

## Prediksi Jumlah Produksi *Coconut Oil* Menggunakan *k-Nearest Neighbor* dan *Backward Elimination*

**Ivo Colanus Rally Drajana**

Fakultas Ilmu Komputer, Jurusan Teknik Informatika, Universitas Ichsan Gorontalo  
Email: ivocolanusrally@gmail.com

### Abstrak

Hampir keseluruhan tanaman pohon kelapa memiliki banyak bagiannya yang dimanfaatkan oleh manusia, sehingga tumbuhan ini dianggap tumbuhan serbaguna. Minyak kelapa (*coconut oil*) dihasilkan oleh buah pohon kelapa. Produksi jumlah *coconut oil* menjadi bagian penting disetiap perusahaan yang bergerak di bidang produksi *coconut oil* dengan tujuan mencapai target hasil produksi. Produksi minyak setiap hari mengalami perubahan fluktuatif. Untuk memenuhi permintaan *sales order* dari *customer* perusahaan, pengembangan *system* perusahaan sangat diperlukan untuk prediksi jumlah produksi *coconut oil*. Penelitian ini *k-Nearest Neighbor* dengan *feature selection* digunakan untuk memprediksi jumlah produksi *coconut oil* berdasarkan data *time series Sales Order* (SO). Metode yang sering digunakan dalam prediksi adalah metode *time series*, dimana nilai masa mendatang dilihat berdasarkan rentet waktu pengamatan. Performa metode *time series* sangatlah unggul dengan melihat nilai-nilai di masa lalu. Algoritma *k-Nearest Neighbor* merupakan metode non-parametrik yang digunakan untuk klasifikasi dan regresi. *k-Nearest Neighbor* menghitung jarak datanya pada semua sampel. Jarak *Euclidean* lebih sering digunakan. Algoritma *k-Nearest Neighbor* adalah metode penerapan algoritma *supervised* di mana algoritma ini terbagi menjadi dua bagian yaitu *supervised learning* dan *unsupervised learning*. *Feature selection* yakni *Backward Elimination* meningkatkan performa yang lebih baik. *Backward Elimination* diterapkan untuk mengeliminasi atribut atau variabel yang dianggap tidak relevan, variabel yang tidak berpengaruh serta tidak signifikan dalam model dihapuskan dari dalam model. Algoritma *k-Nearest Neighbor* dihasilkan model terbaik yang dilihat berdasarkan nilai *error* terkecil yaitu 0.111. Kemudian algoritma *k-Nearest Neighbor* menggunakan *Backward Elimination* dihasilkan model terbaik yang

dilihat berdasarkan nilai *error* terkecil yaitu 0.109. Seleksi fitur yaitu *Backward Elimination* menghasilkan kinerja lebih baik

**Kata kunci:** Metode *k-Nearest Neighbor*; *Backward Elimination*; *Data Mining*; Produksi *Coconut Oil*

### Abstract

*Almost all coconut trees have many parts that are used by humans, so this plant is considered a multipurpose plant. Coconut oil (coconut oil) is produced by the fruit of coconut trees. Production of coconut oil becomes an important part in every company engaged in the production of coconut oil with the aim of achieving production targets. Oil production every day experiences fluctuating changes. To meet the demand for sales orders from corporate customers, the development of the company's system is needed to predict the amount of coconut oil production. This research *k-Nearest Neighbor* with feature selection is used to predict the amount of coconut oil production based on time series Sales Order (SO) data. The method often used in predictions is the time series method, where the future value is seen based on the time series of observations. The performance of the time series method is very superior by looking at values in the past. The *k-Nearest Neighbor* algorithm is a non-parametric method used for classification and regression. *k-Nearest Neighbor* calculates the distance of the data in all samples. Euclidean distance is used more often. The *k-Nearest Neighbor* algorithm is a method of applying a supervised algorithm where this algorithm is divided into two parts, namely supervised learning and unsupervised learning. Feature selection, the *Backward Elimination* improves better performance. *Backward Elimination* is applied to eliminate attributes or variables that are considered irrelevant, variables that have no effect and are not significant in the model are eliminated from the model. The *k-Nearest Neighbor* algorithm is produced by the best model which is based on the smallest error value of 0.111. Then the *k-Nearest Neighbor* algorithm uses the *Backward Elimination* generated by the best model which is seen based on the smallest error value of 0.109. Feature selection that is *Backward Elimination* produces better performance.*

**Keywords:** *k-Nearest Neighbor Method*; *Backward Elimination*; *Data Mining*; *Coconut Oil Production*

## A. PENDAHULUAN

Tumbuhan pohon kelapa memiliki banyak manfaat hampir semua bagian dari pohon digunakan manusia, sehingga tumbuhan ini dianggap tumbuhan serbaguna terutama bagi masyarakat dibagian pesisir. Minyak kelapa (coconut oil) dihasilkan dari buah pohon kelapa. Produksi minyak

kelapa (*coconut oil*) tiap hari mengalami perubahan atau fluktuatif. Metode yang sering digunakan dalam prediksi yaitu metode *time series*, dimana nilai masa mendatang dapat dilihat berdasar urutan waktu pengamatan. Kemampuan dari metode *time series* sangat unggul dengan cara melihat nilai-nilai di masa lalu (Sularno, 2014).

Algoritma *k-Nearest Neighbor* merupakan metode penerapan algoritma *supervised* dimana algoritma ini terbagi menjadi dua bagian yaitu *supervised learning* dan *unsupervised learning*. Algoritma ini memiliki tujuan *supervised learning* dalam menghasilkan pola yang baru, sedangkan tujuan dari *unsupervised learning* adalah menghasilkan pola didalam sebuah data. Kelebihan dari algoritma *k-Nearest Neighbor* adalah terhadap training dataset yang memiliki banyak *noise* serta sangat efektif terhadap jumlah data training besar, namun algoritma *k-Nearest Neighbor* memiliki kekurangan didalam menentukan nilai *k* dan pemilihan atribut terbaik (Bode, 2017).

*Feature Selection* adalah salah satu proses memilih suatu subset dari fitur *original* dengan mengeliminasi fitur yang dianggap tidak relevan. Proses pemilihan fitur akan lebih meningkatkan efisiensi. Adapun tujuan utama seleksi fitur yaitu fokus pada pencarian sebuah data yang relevan. Fitur yang dianggap tidak relevan serta fitur yang berlebihan mempengaruhi terhadap hasil, sehingga adanya proses pemilihan fitur bisa mengidentifikasi data (Poonguzhali, E., Vijayalakshmi, M., Shamily K., Priyadarshini, V., 2014). *Feature selection* yakni *Backward Elimination* meningkatkan performa yang lebih baik (Bode, 2017). *Backward Elimination* digunakan dalam mengeliminasi atribut variabel yang dianggap tidak relevan, jika terdapat variabel tidak berpengaruh dan tidak dianggap tidak signifikan dalam

model akan dihapus dari dalam model. Kinerja metode *backward elimination* lebih baik dibanding cara statistik dalam menyeleksi fitur. Metode *backward elimination* menghasilkan kinerja lebih baik dibandingkan dengan cara statistik dalam penyeleksian fitur. Performa atau kinerja terbaik dihasilkan dari sensitivitas, spesifisitas serta keakuratan (Narin, Isler, dan Ozer, 2014). Dalam penelitian ini bertujuan menerapkan model prediksi dengan menggunakan algoritma *k-Nearest Neighbor* menggunakan fitur seleksi *Backward Elimination* untuk mendapatkan model yang lebih baik didalam melakukan prediksi *time series*.

## **B. METODE**

Tahapan proses preprocessing pengolahan data univariat *time series* yang berupa numerik data harian. Data direkap menjadi data mingguan, agar mempermudah proses pengelolannya, kemudian data di ubah ke *ascending* ke *descending* proses ini dilakukan dengan cara memanfaatkan toolbar *sort & filter* yang ada pada *excel*. Setelah itu data dinormalisasi. Pada database atau basis data, proses normalisasi bertujuan untuk meminimalisir terjadinya berbagai anomaly data dan ketidak konsistensinya data. Tujuan utama normalisasi database yaitu untuk mendapatkan data yang berdimensi kecil yang tetap dapat mewakili data asli tanpa kehilangan karakteristik data (Drajana, 2017).

Persamaan normalisasi sebagai berikut.

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

Dimana:

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

Untuk mengembalikan ukuran dataset yang sudah ternormalisasi ke data asli dengan cara denormalisasi dataset

Persamaan dapat dilihat sebagai berikut.

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

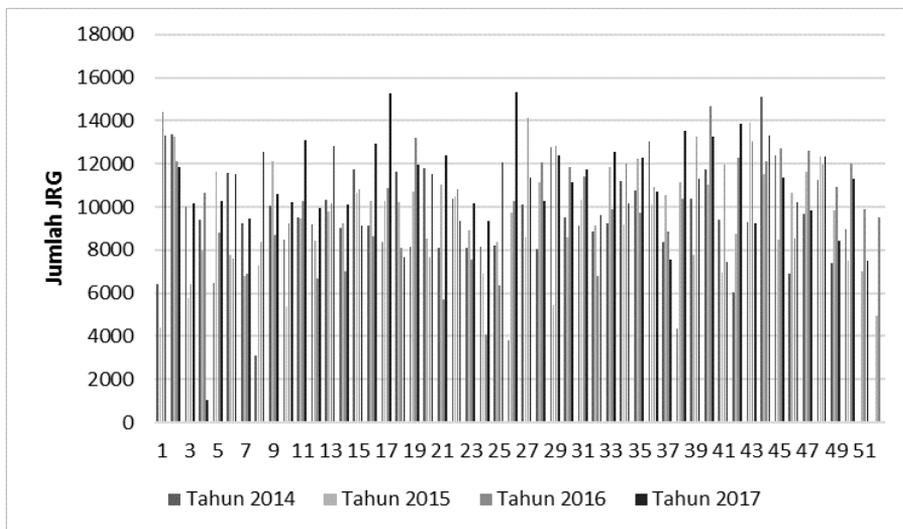
Dimana:

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

Penelitian ini dilakukan beberapa tahap, tahap pertama mengubah data harian menjadi data mingguan, setelah itu data dinormalisasi. Tahap kedua olah data univariat ke multivariat. Tahap ketiga menentukan parameter *k-Nearest Neighbor* dan *k-Nearest Neighbor* menggunakan seleksi fitur *Backward Elimination*. Pengujian ini dilakukan untuk menghasilkan model terbaik dengan melihat nilai *root mean square error* terkecil. Tahap ke-4 mengaplikasikan model terbaik yang dihasilkan algoritma *k-Nearest Neighbor* menggunakan *Backward Elimination* terhadap data testing untuk melakukan prediksi. Dalam tahap eksperimen, beberapa sampel data set digunakan untuk melakukan uji coba, sebelum

dilakukan uji coba dilanjutkan dengan penentuan parameter, nilai parameter yang digunakan proses uji coba yaitu 1, 3, 5, 7, 9, 11, *number of validation* 10, nilai *K* 1, lalu dilakukan proses testing untuk melihat hasil *root mean square error* (RMSE).

Sumber data set pada penelitian ini di ambil PT. Multi Nabati Sulawesi Unit Maleo Kabupaten Puhuwato pada 17 Mei 2018. Data yang dikumpulkan adalah data kuantitatif harian *time series* univariat. Jenis data ini adalah data SO (Sales Order) dari tahun 2014 – 2017. Sesuai dengan ketentuan jam kerja yang ada di perusahaan bahwa dataset yang ada hari libur dan hari minggu tidak terjadi transaksi pertahunnya. Dataset dari yang diambil dari variabel sales order memiliki 208 *record* yang berasal dari setiap pesanan *costumer*.



**Gambar 1. Grafik SO (Sales Order) Minyak Kelapa Tahun 2014-2017**

Gambar 1 menjelaskan bahwa telah terjadi fluktuasi pesanan minyak kelapa (Coconut Oil) dari customer sehingga terjadi masalah dalam persediaan jumlah produksi *coconut oil*.

Data mining yaitu sebuah proses menjalankan satu atau lebih dari satu teknik pembelajaran komputer (machine learning) yang bertujuan menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis (Widayun, H., Nasution, S.D., Silalahi, N., Mesran, 2017). Dalam pengertian lain yaitu *induction-based learning* dimana proses pembentukan definisi-definisi konsep umum yang dikerjakan dengan mengobservasi contoh spesifik dari konsep yang dipelajari. KDD (Knowledge Discovery in Databases) adalah penerapan metode saintifik yang terdapat pada data mining. Dalam sebuah konteks tersebut, data mining adalah langkah dari proses *Knowledge Discovery In Database* (Muktamar, A.B., Setiawan, A.N., Adji, B.T., 2015)

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 (Bode, 2017).

Algoritma *k-Nearest Neighbor* merupakan metode non-parametrik yang dapat digunakan untuk klasifikasi dan regresi. Dengan demikian, *k-Nearest Neighbor* menghitung jarak data dan semua sampel. Jarak *Euclidean* adalah ukuran yang paling banyak digunakan untuk tujuan ini. Sampel pelatihan peringkat dalam urutan menaik menurut jarak yang dihitung, diambil *k* sampel terdekat (tetangga1, tetangga2, ..., tetangga). Kemudian, mereka digunakan untuk menghitung label kelas yang paling dominan. Nilai yang dipilih dari *k* dapat mempengaruhi kinerja dari

teknik ini (Jesus M., Sergio R., Isaac T., Francisco H., 2016). Meskipun algoritma *k-Nearest Neighbor* telah menunjukkan kinerja luar biasa dalam berbagai variasi masalah, algoritma ini tidak memiliki skalabilitas untuk mengelola dataset besekala besar atau tinggi. Ketepatan algoritma *k-Nearest Neighbor* ditentukan adanya data signifikan dan tidak signifikan, atau jika bobot fitur tersebut sama dengan signifikannya terhadap klasifikasi (Wijana, 2015). Algoritma *k-Nearest Neighbor* adalah metode penerapan algoritma *supervised* algoritma tersebut dibagi menjadi dua jenis yaitu *supervised learning* dan *unsupervised learning*. Tujuan algoritma ini yaitu *supervised learning* untuk mendapatkan pola baru, sedangkan tujuan dari *unsupervised learning* yaitu mendapatkan pola dalam sebuah data. Algoritma *k-Nearest Neighbor* memiliki kelebihan terhadap training data set yang banyak noise dan sangat efektif kepada jumlah data training tinggi atau besar (Bode, 2017).

Persamaan dapat dilihat sebagai berikut.

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (3)$$

Keterangan:

$x$  = sampel data

$y$  = data uji

$D$  = jarak

$$(x') = \frac{1}{k} \sum_{i=1}^k y_i(x') \quad (4)$$

Keterangan:

$x'$  = Perkiraan atau estimasi

$k$  = Jumlah tetangga terdekat

$(x')$  = Tetangga terdekat

$y_i$  = Output tetangga terdekat

*K-Fold Cross Validation* yaitu teknik validasi bertujuan untuk membagikan data ke dalam  $k$  bagian, kemudian masing-masing bagian dimasukan kedalam proses klasifikasi. Dengan menggunakan *K-Fold Cross Validation* dapat melakukan testing sebanyak  $k$ . Dalam setiap testing digunakan satu data testing selanjutnya  $k-1$  akan dijadikan data training. Kemudian data testing tersebut masuk kedalam proses penukaran data dengan satu data training, maka tiap testing akan diperoleh hasil dari data testing yang berbeda- beda (Zega, 2014).

Adapun alat dan bahan yang digunakan ddalam penelitian ini terbagi menjadi dua jenis yaitu hardware/software. Hardware meliputi, PC (Laptop), buku, polpen, kertas, dan kalender untuk melihat tanggal dan tahun data saat merekap data. Software berupa tools RapiMiner untuk pengolahan data mencari nilai error terkecil, Microsoft Office Word 2016 untuk penyusunan naskah, Microsoft Office excel 2016 untuk merekap data harian ke mingguan, *ascending* ke *descending*, dan mengubah data ke data multivariate.

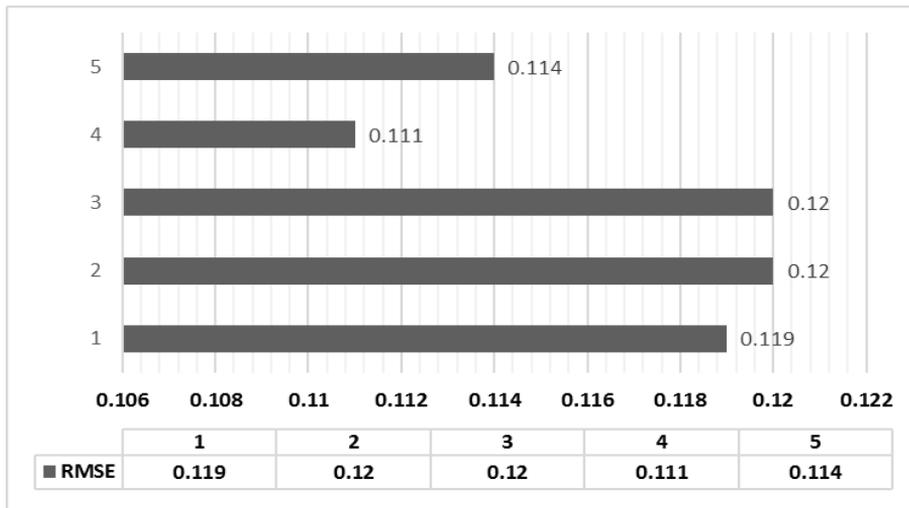
### C. HASIL DAN PEMBAHASAN

Penelitian ini memiliki tujuan untuk menerapkan algoritma *k-Nearest Neighbor* menggunakan *Backward Elimination* pada data SO (Sales Order) *coconut oil* yang bertujuan untuk mencari model yang terbaik didalam memprediksi *time series* jumlah produksi *coconut oil*.

**Tabel 1. Nilai Root Mean Square Error SO Coconut Oil k-Nearest Neighbor**

<b>Variabel Periode</b>	1	2	3	4	5
<b>Of Validation</b>	10	10	10	10	10
<b>K</b>	9	11	5	11	7
<b>RMSE</b>	0.119	0.120	0.120	<b>0.111</b>	0.114

Pada Tabel 1. dibawah ini merupakan tabel ringkasan hasil eksperimen SO (Sales Order) *coconut oil* dengan penentuan model menggunakan algoritma *k-Nearest Neighbor* yang dari 1 sampai 5 variabel periode menggunakan *cross validation* 10 dan nilai *k* 1, 3, 5, 7, 9, 11, yang dilihat berdasarkan nilai *root mean square error* (RMSE) terkecil. Dari hasil ekperimen dihasilkan model terbaik yang dilihat berdasarkan nilai *error* terkecil 0.111 dengan jumlah variabel periode 4 jumlah validasi 10 dan nilai *k* yaitu 11.



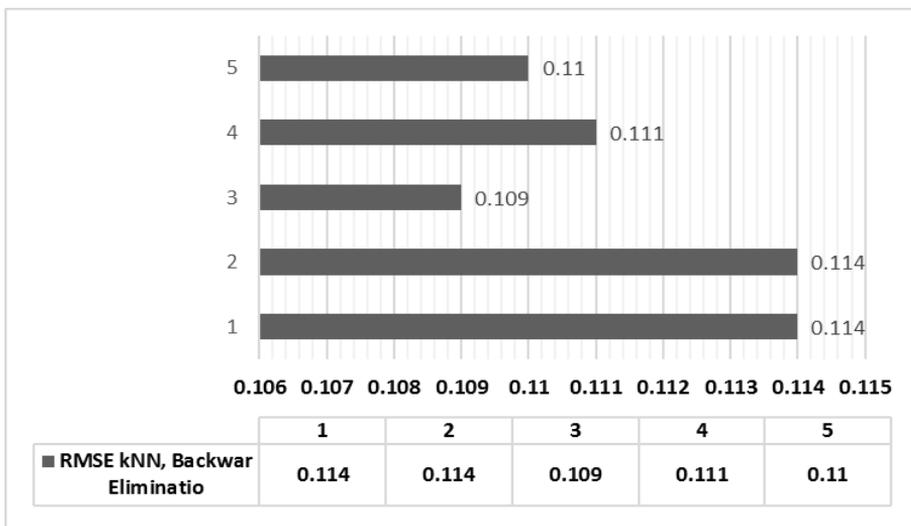
Gambar 2. Grafik Model *k-Nearest Neighbor* SO Coconut Oil

Pada gambar 2 merupakan grafik model terbaik untuk data SO *coconut oil*, bahwa model yang terbaik ada pada variabel periode 4, nilai validasi 10 dengan *K* 11, dengan nilai *root mean square error* sebesar 0.111. Setelah diketahui hasil dari model *k-Nearest Neighbor*, ekperimen berikutnya dengan menambahkan *feature selection* yaitu *backward selection*.

**Tabel 2.** Nilai RMSE SO *Coconut Oil* k-NN menggunakan *Backward Elimination*

Variabel Periode	1	2	3	4	5
Of Validation	10	10	10	10	10
K	11	11	7	7	9
RMSE kNN, Backwar Eliminatio	0.114	0.114	0.109	0.111	0.110

Pada Tabel 2 ini merupakan hasil ringkasan eskperimen penentuan model dengan algoritma *k-Nearest Neighbor* menggunakan *Backward Elimination* dari 1 sampai 5 variabel periode, *cross validation* 10 dengan nilai *K* 1, 3, 5, 7, 9, 11, diambil berdasarkan nilai *root mean square error* terkecil. Maka model terbaik berdasarkan nilai *error* terendahl adalah 0.109 jumlah variabel periode 3, jumlah validasi 10 dengan nilai *k* 7.



**Gambar 3.** Grafik Model *k-Nearest Neighbor* – *Backward Elimination*

Pada gambar 3 merupakan grafik model terbaik *k-Nearest Neighbor* menggunakan *Backward Elimination* untuk data SO *coconut*

*oil*, bahwa model yang terbaik ada pada variabel periode 3, nilai validasi 10 dengan  $K$  7, dan nilai *root mean square error* sebesar 0.109.

Prediksi jumlah produksi *coconut oil* berdasarkan data SO (Sales Order) algoritma *k-Nearest Neighbor* dan *Feature Selection* menggunakan *Backward Elimination* telah selesai dilakukan. Pada eksperimen menggunakan algoritma *k-Nearest Neighbor* dihasilkan model terbaik yang dilihat berdasarkan nilai *error* terkecil yaitu 0.111 dengan variabel periode 4, validasi 10 dan nilai  $k$  11. Kemudian pada eksperimen algoritma *k-Nearest Neighbor* menggunakan fitur seleksi *Backward Elimination* dihasilkan model terbaik yang dilihat berdasarkan nilai *error* terkecil yaitu 0.109 dengan variabel periode 3, validasi 10 dan nilai  $k$  7. Sesuai hasil eksperimen yang ada, menjelaskan bahwa *feature selection* yaitu *Backward Elimination* menghasilkan performa yang lebih baik.

Dalam mendukung kegiatan kinerja bagian produksi dalam memprediksi jumlah produksi *coconut oil* agar lebih akurat, penerapan system yang peneliti telah lakukan yaitu algoritma *k-Nearest Neighbor* menggunakan *Backward Elimination* dapat mempermudah dan lebih efisiensi dalam memprediksi jumlah produksi *coconut oil*. Manfaat dari penelitian ini bagi perusahaan bisa digunakan sebagai masukan bagi bagian produksi dalam bertindak dimasa mendatang.

Penelitian ini memiliki kelemahan yang terletak pada preprosesing data, karena data yang di olah merupakan data mentah yang membutuhkan rekapitulasi data, serta dalam penentuan nilai  $k$  pada saat uji coba data sampel. Kelebihan dari penelitian ini yaitu nilai dari RMSE (*root mean square error*) lebih kecil dibandingkan dengan peneliti-peneliti sebelumnya. Disarankan untuk peneliti berikutnya dapat menggunakan

algoritma lain dengan tujuan utama untuk menghasilkan tingkat *error* yang lebih kecil lagi.

## D. PENUTUP

### Simpulan dan Saran

Prediksi jumlah produksi *coconut oil* dengan menerapkan algoritma *k-Nearest Neighbor* dengan *Feature Selection* menggunakan *Backward Elimination* telah berhasil diselesaikan. Mengacu pada hasil eksperimen yang telah dilakukan, menjelaskan bahwa *feature selection* yaitu *Backward Elimination* meningkatkan performa yang lebih baik lagi dibandingkan algoritma *k-Nearest Neighbor* tanpa *feature selection*. Untuk peneliti selanjutnya agar dapat melakukan uji coba menggunakan *feature selection* selain *Backward Elimination* dengan tujuan mendapatkan hasil *error* yang lebih kecil.

## DAFTAR PUSTAKA

- 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.
- 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.
- Jesus, M., Sergio, R., Isaac, T., & Francisco, H. 2016. *kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors Classifier for Big Data*, *Knowledge-Based Systems*.

- 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.
- Narin, A., Isler, Y., & Ozer, M. 2014. Investigating the performance improment of HRV indices in CHF using feature selection methods based on backward elimination and statistical significance, computer in biology and medicine.
- Poonguzhali, E., Vijayalakshmi, M., Shamily, K., & Priyadarshini. V. 2014. Hybrid Feature Selection Algorithm For High Dimensional Database, Pondicherry University India, *International Journal Of Engineering Trends And Technology (IJETT)* – Volume 8 Number 9.
- Sularno, A. 2014. Prediksi Nilai Saham Menggunakan Pemograman Genetika Dan Pemograman Ekspresi Gen, Universitas Gunadarma Depok. Indonesia.
- 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.
- Wijana, Katon. 2015. Program Bantu Prediksi Penjualan Barang Menggunakan Metode KNN Studi Kasus: U.D. Anang, *Jurnal Eksis*. Vol 08 No 02 November.
- Zega, S. Artaty. 2014. Penggunaan Pohon Keputusan Untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah, Politeknik Negeri Batam, Yogyakarta. ISSN: 1907-5022.