

PERBANDINGAN METODE PREDIKSI SUPPORT VECTOR MACHINE DAN LINEAR REGRESSION MENGGUNAKAN BACKWARD ELIMINATION PADA PRODUKSI MINYAK KELAPA

Andi Bode

Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo

Jl. Achmad Nadjamuddin No. 17 Telp. 0435-829975 Fax. 0435-829976 Kota Gorontalo
andibode22@gmail.com

Pohon kelapa banyak dimanfaatkan oleh manusia, sehingga tumbuhan ini dianggap tumbuhan serbaguna, salah satunya minyak kelapa yang dihasilkan oleh buah pohon kelapa. Produksi jumlah minyak kelapa menjadi bagian penting disetiap perusahaan yang bergerak di bidang produksi dengan tujuan mencapai target hasil produksi. Namaun Produksi minyak setiap hari mengalami perubahan fluktuatif. Perusahaan sangat memerlukan prediksi jumlah produksi. Penelitian ini bermaksud membandingkan metode support vector machine dan linear regression menggunakan fitur seleksi backward elimination berdasarkan data time series Sales Order. Hasil penelitian pada dataset sales order dengan menggunakan metode Support Vector Machine (SVM) didapatkan RMSE 0,127, dengan menggunakan metode SVM dan Backward Elimination (BE) didapatkan RMSE 0,115, dengan metode Linear Regression (LR) didapatkan RMSE 0,118 dan dengan menggunakan metode LR dan Backward Elimination didapatkan RMSE 0,118. Dari hasil perbandingan tersebut dapat disimpulkan bahwa kinerja SVM menggunakan Backward Elimination lebih baik dibanding SVM, LR dan LR menggunakan Backward Elimination.

Kata Kunci: *Produksi Minyak Kelapa, Support Vector Machine, Linear Regression, Backward Elimination.*

I. PENDAHULUAN

Salah satu tumbuhan serba guna bagi masyarakat yaitu pohon kelapa, pohon ini memiliki banyak manfaat hampir keseluruhan bagian dari pohon ini digunakan, sehingga tumbuhan ini merupakan tumbuhan serbaguna terutama bagi penduduk dibagian pesisir. Salah satu hasil dari pohon kelapa yang dimanfaatkan oleh masyarakat adalah minyak kelapa yang dihasilkan dari buah pohon kelapa tersebut. Minyak kelapa merupakan salah satu kebutuhan pokok manusia dan untuk memenuhi permintaan kebutuhan telah banyak perusahaan bersaing dalam mengelola buah kelapa untuk dijadikan produk pakai

yaitu minyak kelapa. Dari hasil permintaan pasar yang tidak bisa di perkirakan maka terjadi permasalahan dalam meramalkan produksi minyak pada perusahaan. Produksi minyak kelapa tiap hari mengalami perubahan atau fluktuatif. Metode yang lajim digunakan didalam prediksi adalah metode *time series*, dimana untuk mengetahui nilai dimasa mendatang dapat dilihat berdasar urutan waktu pengamatan. Untuk performa atau kemampuan dari metode *time series* sangat unggul dengan cara melihat nilai-nilai di masa lalu [1].

Support Vector Machine dan *Linear Regression* merupakan metode yang sering digunakan dalam *forecasting* (prediksi) dengan tingkat keakuratan yang baik. Telah banyak peneliti-peneliti menggunakan metode *forecasting* tersebut. *Support Vector Machine* metode yang berlandaskan pada teori pembelajaran *statistic* dan memberi hasil yang menjanjikan lebih baik dibanding metode lain. *Support Vector Machine* bekerja juga dengan baik terhadap data yang berdimensi tinggi dengan menggunakan teknik *kernel* [2]. Begitu juga dengan Metode *Liner Regression* mampu menghasilkan koefisien korelasi yang sangat baik untuk prediksi [3]. Pada penulisan penelitian kali ini akan terapkan kedua metode tersebut pada bidang *forecasting* dengan permasalahan yang diangkat adalah jumlah produksi yang fluktuatif serta untuk dapat mengetahui keoptimalan dua metode tersebut dalam memprediksi data *sales order*.

II. METODE PENELITIAN

2.1. Preprocessing Data

Pada tahapan preprocessing data univariat *time series* yang berupa numerik data harian, data diubah menjadi data mingguan, untuk mempermudah proses pengelolannya, data di ubah ke *ascending descending*, setelah itu data dinormalisasi. Proses normalisasi dilakukan untuk meminimalisir terjadinya belbagai anomaly dan ketidak konsistensinya data. Namun tujuan utama normalisasi yaitu untuk menghasilkan data yang

berdimensi kecil dan tetap dapat mewakili data asli tanpa kehilangan karakteristik data [2].

Persamaan normalisasi sebagai berikut:

$$N = \frac{(X - \min)}{(\max - \min)} \quad (1)$$

Dimana:

- N = Normalisasi
- x = Data
- Min = Data Minimum
- Max = Data Maksimum

Persamaan denormalisasi dapat dilihat sebagai berikut:

$$D = Y(\max - \min) + \min \quad (2)$$

Dimana:

- D = Denormalisasi
- Y = Hasil Keluaran Dari Pelatihan
- Min = Data Minimum
- Max = Data Maksimum

Sumber data set pada penelitian ini di ambil PT. Multi Nabati Sulawesi Unit Maleo Kabupaten Pohuwato. Data yang dikumpulkan adalah data kuantitatif harian *time series* univariat. Jenis data ini adalah data SO (Sales Order). Dataset dari yang diambil dari variabel sales order memiliki 208 *record* yang berasal dari setiap pesanan *costumer*.

2.2. Data Mining

Data mining yaitu sebuah proses menjalankan satu atau lebih dari satu teknik *machine learning* yang bertujuan menganalisis dan mengekstraksi *knowledge* secara otomatis [4]. Dalam pengertian lain yaitu *induction-based learning* dimana proses pembentukan definisi-definisi konsep umum yang dikerjakan dengan mengobservasi contoh spesifik dari konsep yang dipelajari. *Knowledge Discovery in Databases* yaitu penerapan metode saintifik yang ada pada data mining. Dalam sebuah konteks tersebut, data mining adalah langkah dari proses *Knowledge Discovery In Database* [5].

2.3. Fitur Seleksi

Metode *Backward Elimination* merupakan metode yang memiliki tujuan untuk mengoptimalkan performa suatu model sistem kerja dengan cara pemilihan mundur. Pemilihan sebuah variabel dilakukan dengan system kerja pemilihan kedepan yaitu menguji keseluruhan variabel dan menghapus variabel yang dianggap tidak relevan. Variabel yang telah diproses satu per satu, jika didapati sebuah variabel yang dianggap tidak berpengaruh, tidak relevan atau tidak signifikan didalam model, maka variabel tersebut akan dihapus dari dalam model [6].

2.4. Support Vector Machine

Support Vector Machines (SVM) adalah algoritma *supervised learning* yang sangat baik yang digunakan

untuk klasifikasi atau regresi, algoritma SVM dapat juga digunakan untuk prediksi numerik serta klasifikasi [7]. SVM adalah algoritma yang memiliki fungsi untuk pemetaan nonlinier dan mengubah data uji asli ke dimensi yang lebih tinggi. Dengan pemetaan nonlinier yang sama ke dalam dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan oleh *hyperplane*. SVM menemukan *hyperplane* tersebut dengan menggunakan vektor dukungan dan margin. Support Vector Machine (SVM) secara konseptual adalah mesin linier yang dilengkapi dengan fitur spesial, dan didasarkan pada metode minimalisasi resiko struktural, serta teori pembelajaran statistik. SVM memiliki dua sifat khusus yaitu mencapai generalisasi yang tinggi dengan memaksimalkan margin, dan mendukung pembelajaran yang efisien dari fungsi nonlinier pada trik kernel sehingga membuat kinerja generalisasinya baik dalam menyelesaikan masalah pengenalan pola. SVM memiliki properti yang sama tentang memaksimalkan margin dan trik kernel untuk memetakan data yang nonlinier. Secara singkat sekumpulan data *training* untuk regresi SVM digambarkan sebagai dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ Dimana x_j adalah vector *n-dimensional*, sedangkan y adalah jumlah real untuk setiap x_j . Tugas dari regresi adalah untuk menemukan fungsi antara x_j dengan y_j yang dalam kasus linier [3].

$$y = f(x) = w \cdot x + b \quad (3)$$

2.5. Linear Regression

Linear Regression yaitu pemodelan dan analisis data numerik yang terdiri dari satu atau lebih variabel independen dan nilai variabel dependen. Tujuan menggunakan analisis regresi adalah untuk membuat estimasi nilai variabel dependen dan yang didasarkan pada nilai variabel independen. Model regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antar variabel. Regresi adalah model untuk memprediksi nilai dari masukan data yang diberikan. Metode utama untuk melakukan prediksi dengan membentuk model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor X dengan variabel dependen atau respon Y [3].

Model Linear Regression dinyatakan sebagai berikut:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (4)$$

Dengan :

- Y = adalah variabel dependen
- a = adalah *regression costanta* atau *intercept*
- b = adalah *regression coefficients*
- n = adalah banyaknya variabel independen X_1, X_2, \dots, X_n adalah variabel independen

2.6. Root Mean Square Error (RMSE)

Untuk mendapatkan hasil keakuratan dari sebuah peramalan, diperlukan evaluasi terhadap data yang sebenarnya. Ada beberapa metode yang sering digunakan untuk melakukan perhitungan kesalahan peramalan salah

satunya adalah *root mean square error* (RMSE), metode ini sering digunakan dalam mengevaluasi hasil peramalan. Ukuran yang sering digunakan dari perbedaan antara nilai-nilai diprediksi oleh model atau estimator dan nilai-nilai benar-benar diamati yaitu RMSE. Perbedaan-perbedaan individual disebut residual ketika perhitungan dilakukan atas sampel data yang digunakan untuk estimasi, dan disebut kesalahan prediksi ketika dihitung *out-of-sample*. RMSE memiliki fungsi untuk agregat besaran kesalahan dalam prediksi untuk berbagai kali menjadi ukuran tunggal daya prediksi. RMSE merupakan ukuran akurasi yang baik, tapi hanya untuk membandingkan kesalahan peramalan model yang berbeda untuk variabel tertentu dan tidak antara variabel, karena skala-dependent [8]. MenurutRaharja, A [9], berikut persamaan *root means square error* (RMSE).

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}} \tag{5}$$

Dimana

- RMSE** = *Root Mean Square Error*
- n** = Jumlah Sampel
- y_t** = Nilai Aktual Indeks
- y_t̂** = Nilai Prediksi Indeks

2.7. *Data Time Series*

Data *time series* adalah nilai suatu variabel yang dikumpulkan atau dicatat berdasarkan urutan/deret waktu dalam bentuk satuan harian, mingguan, bulanan, atau tahunan dan di beberapa kasus terdapat jam tergantung keperluan terhadap perilaku data yang akan diobservasi. Data *time series* yang digunakan adalah data historis yang diukur berdasarkan suatu pengamatan tertentu. Karena data yang digunakan adalah data yang terukur, maka peramalan secara *time series*, termasuk ke dalam peramalan kuantitatif. Metode prediksi *time series* beranggapan bahwa data atau kejadian masa lalu akan cenderung berulang dimasa yang akan datang. Fokus prediksi secara *time series* adalah apa yang akan terjadi, bukan mengapa hal itu terjadi [10].

III. HASIL DAN PEMBAHASAN

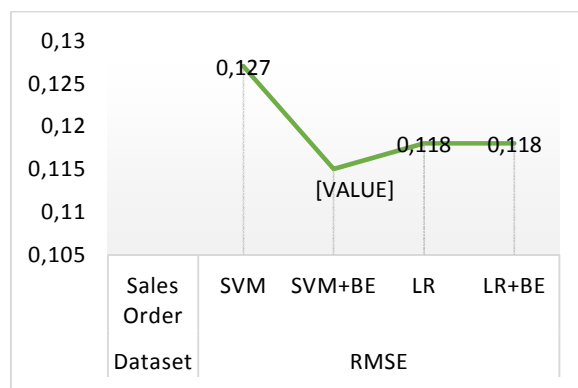
Penelitian ini memiliki tujuan untuk membandingkan metode *Support Vector Machine* dan metode *Linear Regression* dengan menggunakan fitur sleksi *Backward Elimination* pada data *sales order* minyak kelapa yang bertujuan untuk mencari model yang terbaik didalam memprediksi *time series* jumlah produksi minyak kelapa. Tabel 1 merupakan perbandingan metode *Support Vector Machine* (SVM), *Support Vector Machine* dan *Backward Elimination* (BE), *Linear Regression* (LR), *Linear Regression* dan *Backward Elimination* (BE). Tabel tersebut menjelaskan bahwa dengan menggunakan metode *Support Vector Machine* didapatkan RMSE 0,127, dengan metode *Support Vector Machine* menggunakan *Backward Elimination* didapatkan RMSE 0,115, dengan metode *Linear Regression* didapatkan RMSE 0,118 dan

metode *Linear Regression* menggunakan *Backward Elimination* didapatkan RMSE 0,118.

Tabel 1. Nilai *Root Mean Square Error*

Dataset	RMSE			
	SVM	SVM+BE	LR	LR+BE
Sales Order (SO)	0.127	0.115	0.118	0.118

Dari Tabel 1. Menunjukkan tabel ringkasan hasil eksperimen dataset *Sales Order* minyak kelapa dan dapat dilihat bahwa eksperimen telah selesai dilakukan dan model terbaik adalah pada algoritma *Support Vector Machine* menggunakan fitur seleksi *Backward Elimination* dengan hasil nilai RMSE yaitu 0,115. Dengan demikina algoritma *Support Vector Machine* lebih baik dibandingkan dengan algoritma *Linear Regression*.



Gambar 1. Perbandingan *Root Mean Square Error*

Sesuai dengan gambar 1 yaitu merupakan grafik perbandingan algoritma *Support Vector Machine*, *Support Vector Machine* menggunakan *Backward Elimination*, *Linear Regression*, serta *Linear Regression* menggunakan *Backward Elimination* menggambarkan bahwa model terbaik pada Algoritma *Support Vector Machine* menggunakan *Backward Elimination*.

IV. KESIMPULAN

Hasil penelitian pada dataset *sales order* dengan menggunakan metode *Support Vector Machine* didapatkan RMSE 0,127, dengan menggunakan metode *Support Vector Machine* dan *Backward Elimination* didapatkan RMSE 0,115, dengan metode *Linear Regression* didapatkan RMSE 0,118 dan dengan menggunakan metode *Linear Regression* dan *Backward Elimination* didapatkan RMSE 0,118. Dari hasil perbandingan tersebut dapat disimpulkan bahwa kinerja *Support Vector Machine* menggunakan *Backward Elimination* lebih baik dibanding *Support Vector Machine*, *Linear Regression* dan *Linear Regression* menggunakan *Backward Elimination*.

V. SARAN

Adapun saran untuk penelitian selanjutnya dapat menambahkan atau menggunakan metode prediksi yang lain, dengan menambahkan jumlah atribut dan mencoba fitur seleksi yang berbeda.

DAFTAR PUSTAKA

- [1]. Sularno, A. 2014. Prediksi Nilai Saham Menggunakan Pemograman Genetika Dan Pemograman Ekspresi Gen, Universitas Gunadarma Depok. Indonesia.
- [2]. Drajana Ivo, C.R. 2017. Metode *Support Vector Machine* Dan *Forward Selection* Prediksi Pembayaran Pembelian Bahan Baku Kopra, Universitas Ichsan Gorontalo, *ILKOM Jurnal Ilmiah*, Volume 9 Nomor 2 Agustus, ISSN: 2087-1716.
- [3]. Setiyorini. T, Asmono. Rizky. T, 2018. Komparasi Metode Neural Network, Support Vector Machine dan Linear Regression Pada Estimasi Kuat Tekan Beton. *Jurnal TECHNO Nusa Mandiri*, ISSN: 1978-2136, Volume 15, Nomor 1.
- [4]. [Widayun, H., Nasution, S.D., Silalahi, N., & Mesran. 2017. Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5, *Media Informatika Budidarma*, Vol 1, No 2, Juni, ISSN: 2548-8368, Hal 32-37.
- [5]. Muktamar, A.B., Setiawan, A.N., & Adji, B.T. 2015. Pembobotan Korelasi pada Naive Bayes Classifier, STMIK AMIKOM Yogyakarta, *Seminar Nasional Teknologi Informasi dan Multimedia*, Februari, ISSN : 2302-3805.
- [6]. Bode Andi. 2017. *K-Nearest Neighbor* Dengan *Feature Selection* Menggunakan *Backward Elimination* Untuk Prediksi Harga Komoditi Kopi Arabika, Universitas Ichsan Gorontalo, *ILKOM Jurnal Ilmiah*, Vol. 9, No. 2, Agustus, ISSN: 2087-1716.
- [7]. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed.). Morgan Kaufmann.
- [8]. Fikri, A. 2013. Penerapan *Data Mining* Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression, Universitas Dian Nuswantoro Semarang.
- [9]. Alda Raharja dan Wiwik Anggraeni, S.Si, M.Kom dan Retno Aulia Vinarti, S.Kom. 2010. Menerapkan Metode *exponential Smoothing* untuk Peramalan Penggunaan waktu telpon di Pt.telkomsel drive3,SISFO-*Surnal Sistem Informasi*.
- [10]. Purwanto, C. Eswaran dan R. Logeswara. 2011 "Sistem Inferensi Adaptif Neuro-Fuzzy yang Disempurnakan Untuk Prediksi Time Series HIV/AIDS, A. Abd Manaf dkk. (Eds.) : ICIES 2011, Part III, CCIS 253, PP. 1-13.