

METODE *K-NEAREST NEIGHBOR* BERBASIS *FORWARD SELECTION* UNTUK PREDIKSI HARGA KOMODITI LADA

Muis Nanja¹, Purwanto²

^{1,2} Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Imam Bonjol 207, Semarang, 50131 Indonesia

¹muis_abi@yahoo.co.id
²purwanto@dsn.dinus.ac.id

Abstrak: Banyak peneliti termotivasi untuk meningkatkan kinerja prediksi. *K-Nearest Neighbor* (KNN) merupakan salah satu algoritma untuk regresi maupun klasifikasi sudah secara sukses diimplementasikan di berbagai bidang. Di sisi lain, penentuan variabel yang sesuai dapat memberikan performa yang semakin baik pada suatu model. Pada penelitian ini bertujuan mengembangkan model prediksi dengan menggabungkan algoritma *K-Nearest Neighbor* dengan metode seleksi atribut, khususnya *forward selection* untuk memprediksi komoditi lada. Model yang diusulkan dievaluasi dengan data *time series* lada hitam dan lada putih. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor* berbasis *forward selection* memberikan kinerja yang terbaik dibandingkan dengan KNN berbasis *backward elimination* dan SVM berbasis seleksi atribut.

Kata Kunci: *K-Nearest Neighbor*, *forward selection*, *time series* (rentet waktu), prediksi.

Abstrack: *Many researchers are motivated to improve performance predictions. K-Nearest Neighbor (KNN) is one of algorithm to regression and classification which has been successfully implemented in various fields. On the other hand, the determination of the appropriate variables that can provide better performance on a model. This research aims to develop a prediction model by combining the K-Nearest Neighbor algorithm with attribute selection methods, especially forward selection to predict commodity of pepper. The proposed model was evaluated with a data time series of black pepper and white pepper. The results showed that the K-Nearest Neighbor algorithm-based forward selection gives the best performance than KNN based backward elimination and SVM-based attribute selection.*

Keywords: *K-Nearest Neighbor, forward selection, time series, prediction.*

I. PENDAHULUAN

Perkembangan teknologi informasi begitu pesat dan memberikan kontribusi yang cukup besar bagi peradaban dunia dari berbagai aspek kehidupan baik dari segi pendidikan, pemerintahan maupun dunia usaha atau bisnis, dalam perkembangan teknologi informasi menawarkan sebuah konsep keleluasaan dalam mengelola dan mencari informasi, karena informasi merupakan hal yang

dinanti oleh para pengelola informasi untuk dikelola dan menjadi bahan ataupun barometer dalam pengambilan suatu keputusan [1], sebagaimana yang diketahui seiring pesatnya pertumbuhan ekonomi, telah membawa dampak yang begitu besar bagi seluruh aspek kehidupan masyarakat terutama dalam dunia bisnis. Peningkatan taraf hidup harus selalu diupayakan karena kesejahteraan masyarakat merupakan komponen yang sangat penting dalam kemajuan suatu Negara [2]. Indonesia sebagai Negara agraris telah menjadi salah satu negara importir pangan penting di dunia [3], dan juga sebagai eksportir. Dengan demikian, diperkirakan perubahan harga komoditas yang sifatnya fluktuatif membawa dampak terhadap perekonomian Indonesia. Seperti halnya komoditi lada merupakan salah satu komoditi yang cukup berpengaruh pada perekonomian.

Berdasarkan berbagai keunggulan dan kontribusi ekspor perkebunan Indonesia, maka dipandang perlu dan suatu kewajiban dalam

meningkatkan daya saing ekspor untuk mengimbangi impor Indonesia. Dalam menghadapi persaingan global maupun regional spesialisasi ekspor hasil perkebunan yang berdaya saing tinggi diharapkan mampu meningkatkan nilai dan daya saing ekspor produk perkebunan khususnya pada hasil perkebunan lada. Untuk menentukan spesialisasi pada komoditas yang berdaya saing tinggi diperlukan suatu informasi mengenai tingkat daya saing tiap komoditas. Selain itu keeksistensian persaingan regional terutama antara negara-negara ASEAN, juga sangat penting artinya untuk menentukan spesialisasi ekspor produk-produk perkebunan. Hal ini terkait dengan keunggulan komparatif berdasarkan kondisi perekonomian, faktor produksi, letak geografis dan kondisi yang berhubungan erat diantara negara-negara dalam satu kawasan regional seperti ASEAN [4].

Pasar domestik dan internasional semakin terbuka dan persaingan pasar tidak terhelakkan lagi sebagai konsekuensi logis akan menjadi semakin meningkat. Karena semua negara diharuskan melakukan pembukaan akses pasar dan penghapusan tarif dan subsidi. Sektor pertanian tidak dikecualikan dalam agenda perdagangan bebas tersebut [3], sehingga data harga komoditas mengalami perubahan yang fluktuatif dan menunjukkan pola yang tidak stasioner, oleh karena itu diperlukan suatu metode untuk melakukan memprediksi harga suatu komoditi guna sebagai bahan pertimbangan bagi pengusaha ataupun bagi pihak-pihak yang berkepentingan untuk mengambil keputusan atau tindakan guna kelangsungan bisnis dan ekonomi.

Pada umumnya, prediksi menggunakan metode *time series* adalah suatu cara untuk mengetahui sebuah nilai dimasa yang akan datang, dimana pengamatannya dilakukan berdasarkan urutan

waktu. Proses determinasi dari metode *time series* memiliki kemampuan prediksi yang tinggi, nilai yang akan datang dapat diketahui dengan melihat nilai dari masa lalu [5].

Prediksi *time series* telah banyak dilakukan para peneliti dengan menggunakan berbagai model, misalnya *moving average* [6], Box-Jenkins ARIMA dan adaptive *neuro-fuzzy inference system* (ANFIS) model digunakan untuk prediksi di bidang kesehatan [6, 7]. *Artificial neural networks* (ANN) yang memiliki kemampuan menangani data non linier juga telah digunakan di berbagai aplikasi [8]-[11].

Arvianto candra dkk [12] telah membandingkan algoritma bayesian network dan *k-nearest neighbor* (KNN) pada prediksi penyakit paru manusia. KNN sangat dikenal dikarenakan memiliki kesederhanaan dan kemampuan dalam memodelkan beragam masalah di berbagai bidang. Muhammad Ilyas Sikki [13], tentang pengenalan wajah menggunakan KNN dengan pra-proses transformasi wavelet. KNN memiliki kelebihan antara lain yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar, proses mudah direpresentasikan dibandingkan dengan metode lain. Dan penelitian yang dilakukan oleh Neneng Sunengsih [14], tentang seleksi variabel dalam analisis regresi multivariat multipel, menerangkan bahwa seleksi variabel dalam analisis regresi multivariate multiple sebaiknya dilakukan secara simultan dengan alasan akan lebih cepat dalam proses perhitungan dan algoritma *forward selection* dapat dijadikan rujukan untuk mengoptimalkan variabel yang harus dimasukkan ke dalam model dan dalam penelitian tersebut metode *forward selection* menunjukkan performa yang lebih baik dibandingkan dengan metode *backward elimination*. Wardani, dkk [15] juga telah

melaporkan hasil penelitiannya, bahwa seleksi atribut akan meningkatkan akurasi suatu prediksi.

Pada penelitian ini bertujuan untuk mengembangkan model prediksi menggunakan algoritma KNN berbasis *forward selection* untuk memprediksi harga komoditi lada hitam dan lada putih. Peneliti juga membandingkan model yang dikembangkan dengan metode lain yaitu *Support Vector Machine*.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Pada penelitian yang dilakukan sebelumnya seperti yang telah dilakukan oleh Wuwung, dkk [16], bertujuan untuk mengembangkan model ARIMA yang mampu memprediksi harga beras dalam waktu jangka pendek yakni tiga bulan sebelumnya. Hal ini sebagai dasar yang dapat digunakan untuk memonitoring perubahan harga komoditi dari tiap bulannya. Pada penelitian tersebut dijelaskan bahwa dalam ilmu statistika, model ARIMA merupakan model *time series* yang dapat diterapkan untuk memprediksi data yang tidak *stasioner*. Pada penelitian ini, model ARIMA diterapkan untuk memprediksi harga beras sultan dan beras membramo. Berdasarkan data yang diolah model ARIMA cukup efisien untuk digunakan pada jumlah data yang sedikit akan tetapi kurang cocok untuk penggunaannya pada jumlah data yang banyak.

Berdasarkan penelitian yang dilakukan oleh Antti Sorjamaa dkk [17] mengusulkan metodologi global untuk prediksi jangka panjang *time series*. Dalam metodologi ini menggabungkan langsung strategi prediksi dengan masukan kriteria seleksi yaitu metode pendekatan k-terdekat (K-NN), *mutual information* (MI) dan *nonparametric noise estimation* (NNE). Pada penelitian yang dilakukan oleh Siska Ernida Wati dkk [18],

tentang perbandingan metode *fuzzy* dengan regresi *linier* berganda dalam peramalan jumlah produksi, pada penelitian ini menggunakan dua model yakni *fuzzy* dan *linier* regresi dalam meramalakan hasil produksi produksi kelapa sawit di PT. Perkebunan III Medan. Kedua metode tersebut memiliki masing-masing variabel bebas dan variabel terkait akan tetapi kedua metode tersebut memiliki tahap-tahap yang berbeda satu sama lain.

2.2 Seleksi Fitur

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 apakah 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 [19].

2.3 K- Nearest Neighbor

Algoritma KNN merupakan metode yang menggunakan algoritma *supervised*. Algoritma *supervised learning* tujuannya untuk mendapatkan pola baru sedangkan *unsupervised learning* tujuannya untuk mendapatkan pola dalam sebuah data [20].

KNN Regresi, merupakan algoritma yang memberikan pengenalan ke K-terdekat regresi tetangga yang dasar pendekatan *Unsupervised K*

Nearest Neighbor atau UNN, yang mana dalam regresi adalah untuk memprediksi nilai output. Ide KNN didasarkan pada asumsi lokalitas di ruang data. Dalam lingkungan lokal pola x diharapkan memiliki nilai output yang sama y (atau label kelas) untuk $f(x)$. Akibatnya, untuk x' diketahui label harus mirip dengan label dari pola terdekat, yang dimodelkan dengan rata-rata nilai output dari sampel terdekat K [21].

Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. Data training diproyeksikan ke ruang berdimensi banyak, yang mana masing-masing dimensi menjelaskan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data training. Sebuah titik pada ruang ini ditandai kelas c , jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut [22].

Dalam melakukan prediksi harga komoditi lada dengan menggunakan algoritma KNN digunakan model sebagai berikut:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (1)$$

Keterangan:

x = sampel data

y = data uji

D = Jarak

$$f_{knn}(x') = \frac{1}{K} \sum_{i \in N_k(x')} y_i \quad (2)$$

x' = Perkiraan atau estimasi

K = Jumlah tetangga terdekat

$N_k(x')$ = Tetangga terdekat

y_i = Output tetangga terdekat

2.4 Root Mean Square Error (RMSE)

Dalam melihat hasil suatu prediksi dilakukan suatu evaluasi dimana 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 mengakarkan 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 [6] [7]. Semakin kecil nilai yang dihasilkan semakin bagus pula hasil peramalan yang dilakukan. Rumus RMSE dipresentasikan pada persamaan (3):

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}} \quad (3)$$

y_t = Nilai aktual indeks

\hat{y}_t = Nilai prediksi indeks

n = Jumlah sampel

2.5 Data Time series

Data adalah segala fakta, angka, atau teks yang dapat diproses oleh komputer. Dimana saat ini, akumulasi perkembangan jumlah data berjalan dengan cepat dan pesat dalam format dan basis data yang berbeda. Data-data tersebut antara lain yaitu data oprasional atau penjualan, seperti penjualan, inventaris, penggajian, ankuntansi, dan sebagainya, data nonoprasional, seperti industri penjualan, peramalan, dan data ekonomi mikro, serta metadata yaitu data mengenai data itu sendiri,

seperti desain logika basis data atau defenisi kamus data[23].

Dalam perkembangan analisis data *time series*, banyak fenomena yang menarik dan sederhana merupakan fenomena yang non-linear (hubungan antara kejadian di masa lalu dan saat ini adalah non-linear). Sehingga pemodelan data *time series* yang linear tidaklah cukup dan sesuai untuk menangani kasus-kasus tersebut. Sebagai konsekuensinya, model-model *time series* non-linear telah menjadi perhatian utama para peneliti prediksi *time series* pada beberapa tahun terakhir ini. Beberapa bentuk model non-linear telah dikembangkan dan diaplikasikan pada beberapa kasus *time series*, dan sebagai *overview*[24].

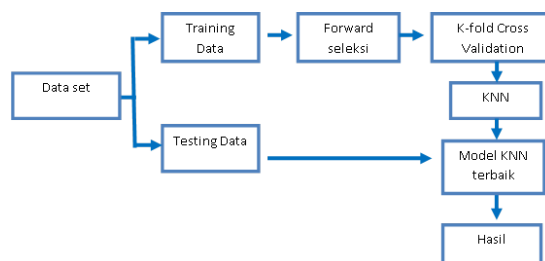
III. METODOLOGI

3.1 Pengumpulan Data

Pengumpulan data merupakan langkah awal pada suatu penelitian, dimana sumber data pada penelitian ini diambil dari *data set* BAPPEBTI atau Badan Pengawasan Perdagangan Berjangka Komoditi. Data yang diperoleh berupa data harian *time series* univariate. Data tersebut berupa data harga komoditi lada yaitu *data set* dari tahun 2010-2013. *Data set* tersebut untuk hari sabtu dan minggu tidak terjadi transaksi serta tanggal merah setiap tahunnya kosong dan terdiri dari variabel harga dengan 961 record untuk lada hitam (LH) dan 958 record untuk lada putih (LP).

3.2 Metode Yang Diusulkan

Tahapan ini akan membahas metode yang akan digunakan pada penelitian. Pada penelitian ini kami mengusulkan sebuah metode untuk menjawab pertanyaan penelitian, serta untuk mencapai tujuan penelitian.



Gambar 1 Metode yang diusulkan

Berdasarkan dari Gambar 1, maka dapat dijelaskan dari sistem kerja dari metode yang diusulkan yaitu dari *data set* yang diperoleh data akan dibagi menjadi dua buah *data set*, yaitu data *training* dan data *testing*. Setelah melakukan pembagian data maka langkah selanjutnya adalah mengubah data training dari data univariat menjadi data multivariat seperti data *training* dijadikan dalam beberapa variabel *independent*, seperti 1 periode sampai 10 periode dengan masing-masing 1 variabel *dependent*. Contoh data 1 periode seperti $x_{t-1} : x_t$, 2 variabel *input* seperti $x_{t-2}, x_{t-1} : x_t$ dan data 10 periode seperti $x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1} : x_t$ begitu pula pada data *testing*. Data *training* diubah menjadi beberapa periode guna mendapatkan model yang tepat untuk melakukan prediksi harga komoditi lada. Selanjutnya data *training* akan diseleksi dengan menggunakan fitur seleksi yaitu *forward selection* guna mempercepat proses dan memilih variabel yang sesuai. Hasil data seleksi akan dilanjutkan ke proses berikutnya yaitu masuk dalam proses algoritma KNN, dalam proses ini dilakukan percobaan-percobaan dengan mengganti-ganti nilai ketetanggan atau nilai *k* pada setiap percobaan data training mulai dari 1 sampai pada 10, hal ini diperlukan untuk mendapatkan model yang baik sehingga diperoleh nilai RMSE yang lebih kecil. Setelah diperoleh model atau formula yang tepat maka data *testing* akan diproses dan didapatlah nilai prediksi.

3.3 Eksperimen Dan Pengujian Metode

Pada tahapan eksperiment dan pengujian metode dilakukan beberapa tahapan yakni:

3.3.1 Tahap Pengolahan Data

Data merupakan segala fakta, angka, teks dan gambar yang dapat diproses oleh komputer. Dalam tahap ini data yang diolah merupakan data univariat *time series* yang berupa numerik. Dilakukan pengolahan data dengan mengubah data tersebut dari data univariat menjadi multivariat, yaitu mengubah data mulai dari 1 periode sampai 10 periode. *Time series multivariate* memodelkan peubah-peubah yang berkorelasi dan tercatat dari waktu ke waktu.

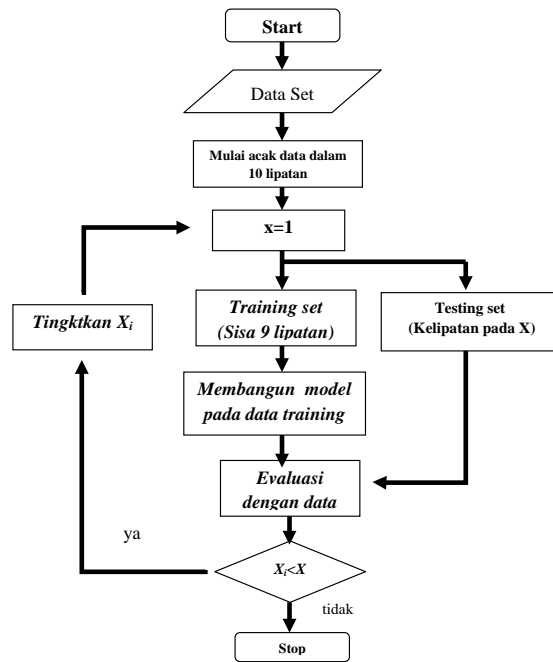
Pola data yang digunakan dalam Tabel 1 mengubah data univariat menjadi multivariat, ditunjukkan pada pola atau modelnya sebagai berikut

Tabel 1. Pola data *time series* (univariate ke multivariate) [7]

| Pattern | Input lag | Output/ Target |
|---------|-------------------------------------|----------------|
| 1 | $x1, x2, x3, x4, \dots, xp$ | $xp+1$ |
| 2 | $x2, x3, x4, x5, \dots, xp+1$ | $xp+2$ |
| 3 | $x3, x4, x5, x6, \dots, xp+2$ | $xp+3$ |
| ... | ... | |
| $m-p$ | $xm-p, xm-p+1, xm-p+2, \dots, xm-1$ | xm |

3.3.2 Tahap Seleksi Data

Dalam tahap ini digunakan *forward selection* yaitu memilih variabel yang benar-benar memberikan informasi dalam keakuratan prediksi agar performa suatu algoritma dapat dioptimalkan. *Forward selection* prinsip kerjanya yaitu pendekatan ini membangun model dimulai dengan tidak ada variabel dalam model dan menambahkan variabel yang berguna satu per satu. Skema umum adalah sebagai berikut:



Gambar 2. Flowchart forward selection

Data yang di-*training* dilakukan secara bertahap yakni dimulai dari 1 variabel sampai pada tingkat atau jumlah variabel yang menghasilkan performa atau nilai akurasi paling baik atau *error* terkecil. Misalnya pada pengujian data dengan 2 variabel menghasilkan *error* lebih kecil dan ketika diujikan lagi dengan 3 variabel dan menghasilkan nilai *error* lebih besar dibandingkan dengan 2 variabel maka *error* terkecil didapatkan pada variabel ke 2 yang berarti variabel kedua signifikan, proses dihentikan bila semua variabel *independen* sudah diujikan. Algoritma *forward selection* akan diujikan pada setiap data, mulai dari data 1 variabel periode sampai pada data 10 variabel periode untuk dibandingkan data mana yang menghasilkan akurasi yang paling baik.

3.3.3 Tahap Pengujian Algoritma

Dalam tahap pengujian algoritma yang akan digunakan adalah algoritma KNN. Algoritma KNN merupakan metode yang menggunakan algoritma supervised dimana algoritma tersebut terbagi atas dua jenis yaitu *supervised learning* dengan

unsupervised learning. Algoritma *supervised learning* tujuannya untuk mendapatkan pola baru sedangkan *unsupervised learning* tujuannya untuk mendapatkan pola dalam sebuah data. Pada pengujianya untuk setiap data yang diuji agar mendapatkan hasil yang maksimal dalam setiap pengujian data, nilai ketetanggan atau K dari algoritma ini dilakukan perubahan sampai mendapatkan nilai yang paling baik. Sehingga diketahui data yang mana dan nilai ketetanggan berapa yang menghasilkan nilai yang paling baik.

Selain algoritma KNN peneliti juga akan menggunakan algoritma lain seperti algoritma *Support Vector Machine* (SVM) sebagai perbandingan algoritma yang memiliki performa yang paling baik dalam memprediksi harga lada. Model SVM yang digunakan dengan tipe kernel polynomial.

3.4. Evaluasi

Tahapan evaluasi *data set* yaitu data testing dievaluasi dengan mencari nilai *Root Mean Square Error* (RMSE) terkecil. RMSE yaitu dilakukan dalam melihat hasil suatu peramalan dilakukan suatu evaluasi dimana evaluasi tersebut digunakan untuk mengetahui keakuratan hasil peramalan yang telah dilakukan terhadap data yang sebenarnya.

IV. HASIL DAN PEMBAHASAN

4.1 Hasil Eksperimen

Peneliti melakukan eksperimen beberapa tahap. Tahap pertama pengolahan data yaitu mengubah data harian menjadi data mingguan. Tahap kedua dari merubah data *time series* univariat menjadi multivariat. Tahap ketiga menentukan parameter k pada KNN dan KNN berbasis *forward selection* dan *backward elimination* serta membandingkan SVM berbasis *forward selection* dan *backward elimination*.

Dalam menentukan parameter k pada KNN dilakukan beberapa pengujian yaitu penentuan jumlah variabel independent atau variabel periode x_{t-i} . Pengujian ini dilakukan untuk mendapatkan model yang terbaik dengan melihat nilai *root meansquare error* terkecil. Tahap keempat menentukan parameter SVM. Dalam menentukan parameter SVM hampir sama tahap kedua yaitu dilakukan beberapa pengujian yaitu penentuan jumlah variabel input untuk mendapatkan model yang terbaik, dengan melihat nilai *root meansquare error* terkecil. Tahap kelima yaitu membandingkan model yang terbaik menggunakan algoritma KNN dan SVM berbasis seleksi atribut menggunakan *forward selection* dan *backward elimination*.

Pada proses eksperimen, digunakan *number of validation* 10 (*10-fold validation*), variabel input dari 1 sampai 10 dan nilai ketetanggan pada KNN, dari $k=1$ sampai dengan 10. Kemudian dilakukan proses testing untuk mendapatkan *root mean square error*.

A. *K-Nearest Neighbor* (KNN)

Dengan menggunakan metode KNN, dihasilkan kinerja prediksi seperti ditunjukkan pada Tabel 2 berikut ini.

Tabel 2. Nilai RMSE menggunakan KNN untuk Lada Hitam

| Nilai variabel periode | Number of validation | k | RMSE |
|------------------------|----------------------|-----|----------|
| 1 | 10 | 5 | 1722,406 |
| 2 | 10 | 4 | 1733,835 |
| 3 | 10 | 4 | 1777,334 |
| 4 | 10 | 5 | 1828,186 |
| 5 | 10 | 4 | 1764,610 |
| 6 | 10 | 3 | 1749,059 |
| 7 | 10 | 2 | 1778,979 |
| 8 | 10 | 3 | 1669,635 |
| 9 | 10 | 4 | 1790,217 |
| 10 | 10 | 3 | 1915,523 |

Berdasarkan pada Tabel 2, yang merupakan tabel rangkuman hasil eksperimen untuk data *time series* lada hitam. Perlu dicatat bahwa Tabel 2 merupakan nilai *k* pada KNN yang memiliki RMSE terbaik (dipilih berdasarkan nilai *root mean square error* (RMSE) terkecil). dari nilai *k* dimulai dari 1 sampai 10. Maka didapatkan model terbaik dengan RMSE terkecil 1669,635, dengan jumlah variabel input sebanyak 8 dan nilai *k* pada KNN sama dengan 3.

Untuk data *time series* lada putih, dari eksperimen dihasilkan nilai RMSE seperti ditunjukkan Tabel 3.

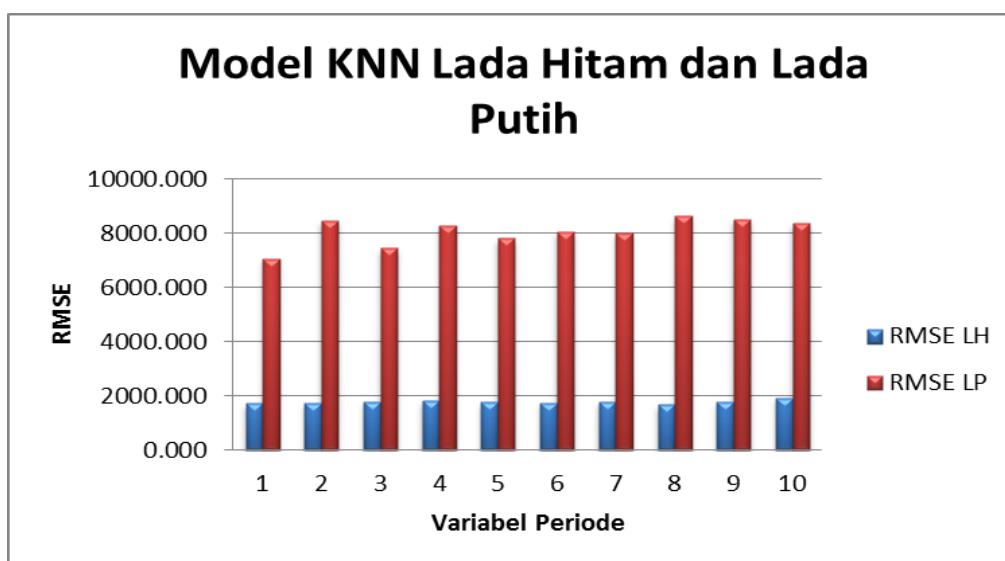
Tabel 3. Nilai RMSE pada KNN untuk Lada Putih

| Nilai variabel periode | Number of validation | <i>k</i> | RMSE |
|------------------------|----------------------|----------|-----------------|
| 1 | 10 | 1 | 7062,539 |
| 2 | 10 | 10 | 8472,747 |
| 3 | 10 | 1 | 7434,537 |
| 4 | 10 | 9 | 8287,635 |
| 5 | 10 | 10 | 7803,198 |
| 6 | 10 | 10 | 8049,394 |
| 7 | 10 | 10 | 7998,275 |
| 8 | 10 | 10 | 8614,896 |

| | | | |
|----|----|----|----------|
| 9 | 10 | 10 | 8510,118 |
| 10 | 10 | 10 | 8378,711 |

Berdasarkan pada Tabel 3, yang merupakan tabel rangkuman hasil eksperimen lada putih dari penentuan model dengan menggunakan algoritma KNN yang dimulai dari 1 sampai 10 variabel input dengan menggunakan *10 fold validation* dan nilai *k* pada KNN dari 1 sampai dengan 10 yang dipilih berdasarkan nilai *root mean square error* (RMSE) terkecil, maka didapatkan model terbaik yang ditunjukkan berdasarkan nilai RMSE terkecil 7062,539, yaitu 1-NN dengan jumlah variabel input sebanyak 1.

Berdasarkan Gambar 3 yang merupakan grafik perbandingan model terbaik untuk data lada Hitam dan lada putih, menunjukkan bahwa model yang paling baik terdapat pada variabel input sebanyak 8 pada 3-NN untuk data lada hitam, dengan nilai RMSE-nya sebesar 1669,635. Sedangkan untuk data lada putih, model yang paling baik terdapat pada variabel input sebanyak 1 pada 1-NN dengan nilai RMSE-nya sebesar 7062,539.



Gambar 3 Grafik Kinerja Model KNN untuk Lada Hitam dan Lada Putih

A. *K-NEAREST NEIGHBOR* BERBASIS *FORWARD SELECTION*

Dari hasil percobaan yang dilakukan mulai dari variabel input sebesar 1 sampai dengan 10 maka dapat ditentukan model yang akan digunakan untuk melakukan prediksi harga lada hitam dan lada putih.

Tabel 4. RMSE Nilai Model Lada Hitam dengan KNN berbasis *Forward selection*

| Nilai variabel periode | Number of validation | K | RMSE Forward selection |
|------------------------|----------------------|----------|------------------------|
| 1 | 10 | 5 | 1722,406 |
| 2 | 10 | 6 | 1682,326 |
| 3 | 10 | 5 | 1710,424 |
| 4 | 10 | 3 | 1611,784 |
| 5 | 10 | 3 | 1559,741 |
| 6 | 10 | 3 | 1634,526 |
| 7 | 10 | 4 | 1563,856 |
| 8 | 10 | 3 | 1565,487 |
| 9 | 10 | 4 | 1610,171 |
| 10 | 10 | 4 | 1605,271 |

Berdasarkan pada Tabel 4, yang merupakan tabel rangkuman hasil eksperimen dari penentuan model dengan menggunakan algoritma KNN berbasis *Forward selection* yang dimulai dari 1 sampai 10 variabel input untuk data lada hitam dengan menggunakan *cross validation* 10 dan nilai *k* pada KNN dari 1 sampai dengan 10 yang dipilih berdasarkan nilai *root mean square error* (RMSE) terkecil. Maka didapatkan model terbaik yang ditunjukkan berdasarkan nilai *error* terkecil 1559,741 dengan jumlah variabel input sebesar 5,

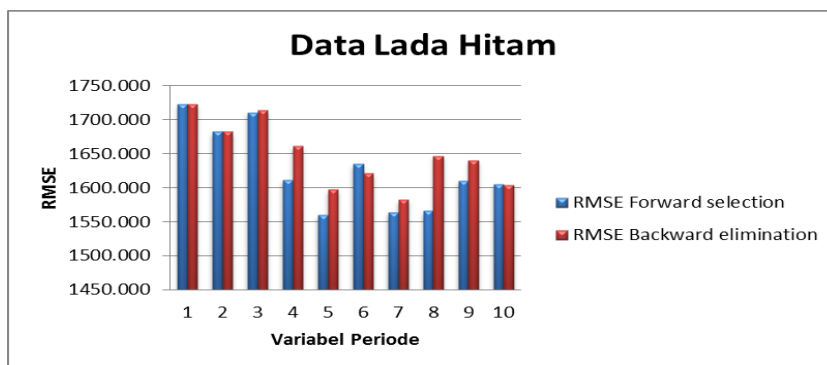
pada 3-NN dengan variabel yang terpilih yaitu x_{t-1} dan x_{t-4} .

Apabila digunakan seleksi variabel *Backward Elimination*, hasil kinerja model ditunjukkan pada Tabel 5 berikut:

Tabel 5. RMSE Nilai Model Lada Hitam dengan KNN berbasis *Backward Elimination*

| Nilai variable periode | Number of validation | K | RMSE Bacward elimination |
|------------------------|----------------------|----------|--------------------------|
| 1 | 10 | 5 | 1722,406 |
| 2 | 10 | 6 | 1682,326 |
| 3 | 10 | 4 | 1713,734 |
| 4 | 10 | 3 | 1660,789 |
| 5 | 10 | 3 | 1596,787 |
| 6 | 10 | 3 | 1620,848 |
| 7 | 10 | 4 | 1582,273 |
| 8 | 10 | 3 | 1646,423 |
| 9 | 10 | 4 | 1640,445 |
| 10 | 10 | 4 | 1604,088 |

Berdasarkan pada Tabel 5, yang merupakan tabel rangkuman hasil eksperimen dari penentuan model dengan menggunakan algoritma KNN berbasis *Backward elimination* yang dimulai dari 1 sampai 10 variabel input dan nilai *k* pada KNN dari 1 sampai dengan 10 yang dipilih berdasarkan nilai *root mean square error* (RMSE) terkecil, maka didapatkan model terbaik yang ditunjukkan berdasarkan nilai *error* terkecil senilai 1582,273 dengan jumlah variabel input sebesar 7, pada 4-NN dan variabel input yang dipilih yaitu x_{t-1} , x_{t-2} , x_{t-4} dan x_{t-7} .



Gambar 4. Grafik Model KNN Lada Hitam

Berdasarkan Gambar 4, yang merupakan grafik perbandingan model terbaik untuk data lada hitam, antara KNN berbasis *Forward selection* dengan KNN berbasis *Backward elimination* menunjukkan bahwa model yang paling baik terdapat pada variabel input sebanyak 5 pada 3-NN dengan menggunakan *forward selection* yaitu nilai RMSE-nya sebesar 1559,741.

Untuk data *time series* lada putih, dengan menggunakan KNN berbasis *forward selection* dihasilkan nilai RMSE seperti ditunjukkan Tabel 5

Tabel 6. Nilai RMSE dengan KNN berbasis *Forward selection* untuk Lada Putih

| Nilai variabel periode | Number of validation | K | RMSE Forward selection |
|------------------------|----------------------|----------|------------------------|
| 1 | 10 | 1 | 7062,539 |
| 2 | 10 | 1 | 6861,756 |
| 3 | 10 | 1 | 7203,587 |
| 4 | 10 | 1 | 7133,435 |
| 5 | 10 | 1 | 6879,243 |
| 6 | 10 | 1 | 7190,222 |
| 7 | 10 | 1 | 6416,486 |
| 8 | 10 | 1 | 6632,819 |
| 9 | 10 | 1 | 6364,136 |
| 10 | 10 | 1 | 6328,376 |

Berdasarkan pada Tabel 6, didapatkan model terbaik yang ditunjukkan berdasarkan nilai *error* terkecil senilai 6328,376 dengan jumlah variabel input sebesar 10 pada 1-NN dan variabel yang dipilih yaitu x_{t-1} , x_{t-4} , x_{t-6} dan x_{t-10} .

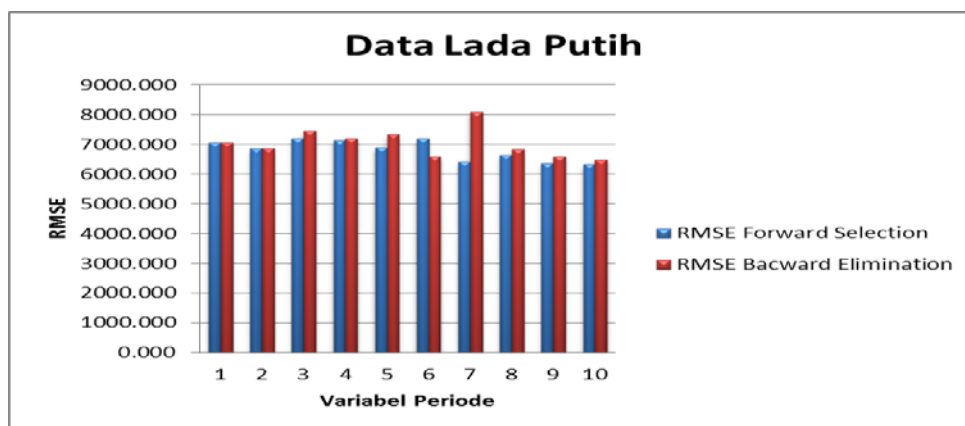
Apabila digunakan seleksi variabel *Backward Elimination*, diperoleh kinerja model sebagai berikut:

Tabel 7. Nilai RMSE dengan KNN berbasis *Backward Elimination* untuk Lada Putih

| Nilai variabel periode | Number of validation | K | RMSE Bacward elimination |
|------------------------|----------------------|----------|--------------------------|
| 1 | 10 | 1 | 7062,539 |
| 2 | 10 | 1 | 6861,756 |
| 3 | 10 | 1 | 7434,537 |
| 4 | 10 | 1 | 7200,468 |
| 5 | 10 | 1 | 7334,363 |
| 6 | 10 | 1 | 6574,156 |
| 7 | 10 | 9 | 8077,722 |
| 8 | 10 | 1 | 6828,468 |
| 9 | 10 | 1 | 6588,428 |
| 10 | 10 | 1 | 6486,606 |

Berdasarkan pada Tabel 7, didapatkan model terbaik yang ditunjukkan berdasarkan nilai *error* terkecil senilai 6486,606 dengan jumlah input sebanyak 10 pada 1-NN dan variabel yang dipilih yaitu x_{t-1} , x_{t-2} , x_{t-3} , x_{t-4} , x_{t-8} , x_{t-9} dan x_{t-10} .

Berdasarkan Gambar 5, yang merupakan grafik perbandingan model terbaik untuk data lada putih, antara KNN berbasis *Forward selection* dengan KNN berbasis *Backward elimination* menunjukkan bahwa model KNN berbasis *Forward selection* lebih baik daripada KNN berbasis *Backward elimination*



Gambar 5. Grafik Model KNN Lada Putih

B. PERBANDINGAN MODEL

Peneliti juga melakukan eksperimen dengan menggunakan model SVM berbasis *forward selection* dan *Backward Elimination*. Hasil perbandingan RMSE dari model SVM berbasis seleksi atribut ditunjukkan pada gambar 5.

Berdasarkan pada gambar 6, yang merupakan grafik perbandingan model SVM berbasis *Forward selection* dengan SVM berbasis *Backward elimination* menunjukkan bahwa model yang paling baik terdapat pada variabel input sebesar 10 dan *type kernel polynomial* serta seleksi atribut menggunakan *backward elimination* yaitu nilai RMSE-nya sebesar 11880,122.

Sedangkan untuk data *time series* lada putih, model SVM berbasis *Forward selection* adalah model terbaik dengan 10 variabel input dan *type kernel polynomial* dengan nilai RMSE-nya sebesar 19997,880.

Hasil perbandingan kinerja prediksi (RMSE) menggunakan algoritma KNN dan SVM ditunjukkan pada Tabel 8.

Berdasarkan tabel 8, dapat dilihat model yang dapat digunakan untuk melakukan prediksi harga komoditi lada hitam dan lada putih yaitu model pada algoritma KNN berbasis *forward selection* karena memiliki RMSE terkecil.

Tabel 8. Hasil perbandingan algoritma KNN, KNN+ forward , KNN+ backward, SVM + forward dan SVM + backward

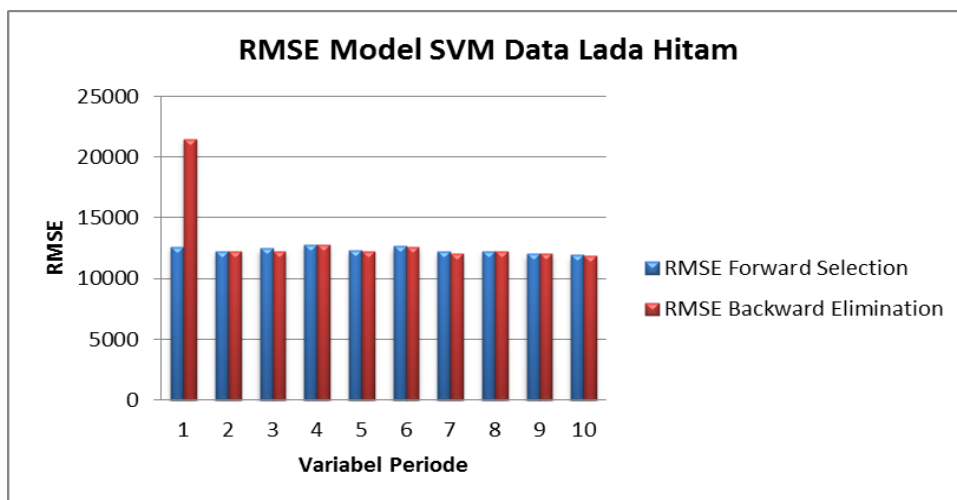
| Model | Jumlah variabel input | | RMSE Lada Hitam | RMSE Lada putih |
|--------------------|-----------------------|----------|-----------------|-----------------|
| | LH | LP | | |
| KNN | | | 1669,635 | 7062,539 |
| KNN Forward | 8 | 1 | 1559,741 | 6328,376 |
| KNN Bacward | 5 | 10 | 11920,611 | 19997,880 |
| SVM Forward | 7 | 10 | 11880,122 | 20140,984 |
| SVM Backward | 10 | 10 | | |
| | 10 | 8 | | |

V. KESIMPULAN

Dari hasil eksperiment yang dilakukan prediksi harga komoditi lada dengan menggunakan algoritma KNN berbasis *forward selection* berhasil dilakukan.

Berdasarkan hasil eksperiment yang telah dilakukan peneliti, menunjukkan bahwa fitur seleksi yakni *forward selection* merupakan model yang lebih baik dalam melakukan seleksi variabel yang signifikan dibandingkan dengan bacward elimination.

Algoritma KNN berbasis *forward selection* telah menunjukkan performa yang lebih baik dibandingkan algoritma KNN, KNN berbasis *backward elimination* dan SVM.



Gambar 6. Grafik RMSE Model SVM Lada Hitam

REFERENSI

- [1] Sugiharto, *Pendekatan informasi sebagai komoditi Dalam proses diseminasi informasi*, Subid. Jasa Kemasan Informasi, 2011.
- [2] N. I. Wulandari, *Penentuan agribisnis unggulan komoditi Pertanian berdasarkan nilai produksi di Kabupaten Grobogan*. Universitas Diponegoro Semarang, 2010.
- [3] M. M. Mustadjab and H. Tiawon, Ketahanan pangan utama Indonesia pada Era liberalisasi perdagangan, ISSN: 0853-5167, *HABITAT*, Vol. XX (2), 2009.
- [4] Rosihan Asmara, Nesia Artdiyasa, Analisis Tingkat Daya Saing Ekspor Komoditi Perkebunan Indonesia, *AGRISE*, Vol. VIII (2), 2008
- [5] A. Sularno, *Prediksi Nilai Saham Menggunakan Pemrograman Genetik Dan Pemrograman Ekspresi Gen*, Universitas Gunadarma Depok. Indonesia, 2014.
- [6] 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
- [7] 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
- [8] B. A. Jain and B.N. Nag, Performance Evaluation of Neural Network Decision Models. *Manage Information Systems* 14, 201–216, 1997
- [9] H. Niskaa, T. Hiltunena, A. Karppinenb, J. Ruuskanena, and M. Kolehmaine, Evolving the Neural Network Model for Forecasting Air Pollution Time series. *Engineering Applications of Artificial Intelligence* 17, 159–167, 2004
- [10] S. Georgakarakos, D. Koutsoubas, V. Valavanis, Time series Analysis and Forecasting Techniques Applied on Loliginid and Ommastrephid Landings in Greek Waters, *Fisheries Research* 78, 55–71, 2006
- [11] F. Aminian, E. D. Suarez, M. Aminian, D.T. Walz, Forecasting Economic Data with Neural Networks, *Computational Economics* 28, 71–88, 2006.
- [12] A. Candra, A. Syaifudin, B. Nurjayanti and A. Putriani, *Perbandingan algoritme bayesian network Dan k-nearest neighbor (KNN) Pada prediksi penyakit paru manusia*, Institut Pertanian Bogor, 2010.
- [13] M. I. Sikki, Pengenalan wajah menggunakan k-nearest neighbour dengan praproses transformasi wavelet, *Jurnal Paradigma*, Vol X. No. 2 Desember 2009.
- [14] N. Sunengsih, Seleksi Variabel Dalam Analisis Regresi Multivariat Multipel, Staf Jurusan Statistika FMIPA UNPAD, *Seminar Nasional Matematika dan Pendidikan Matematika*, 5 Desember 2009
- [15] R. S. Wardani, Purwanto, Model Pengambilan Keputusan dalam Prediksi Kasus Tuberkulosis Menggunakan Regresi Logistik Berbasis Backward elimination, *Prosiding Seminar Nasional Hasil-hasil Penelitian dan Pengabdian*, 2014
- [16] V. Wuwunga, N. Nainggolana, and M. Paendonga, *Prediksi Harga Beras Sultan dan Membramo di Kota Manado dengan Menggunakan Model ARIMA*, FMIPA UNSRAT, 2013.
- [17] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse, *Methodology for long-term prediction of time series*, Helsinki University of Technology, 21 May 2007
- [18] S. E. Wati, D. Sebayang, R. Sitepu, Perbandingan metode fuzzy dengan Regresi linier berganda dalam Peramalan jumlah produksi, *Mathematics Subject Classification*, 2013.
- [19] 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.
- [20] N. Krisandi, Helmi, and B. Prihandono, Algoritma k-nearest neighbor dalam klasifikasi data hasil produksi kelapa sawit pada pt. Minamas Kecamatan Parindu, *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster*, Volume 02, No.1, 2013.
- [21] O. Kramer, Unsupervised K-Nearest Neighbor Regression, Carl von Ossietzky Universit'at Oldenburg 26111 Oldenburg, Germany arXiv: 1107.3600v2 [stat.ML] 26 Sep 2011
- [22] R. N. Whidhiasih, N. A. Wahanani and Supriyanto, Klasifikasi buah belimbing berdasarkan citra red-green-blue menggunakan knn dan lda, *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic* 1(1) : 29-35, 2013.
- [23] E. Prasetyo, *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*, ANDI Jl. Beo 38-40 Yogyakarta, 2012.
- [24] Suhartono, *Feedforward Neural Networks Untuk Pemodelan Runtun Waktu*, Universitas Gadjah Mada Yogyakarta, 25 September 2007.