

# PENERAPAN DATA MINING MENGGUNAKAN PERBANDINGAN ALGORITMA GREEDY DENGAN ALGORITMA GENETIKA PADA PREDIKSI RENTET WAKTU HARGA CRUDE PALM OIL

**Desy Ika Puspitasari**

Teknik Informatika, Fakultas Teknologi Informasi, Universitas Islam Kalimantan MAB Banjarmasin  
Email : smile4desyka@gmail.com

## ABSTRACT

*This research applying predictive data mining on CPO (Crude Palm Oil) prices by comparing the modeling optimization feature selection genetic algorithm and greedy algorithm in neural network (NN) method. CPO price predictions done to meet the needs of oil palm investors, through problem analysis CPO price fluctuations time series is uncertain. In order to facilitate the calculation, the steps of the algorithm Genetics and Greedy algorithms implemented by a computer program Rapid Miner Studio. The purpose of this study is compare the accuracy of the evaluation parameters RMSE is generated and program execution time required by the Genetics and Greedy algorithms in solving problems CPO price predictions. The test results indicate that the use of the method accuracy NN Genetic algorithm optimization generally provide better value RMSE (0.096) compared to the algorithm Greedy-forward selection (0.111) and Greedy algorithms-backward selection (0.101). But if the review of program execution time required to resolve problems CPO price forecast, then the algorithm Genetics takes longer than the Greedy algorithm.*

**Keywords:** CPO price forecast, Genetic algorithms, Greedy algorithms, neural network, time series.

## ABSTRAK

Penelitian ini menerapkan data mining pada prediksi harga CPO (Crude Palm Oil) dengan membandingkan pemodelan optimasi seleksi fitur algoritma genetika dan algoritma greedy pada metode neural network (NN). Prediksi harga CPO dilakukan untuk memenuhi kebutuhan investor kelapa sawit, melalui analisa masalah fluktuasi harga CPO time series yang tidak pasti. Guna mempermudah dalam melakukan perhitungan, langkah-langkah dari algoritma Genetika dan algoritma Greedy diimplementasikan dengan program komputer Rapid Miner Studio. Adapun tujuan penelitian ini yaitu mengetahui perbandingan akurasi dengan parameter evaluasi RMSE yang dihasilkan dan waktu eksekusi program yang dibutuhkan oleh algoritma Genetika dan algoritma Greedy dalam menyelesaikan masalah prediksi harga CPO. Hasil pengujian akurasi menunjukkan bahwa penggunaan metode NN optimasi algoritma Genetika secara umum memberikan nilai RMSE yang lebih baik (0,096) dibandingkan algoritma Greedy-forward selection (0,111) dan algoritma Greedy-backward selection (0,101). Akan tetapi jika ditinjau dari waktu eksekusi program yang dibutuhkan dalam menyelesaikan masalah prediksi harga CPO, maka algoritma Genetika membutuhkan waktu yang lebih lama dari pada algoritma Greedy.

**Kata Kunci:** algoritma Genetika, algoritma Greedy, neural network, prediksi harga CPO, time series.

## PENDAHULUAN

Harga minyak kelapa sawit mentah atau yang lebih dikenal dengan CPO (*Crude Palm Oil*) secara historis terus meningkat. Peningkatan harga CPO ini merupakan salah satu indikator penting yang dapat menggambarkan tingkat kesejahteraan petani kelapa sawit (Pasaribu, 2010).

Kebutuhan akan minyak kelapa sawit terus meningkat seiring peningkatan konsumsi minyak nabati dunia. Bahkan, minyak sawit telah dikembangkan sebagai salah satu bahan

bakar nabati untuk mengurangi dampak pemanasan global (Sugiyono, 2009).

Hal tersebut yang melatar belakangi bahwa investasi CPO sangat menjanjikan. Investasi pada CPO digolongkan pada investasi riil (dalam bentuk yang bisa dilihat secara fisik). Proses berinvestasi ini memiliki beberapa tahapan, tahap yang paling penting yaitu melakukan analisis, yaitu mengidentifikasi perkiraan harga, kapan saat melepas dan berapa lama menahan kontrak sampai pada penyusunan kontrak jual-beli ke depan.

Untuk bisa mendapatkan informasi

fluktuasi harga CPO kelapa sawit semacam ini, tidak cukup hanya dengan mengandalkan informasi harga CPO bersifat saat ini saja. Informasi harga CPO dari waktu yang lampau (*past*) harus juga diketahui. Dari informasi-informasi inilah kita dapat membuat sebuah model yang menggambarkan bagaimana sifat informasi harga CPO tersebut dan informasi harga CPO dapat terbentuk sedemikian rupa sampai dengan informasi harga CPO pada saat ini (*present*). Dari model inilah informasi harga CPO dapat diprediksi/diramalkan.

Dalam hal ini *neural network* dipilih karena merupakan model non-linier yang dapat ditraining untuk dapat memetakan data historikal dan data masa depan dari data *time series* dengan cara demikian ekstrak struktur hidden dan hubungannya yang dapat menentukan data yang diramalkan (Lawrence, 1997).

Algoritma yang termasuk penggunaannya dalam permasalahan optimasi adalah algoritma Genetika. Salah satu daya tarik algoritma Genetika terletak pada kesederhanaan dan kemampuan untuk mencari solusi yang baik dan cepat untuk masalah yang kompleks.

Algoritma greedy adalah algoritma yang membentuk solusi, langkah per langkah dan di setiap langkah tersebut akan dipilih keputusan yang paling optimal. Prinsip utama algoritma Greedy adalah “*take what you can get now!*”. Maksudnya ialah, pada setiap langkah dalam algoritma Greedy, diambil keputusan yang paling optimal untuk langkah tersebut tanpa memperhatikan konsekuensi pada langkah selanjutnya. Solusi tersebut dinamakan dengan optimum lokal (Prasetyadi, 2013). Lalu pada saat pengambilan nilai optimum lokal pada setiap langkah, diharapkan tercapai optimum global, yaitu tercapainya solusi optimum yang melibatkan keseluruhan langkah dari awal sampai akhir (Larose, 2006).

Data yang digunakan dalam penelitian ini, sebagai data historikal adalah data harga CPO kelapa sawit Dinas Perkebunan Kalimantan Timur. Data historikal ini nantinya akan digunakan sebagai data training. Data harga tandan buah sawit yang digunakan adalah

tanaman yang berumur  $\geq 10$  tahun, karena pada usia tersebut, tandan sawit siap untuk dipanen.

### Algoritma Genetika

Algoritma genetika berasal dari himpunan solusi yang dihasilkan secara acak yang disebut populasi. Sedangkan setiap individu dalam populasi disebut kromosom, merupakan representasi dari solusi dan masing – masing dievaluasi tingkat kinerjanya (*fitness*) oleh fungsi yang telah ditentukan. Melalui proses seleksi alam atas operator genetik, gen-gen dari 2 (dua) kromosom – disebut parent, diharapkan mampu menghasilkan kromosom baru dengan tingkat fitness yang lebih tinggi sebagai generasi baru atau keturunan (*offspring*) berikutnya. Kromosom – kromosom tersebut akan mengalami iterasi yang disebut generasi (*generation*). Pada setiap generasi, kromosom dievaluasi berdasarkan nilai fungsi fitness. Setelah beberapa generasi maka algoritma genetika akan meng-konvergen untuk mendapat kromosom terbaik, yang merupakan solusi optimal (Hall, Ian H. Witten, 2011).

### Algoritma Greedy

Algoritma Greedy disusun oleh elemen-elemen berikut :

1. Himpunan Kandidat : berisi elemen – elemen pembentuk solusi.
2. Himpunan Solusi : berisi kandidat – kandidat yang terpilih sebagai solusi persoalan.
3. Fungsi Seleksi (*selection function*): memilih kandidat yang paling memungkinkan untuk mencapai solusi optimal. Kandidat yang sudah dipilih pada suatu langkah, tidak pernah dipertimbangkan lagi pada pada langkah selanjutnya.
4. Fungsi kelayakan (*feasible*) : memeriksa apakah suatu kandidat yang telah dipilih dapat memberikan solusi yang layak, yakni kandidat tersebut bersama-sama dengan himpunan solusi yang sudah terbentuk tidak melanggar kendala (*constraints*) yang ada. Kandidat yang layak dimasukkan ke

dalam himpunan solusi, sedangkan kandidat yang tidak layak dibuang dan tidak pernah dipertimbangkan lagi.

5. Fungsi obyektif, yaitu fungsi yang memaksimalkan atau meminimumkan nilai solusi, misalnya panjang lintasan, keuntungan, dan lain-lain (Teng, 2013).

### Root Mean Square Error (RMSE)

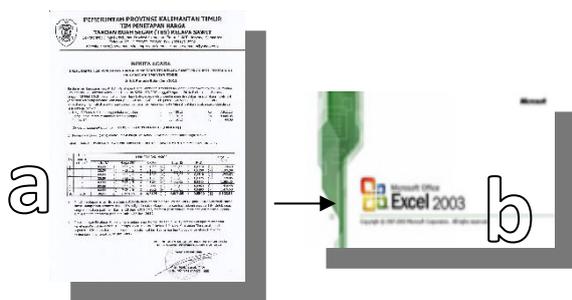
Root Mean Square Error adalah penjumlahan kuadrat error atau selisih antara nilai sebenarnya (aktual) dan nilai prediksi, kemudian membagi jumlah tersebut dengan banyaknya waktu data peramalan dan kemudian menarik akarnya. Jika nilai RMSE semakin kecil maka prediksi model atau variabel tersebut semakin valid. Nilai RMSE dapat dirumuskan sebagai berikut (Pang & Zhao, 2008) :

$$RMSE = \sqrt{\frac{\sum(Aktual - Prediksi)^2}{n}}$$

### METODE PENELITIAN

#### Metode Pengolahan Data Awal

Data *print out* yang didapat dari hasil rapat Tim Penetapan Harga Pembelian TBS kelapa sawit provinsi Kalimantan Timur, dibuat model pengolahan data awal seperti gambar dibawah ini.

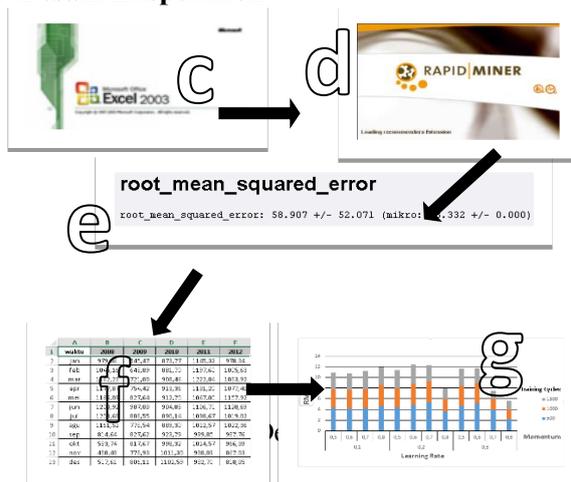


Gambar 1. Metode Pengolahan Data Awal

Keterangan :

- a : adalah data harga TBS hasil rapat TIM Penetapan Harga Pembelian TBS kelapa sawit produksi pekebun di provinsi Kalimantan Timur periode bulanan, dari usia tanaman umur  $\geq 10$  tahun.
- b : adalah aplikasi pengolah angka (*microsoft excel*) untuk mengolah data point “a” sebagai *dataset*.

### Desain Eksperimen



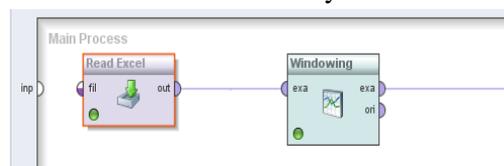
Keterangan Gambar 2 :

- c : file excel berisi dataset yang telah siap untuk di-mining.
- d : software rapidminer melakukan proses perhitungan menggunakan algoritma NN.
- e : hasil evaluasi error menggunakan *root mean square error* (RMSE).
- f : mencatat hasil RMSE dari beberapa parameter yang digunakan pada setiap dataset.
- g : melakukan analisis manual dengan membandingkan nilai RMSE pada setiap parameter dengan *microsoft excel*.

### HASIL DAN PEMBAHASAN

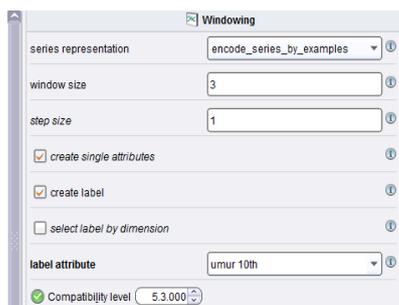
#### Desain Eksperimen dan Pengujian Metode

Algoritma NN akan diuji dengan menggunakan operator windowing. Operator ini digunakan untuk mengolah hasil prediksi *time series*. Berikut contoh tampilannya beserta alur koneksinya.



Gambar 3. Tampilan Main Process dengan Operator Windowing

Parameter windowing yang digunakan yaitu *window size = 3*, kemudian di centang pada *Create label*, pilih *label attribute = umur 10th* untuk menentukan harga CPO sebagai label-nya.



Gambar 4. Jendela Parameter Windowing

Ide peramalan data *time series* pada RapidMiner dengan menggunakan konsep *windowing* memungkinkan untuk mengambil data *time series* dan mengubahnya menjadi format “cross-sectional”. Gambar berikut ini mengilustrasikan idenya. Dalam kasus contoh diatas, menggunakan ukuran jendela (*window size*) 3, ukuran langkah (*step size*) 1.

Row No.	umur 3 th	Bulan	Label	umur 3 th-2	umur 3 th-1	umur 3 th-0
41	1097	01 Mei 11				
42	1106.710	01 Jun 11				
43	1098.670	01 Jul 11				
44	1012.570	01 Agu 11				
45	999.850	01 Sep 11				
46	1014.570	01 Okt 11				
47	938.090	01 Nov 11				
48	932.700	01 Des 11				
49	978.340	01 Jan 12				
50	1005.630	01 Feb 12				
51	1033.920	01 Mar 12				
52	1077.400	01 Apr 12				
53	1157.920	01 Mei 12				
54	1120.690	01 Jun 12				
55	1019.020	01 Jul 12				
56	1022.360	01 Agu 12				
57	987.760	01 Sep 12				
58	966.390	01 Okt 12				
59	867.030	01 Nov 12				
60	818.050	01 Des 12				

Gambar 1. Ilustrasi konsep *windowing*

Keterangan :

a : original data rentet waktu.

b: data yang ter-*window* dengan format “cross sectional”.

Gunakan data dari 3 baris :Jul, Agu, Sep, untuk memprediksi Oct (warna orange).

Gunakan data dari 3 baris : Agu, Sep, Okt, untuk memprediksi Nov (warna biru).

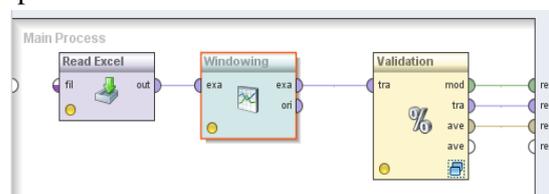
Gunakan data dari 3 baris : Sep, Okt, Nov, untuk memprediksi Des (warna merah).

c: data harga TBS hasil yang ter-prediksi.

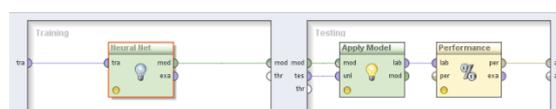
### Pengujian metode menggunakan NN

Pengujian pada tahap ini adalah untuk mengetahui apabila prediksi harga CPO yang menggunakan metode Neural Network,

bagaimanakah hasil akurasi sebelum dilakukan optimasi. Hal ini untuk melihat perbandingan nilai RMSE sebelum dan sesudah optimasi.



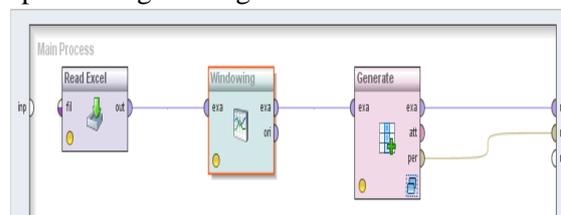
Gambar 6. Proses Pengujian Metode menggunakan Neural Network



Gambar 7. Jendela Training & Testing Metode Neural Network

### Pengujian metode menggunakan NN dan dioptimasi Algoritma Genetika

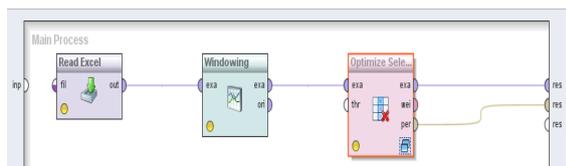
Langkah selanjutnya yaitu prediksi harga CPO pengujian metode menggunakan NN yang dioptimasi dengan algoritma Genetika. Berikut ini tampilan jendela Main Process pengujian NN optimasi algoritma genetika :



Gambar 8. Proses Pengujian metode Neural Network dengan optimasi algoritma Genetika

### Pengujian metode menggunakan NN dan dioptimasi Algoritma Greedy – Forward Selection

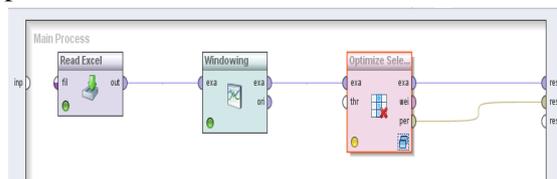
Tahap selanjutnya adalah pengujian mengoptimasi metode NN dengan algoritma Greedy forward selection. Pada jendela main process, operator Optimize Selection, pilih parameter *Selection direction = forward*.



Gambar 9. Proses Pengujian metode Neural Network dengan optimasi algoritma Greedy – Forward selection

**Pengujian metode menggunakan NN dan dioptimasi Algoritma Greedy – Backward Elimination**

Tahap yang terakhir adalah pengujian mengoptimasi metode NN dengan algoritma Greedy backward elimination. Pada jendela main process, operator *Optimize Selection*, pilih parameter *Selection direction = backward*.



Gambar 10. Proses Pengujian metode Neural Network dengan optimasi algoritma Greedy – Backward

**Analisa dan Hasil Eksperimen**

Tabel 1. Nilai RMSE pada masing-masing Metode

No.	Implementasi Metode	RMSE	SE	Jumlah atribut yang dihasilkan	Waktu eksekusi
Window size = 3, validation = 10					
1	NN tanpa optimasi	0,101	0,013	3	1 detik
2	NN optimasi Algoritma Genetika	0,096	0,012	1	30 detik
3	NN optimasi Algoritma Greedy_forward	0,111	0,013	2	1 detik
4	NN optimasi Algoritma Genetika_backward	0,101	0,013	3	1 detik

Dari hasil nilai RMSE pada masing-masing tahapan metode diatas, dengan menggunakan parameter window size = 1 dan x-validation-nya 10, menunjukkan untuk optimasi neural network pada prediksi harga CPO, algoritma Genetika terbukti paling akurat dibanding metode neural network yang dioptimasi dengan algoritma greedy forwad dan algoritma greedy backward. Dengan demikian,

**SIMPULAN**

Hasil pengujian akurasi menunjukkan bahwa penggunaan metode NN setelah di optimasi menggunakan algoritma Genetika, secara umum memberikan nilai RMSE yang lebih baik (0,096) dibandingkan algoritma Greedy-forward selection (0,111) dan algoritma Greedy-backward selection (0,101), terjadi peningkatan akurasi dari metode NN sebelum dioptimasi. Akan tetapi jika ditinjau dari waktu eksekusi program yang dibutuhkan

optimasi NN dengan algoritma Genetika dapat memberikan pemecahan untuk permasalahan

prediksi harga CPO. Namun apabila ditinjau dari waktu eksekusi program yang dibutuhkan dalam menyelesaikan masalah prediksi harga CPO, maka algoritma Genetika membutuhkan waktu yang lebih lama dari pada algoritma Greedy.

dalam menyelesaikan masalah prediksi harga CPO, maka algoritma Genetika membutuhkan waktu yang lebih lama dari pada algoritma Greedy.

**DAFTAR RUJUKAN**

Hall, Ian H. Witten, E. F. M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. (ELSEVIER, Ed.) (Third

- Edit). Morgan Kaufmann (MK).
- Larose, D. T. (2006). *Data Mining Methods and Models*. (A. J. W. & Sons, Ed.). Wiley-Interscience.  
<http://doi.org/10.1002/0471756482>
- Lawrence, R. (1997). Using Neural Networks to Forecast Stock Market Prices, 1–21.
- Pang, M. B., & Zhao, X. P. (2008). Traffic Flow Prediction of Chaos Time Series by Using Subtractive Clustering for Fuzzy Neural Network Modeling. *Proceedings - Second International Symposium on Intelligent Information Technology Application, IEEE Computer Society, 1*, 23–27. <http://doi.org/10.1109/IITA.2008.50>
- Pasaribu, W. P. A. (2010). *Analisa Harga Pembelian TBS Kelapa Sawit Produksi Petani Rakyat*. Fakultas Pertanian. Universitas Sumatera Utara.
- Prasetyady, D. (2013). Penerapan Algoritma Greedy pada Perbaikan Sudut Pergerakan Awal dalam Snellius Ray Path Tracing Tomografi. Retrieved August 10, 2015, from <http://informatika.stei.itb.ac.id/~rinaldi.mu/nir/Stmik/2012-2013/Makalah2012/Makalah-IF3051-2012-016.pdf>
- Sugiyono, A. (2009). *Pengembangan Bahan Bakar Nabati untuk Mengurangi Dampak Pemanasan Global*. Jogjakarta.
- Teng, M. O. (2013). Penerapan Algoritma Greedy dalam Permainan Tetris. Retrieved August 10, 2015, from <http://informatika.stei.itb.ac.id/~rinaldi.mu/nir/Stmik/2012-2013/Makalah2012/Makalah-IF3051-2012-015.pdf>