. Prediksi Harga Kedelai Impor

Label	(prediksi)	yt-4	yt-3	yt-2	yt-1	yt-0
11090	11,139	10970	10970	11070	11090	11090

Dari tabel prediksi di atas harga prediksi pada tanggal 1 bulan february 2016 adalah 11,139 sedangkan harga riil pada tanggal tersebut adalah 11,050, jadi ada selisih harga sebesar 89 rupiah. Jadi tingkat keakurasiannya adalah 0,81%, dan berikut adalah gambar prediksi harga kedelai impor

Sedangkan hasil prediksi dari kedelai lokal adalah:

Tabel 7. Prediksi Harga Kedelai Lokal

label	prediksi	yt-3	yt-2	yt-1	yt-0
11020	10,966	11080	11030	11020	11020

Dari tabel prediksi di atas harga prediksi pada tanggal 1 bulan Februari 2016 adalah 10,966 sedangkan harga riil pada tanggal tersebut adalah 11,020, jadi ada selisih harga sebesar 64 rupiah. Jadi tingkat keakurasiannya adalah 0,49%. Berikut gambar grafis harga prediksi dan harga riil.

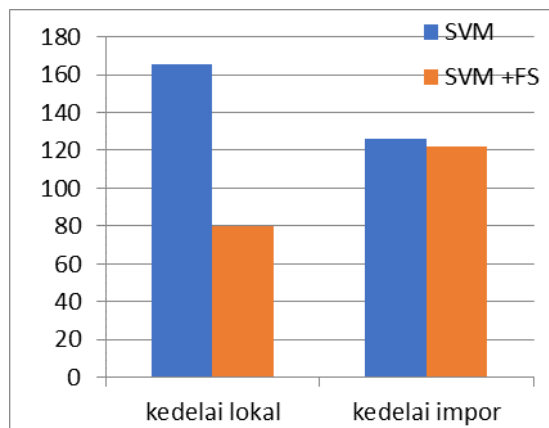
### 4.3. Analisis dan Komparasi

Untuk meningkatkan nilai RMSE tersebut, maka dilakukan penyeleksian pada atribut tersebut, penelitian ini menggunakan *Forward Selection*. Pola kerja dari optimasi ini adalah dengan dimulai dari tidak ada nilai variabel, kemudian memulai dengan menambah nilai variabel satu demi satu dalam model tersebut. *Data training*, dalam penelitian ini dimulai dari 2 *inputan* sampai dengan *inputan* yang menghasilkan nilai terbaik, dalam penelitian ini dengan 30 *inputan* dan hasilnya sebagai berikut.

Tabel 8. SVM + *Forward Selection*

No	Komoditi	Nilai RMSE	
		SVM	SVM + FS
1	Kedelai lokal	165.742 +/- 101.407	79.749 +/- 16.051
2	Kedelai Impor	126.008 +/- 78.371	122.270 +/- 56.049

Berikut grafik dari perbandingan hasil RMSE antara SVM dengan SVM + FS



Gambar 2. Perbandingan SVM dengan SVM+FS

## 5. PENUTUP

### 5.1. Kesimpulan

Dari uraian pembahasan percobaan yang dilakukan di atas dapat disimpulkan hasil sebagai berikut.

- Untuk prediksi kedelai lokal dengan menggunakan parameter *inputan* data 4 (empat) hari sebelumnya, *K-fold*=10, nilai *C*= 0,1 diperoleh nilai RMSE terkecil sebesar 154.025 +/- 114.993. Setelah dilakukan seleksi atribut menggunakan *Forward Selection* diperoleh nilai RMSE sebesar 79.749 +/- 16.051, terdapat peningkatan RMSE sebesar= 74.276
- Untuk prediksi kedelai lokal dengan menggunakan parameter *input-an* data 5 (lima) hari sebelumnya, *K-fold*=15, nilai *C*= 0,1 diperoleh nilai RMSE terkecil sebesar 126.008 +/- 78.371,

setelah dilakukan optimasi menggunakan *Forward Selection* diperoleh nilai RMSE sebesar 122.270 +/- 56.049, terdapat peningkatan RMSE sebesar= 3.738.

- c. Dibandingkan dengan penelitian sebelumnya yang hanya menggunakan SVM dengan model SVM yang berbasis *Forward Selection* menunjukkan performa yang lebih baik.

## 5.2. Saran

Penelitian ini bisa ditindaklanjuti dengan:

- mengujicobakan nilai parameter yang lain pada *Support Vector Machine* yang lainnya seperti halnya: tipe kernel, nilai *C*, *K-fold* dan juga jumlah *inputan*.
- Untuk optimasi hasil bisa menggunakan algoritma yang lainya seperti PSO dengan memilih nilai bobot dari atributnya..

## PERNYATAAN ORISINALITAS

“ Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa artikel ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya”

[Fatkhuroji]

## DAFTAR PUSTAKA

- [1] Supadi, 2009. Dampak Impor Berkelanjutan Terhadap Ketahanan Pangan. Jurnal. Analisis Kebijakan Pertanian. Vol 7 No 1 : 87-102
- [2] Sudaryanto, T. dan D. K. S. Swastika, 2007 Ekonomi Kedelai di Indonesia. Forum Agro Ekonomi (FAE) 12 (3) : 1-27.
- [3] Varra Wuwung, Nelson Nainggola dan Marline Paendongan, *Prediksi harga beras membramo dan beras sultan di kota Manado pata tahun 2013*, penelitian ini menggunakan metode ARIMA. 2013
- [4] Sarim , *Prediksi Harga Bawang Merah dan Cabe di Kab.Brebes dengan SVM*, Jurusan Pasca Sarjana Fakultas Teknik Universitas Dian Nuswantoro (Udinus ) Semarang, 2014.
- [5] Dewi sinta, *Metode Ensemble K-Nearest Neighbor Untuk Prediksi Harga Beras Di Indonesia*, Sekolah Pascasarjana Institut Pertanian Bogor Bogor 2015
- [6] Muhammad Asri,Amran, Erna Tri Herdiani, *Modifikasi Model Faktor Koreksi Pada Metode Ensemble K-Nearest Neighbor Untuk Pola Data Trend yang Mengandung Musiman*, Program Studi Statistika FMIPA Universitas Hasanuddin. 2015.
- [7] Nugroho, *SVM:Paradigma Baru dalam Softcomputing Dan Aplikasinya*, Konferensi Nasional Sistem & Informatika 2008, Bali, 2008.
- [8] Fatimah Wulandini and Anto Satriyo Nugroho, *Text Classification Using Support Vector Machine for Webmining Based Spatio Temporal Analysis of the Spread of Tropical Disease*, Swiss German University, International Conference on Rural Information and Communication Technology 2009
- [9] Ariestya Arlene, Judy Retti Witono ,Maria Fransisca, *Pembuatan roti tawar dari tepung singkong dan tepung kedelai* Jurusan Teknik Kimia, Fakultas Teknologi Industri, Universitas Katolik Parahyangan, (2009)
- [10] Purwanto, C. Eswaran, and R. Logeswaran, *An Optimally Configured Hybrid Model for Healthcare Time Series Prediction*, Asian Journal of Information Technology, 10(6), pp. 209-217, 2010
- [11] Purwanto, C. Eswara, and R. Logeswara., *Improved Adaptive Neuro-fuzzy Inference System for HIV/AIDS Time Series Prediction*. In: Informatics Engineering and Information Science, 253, Springer-Verlag Berlin Heidelberg , pp. 1-13, 2011
- [12] I. M. M. Ghani and S. Ahmad, *Comparison Methods of Multiple Linear Regressions in Fish Landing*, Australian Journal of Basic and Applied Sciences, 5(1): 25-30, 2011



# **RALAT**

Diumumkan bahwa terdapat kesalahan cetak  
Nomor ISSN Jurnal Teknologi Informasi CyberKu  
Pada edisi sebelumnya

Nomor ISSN yang benar adalah:  
**1907-3380**

Terimakasih atas perhatiannya dan mohon maaf atas  
kesalahan ini

**Jurnal Teknologi Informasi CyberKu**

Pengelola