

Input Variable Selection for Oil Palm Plantation Productivity Prediction Model

Input Variable Selection untuk Model Prediksi Produktivitas Perkebunan Kelapa Sawit

Andiko Putro Suryotomo¹, Agus Harjoko²

¹ Jurusan Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

² Departemen Ilmu Komputer dan Elektronika, Universitas Gadjah Mada, Yogyakarta, Indonesia

^{1*} andiko.ps@upnyk.ac.id, ²aharjoko@ugm.ac.id

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: Desember 2022

Revised: January 2023

Accepted: February 2023

Published: February 2023

Keywords: IVS; oil palm; expert knowledge

Kata kunci: IVS; kelapa sawit; pengetahuan ahli

Abstract (menggunakan style abstract)

Purpose: This study aims to implement and improve a wrapper-type Input Variable Selection (IVS) to the prediction model of oil palm production utilizing oil palm expert knowledge criteria and distance-based data sensitivity criteria in order to measure cost-saving in laboratory leaf and soil sample testing.

Methodology: The proposed approach consists of IVS process, searching the best prediction model based on the selected variables, and analyzing the cost-saving in laboratory leaf and soil sample testing.

Findings/result: The proposed method managed to effectively choose 7 from 19 variables and achieve 81.47% saving from total laboratory sample testing cost.

Value: This result has the potential to help small stakeholder oil palm planter to reduce the cost of laboratory testing without losing important information from their plantation.

Abstrak (menggunakan style abstrak)

Tujuan: Penelitian ini bertujuan untuk mengimplementasikan dan memperbaiki *Input Variable Selection (IVS)* tipe *wrapper* pada model prediksi produksi kelapa sawit dengan menggunakan kriteria pengetahuan ahli kelapa sawit dan kriteria sensitivitas data berbasis jarak untuk mengukur penghematan biaya pengujian sampel daun dan tanah di laboratorium.

Metode: Pendekatan yang diusulkan terdiri dari proses IVS, mencari model prediksi terbaik berdasarkan variabel yang dipilih, dan menganalisis penghematan biaya dalam pengujian sampel daun dan tanah di laboratorium.

Hasil: Metode yang diusulkan berhasil dengan efektif memilih 7 dari 19 variabel dan mencapai penghematan 81,47% dari total biaya pengujian sampel laboratorium. Nilai tambah: Hasil penelitian ini berpotensi membantu perkebunan kelapa sawit, khususnya perkebunan rakyat, untuk mengurangi biaya pengujian laboratorium dan tanpa kehilangan informasi penting dari perkebunan mereka.

1. Pendahuluan

Sebagai salah satu negara penghasil kelapa sawit terbesar di dunia, Indonesia menghasilkan 85 % dari total produksi kelapa sawit di dunia[1]. Perkembangan perkebunan kelapa sawit yang sangat pesat di Indonesia diiringi dengan munculnya banyak permasalahan dan konflik lingkungan serta sosio kultural. Untuk mengurangi dampak negatif terhadap lingkungan dan sosio kultural tersebut, perusahaan dan petani perkebunan kelapa sawit mulai memikirkan cara untuk memaksimalkan produksi kelapa sawit mereka tanpa membuka lahan baru. Salah satu caranya adalah mengadopsi langkah implementasi teknologi informasi yang telah terbukti mampu membantu optimasi dan mempermudah operasional pengguna di dunia pertanian dan perkebunan.

Beberapa contoh implementasi teknologi informasi pada dunia perkebunan dan pertanian antara lain adalah pemanfaatan penggunaan pengolahan citra digital dan pengenalan pola untuk memprediksi tingkat kemasakan buah pisang [2], pemanfaatan pengolahan citra digital dan jaringan syaraf tiruan untuk mengenali spesies gulma [3], pemanfaatan pengolahan citra digital dan algoritma Viola-Jones untuk menghitung cacah pohon kelapa sawit dari foto udara secara otomatis [4], dan pemanfaatan jaringan syaraf tiruan untuk membangun model prediksi produksi pada perkebunan kelapa sawit[5]–[10].

Pengumpulan dan pengujian data sampel tanah dan daun tanaman di perkebunan kelapa sawit lapangan menjadi penghambat petani kebun rakyat dan koperasi karena biaya yang besar. Hal ini menyebabkan perlunya seleksi variabel input dan data untuk mendapatkan tingkat kepentingan (*importance*) dari masing-masing variabel sehingga petani dapat memilih variabel input yang akan digunakan dan pada akhirnya mampu mengurangi biaya dan menghemat waktu pengumpulan sampel data dan pengujian sampel data di laboratorium.

Penelitian ini mengembangkan dan memperbaiki sebuah metode seleksi variabel input yang menggabungkan dua buah metode berdasarkan pada pendapat pengetahuan ahli perkebunan kelapa sawit dan juga sensitivitas data berbasis jarak [11] (selanjutnya akan disebut dengan nama metode CZCL). Model prediksi produksi tanaman kelapa sawit yang dikembangkan menggunakan jaringan syaraf tiruan *feedforward* dengan metode pembelajaran *backpropagation*.

Metode CZCL digunakan untuk melakukan seleksi variabel input pada permasalahan model prediksi kualitas tekstil[11]. Karakteristik data tekstil yang digunakan pada penelitian tersebut memiliki banyak kemiripan dengan data sampel tanah dan daun tanaman perkebunan kelapa sawit. Hasil yang diperoleh adalah terpilihnya 6 dari 9 variabel dengan tingkat akurasi model prediksi menyamai model prediksi dengan 9 variabel input. Sementara pada [12] dilakukan

modifikasi kriteria sensitivitas data berbasis jarak dengan menambahkan kriteria *disagreement* dimana perhitungan sensitivitas data hanya dilakukan pada data input yang memiliki output berbeda sehingga mengurangi kompleksitas komputasi. Metode seleksi variabel input yang digunakan mampu memilih 4 variabel signifikan dari kandidat 28 variabel input.

Pada penelitian ini kemudian akan dilakukan analisis penghematan biaya pengujian laboratorium untuk sampel daun tanaman dan tanah perkebunan kelapa sawit. Model prediksi yang memiliki jumlah variabel input paling sedikit dan memiliki performa prediksi (diukur dari nilai RMSE dan MAPE) terbaik akan dihitung total kebutuhan biaya pengujian laboratoriumnya.

2. Metode/Perancangan

Proses seleksi variabel input yang diusulkan pada penelitian ini meliputi tiga proses utama yaitu proses seleksi variabel input, proses pencarian model prediksi terbaik, dan proses perhitungan persentase penghematan biaya pengujian laboratorium. Dalam definisi IVS, model seleksi variabel input yang digunakan adalah IVS j tipe/jenis *wrapper* [13].

2.1. Proses Input Variable Selection (IVS)

Metode seleksi variabel input (IVS) yang akan digunakan merupakan metode yang menggabungkan dua metode yaitu metode IVS yang mempertimbangkan kriteria pengetahuan ahli tentang relasi antara variabel input dan variabel output serta kriteria sensitivitas data yang berbasis pada metode jarak (*distance*). Pengetahuan ahli pada penelitian ini didapatkan dari praktisi perkebunan kelapa sawit yang sudah memiliki pengalaman menjadi manajer kebun dan konsultan perkebunan kelapa sawit selama puluhan tahun.

Pengetahuan ahli yang akan digunakan dalam metode ini adalah pengetahuan terkait dengan korelasi relasi antara fluktuasi (naik atau turunnya) nilai sebuah variabel input dengan fluktuasi nilai sebuah variabel output yang terkait dengan variabel input tersebut. Pengambilan pengetahuan tersebut berdasarkan pernyataan-pernyataan sebagai berikut :

Pernyataan 1 : IF x_i is increasing AND y_i is increasing then $R(x_i, y_i) = 1$

Pernyataan 2 : IF x_i is increasing AND y_i is decreasing then $R(x_i, y_i) = -1$

Pernyataan 3 : IF x_i is decreasing AND y_i is increasing then $R(x_i, y_i) = -1$

Pernyataan 4 : IF x_i is decreasing AND y_i is decreasing then $R(x_i, y_i) = 1$

Setiap variabel x_i akan memiliki nilai $R(x_i, y_i)$ masing-masing dan akan digunakan dalam perhitungan va_p . Kemudian untuk perhitungan nilai bobot kriteria pengetahuan ahli $VA_k(x_i, y_i)$ dilakukan dengan persamaan:

$$VA_k(x_k, y_j) = \frac{1}{t-1} \sum_{p=1}^{t-1} va_p \quad (1)$$

Dengan va_p adalah bobot kriteria pengetahuan ahli untuk setiap selang t pada *output space*. Kemudian untuk kriteria sensitivitas data berbasis jarak dihitung dengan menggunakan persamaan (5) dan (6).

$$T_k = \sum_{i \neq l} \frac{d(y_{ij}, y_{lj})}{d'_k(X_i, X_l)} \quad (2)$$

$$S_k = \frac{(\max_{k \in \{1, \dots, n\}} (T_k)) - T_k}{\max_{k \in \{1, \dots, n\}} (T_k) - \min_{k \in \{1, \dots, n\}} (T_k)} \quad (3)$$

Kemudian kedua kriteria tersebut digabungkan dengan persamaan:

$$F_k = g_1 \cdot VA_k(x_k, y_j) + g_2 \cdot S_k, [k \in (1, \dots, n), j \in 1, \dots, m) \quad (4)$$

Dengan F_k adalah nilai kriteria hasil seleksi variabel input untuk variabel x_k , sedangkan g_1 dan g_2 merupakan dua koefisien positif yang dihitung berdasarkan nilai kriteria VA_k (untuk g_1) dan S_k (untuk g_2).

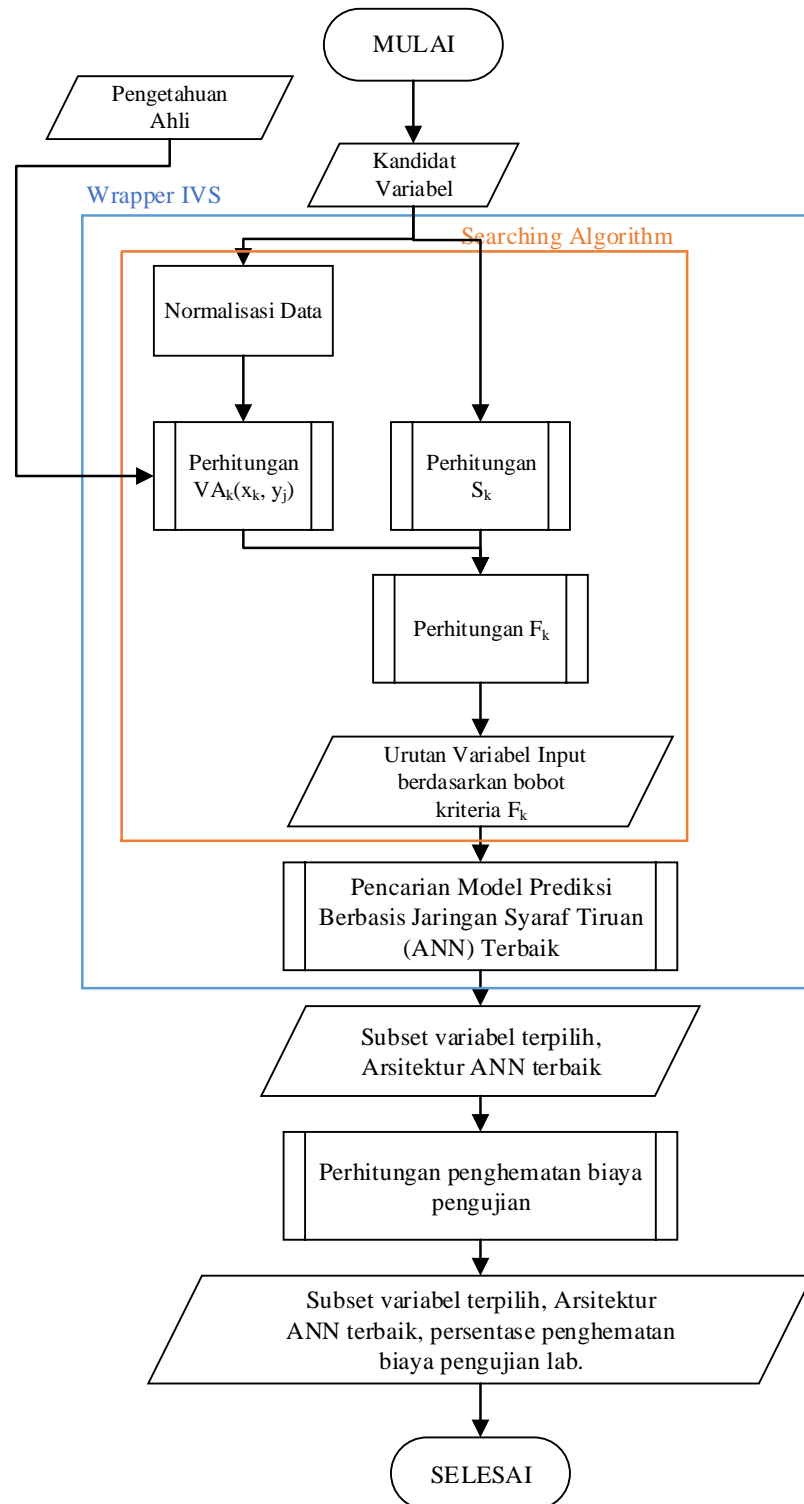
Secara umum sistem IVS yang dikembangkan digambarkan dengan rancangan seperti disajikan pada diagram rancangan sistem seleksi variabel input pada **Gambar 1**.

2.2. Proses Pencarian Model Prediksi Terbaik

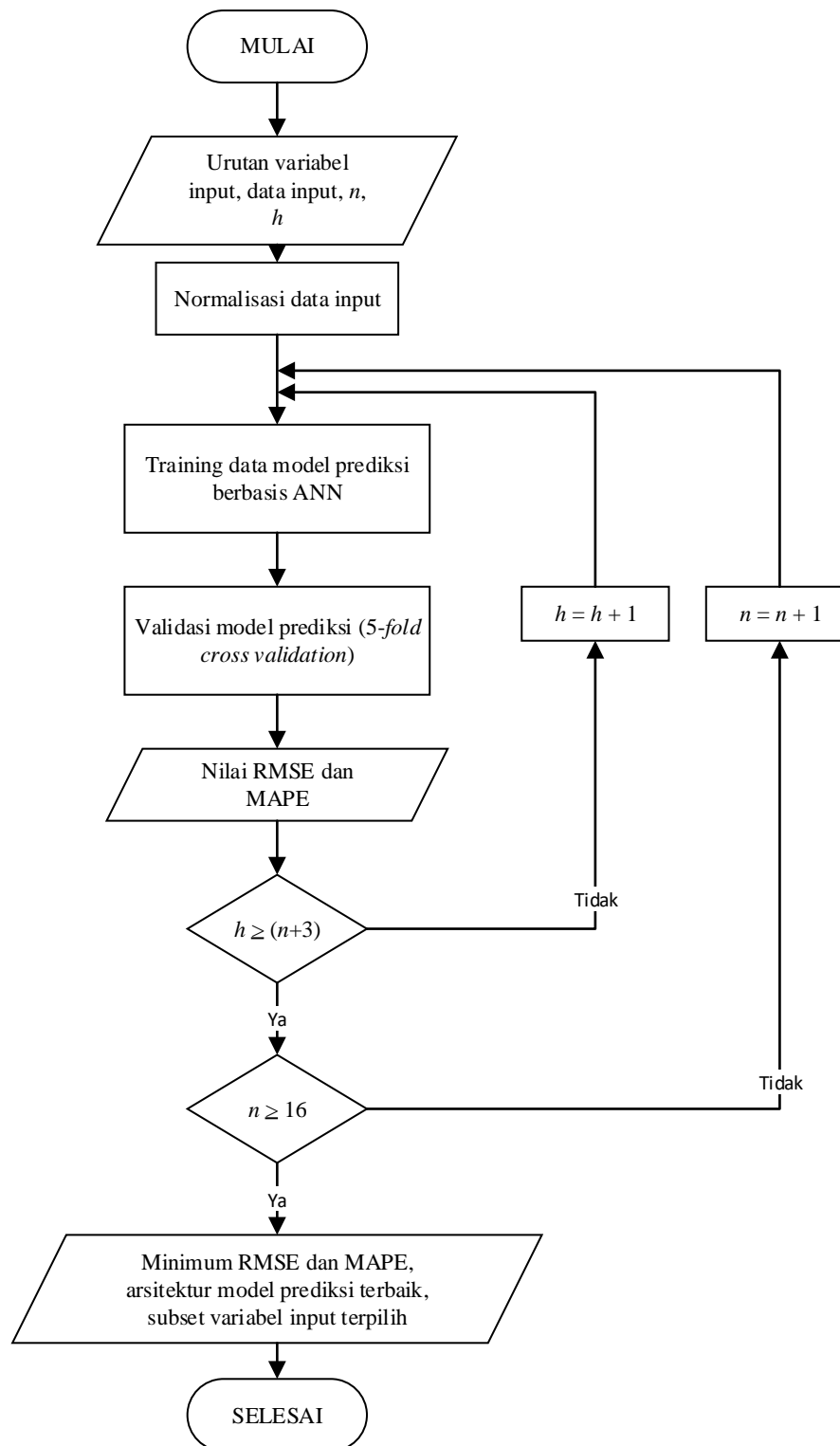
Proses kedua yaitu pencarian model prediksi berbasis jaringan syaraf tiruan yang terbaik dengan ukuran performa berupa nilai RMSE dan MAPE. Jaringan syaraf tiruan yang digunakan menggunakan model feedforward dengan metode pembelajaran backpropagation. Model feedforward dipilih karena dengan model ini training data relatif lebih mudah dilakukan dan arsitektur jaringan syaraf tiruan yang digunakan tidak terlampau rumit sementara backpropagation digunakan sebagai metode pembelajaran karena metode ini terbukti cepat mencapai konvergensi dengan tidak banyak parameter yang harus diatur untuk mencapai model terbaik [10]. Model prediksi dengan nilai RMSE dan MAPE terendah akan dipilih menjadi model prediksi terbaik. Proses pengujian mengujikan berbagai jumlah subset variabel input dimulai 5 sampai dengan 16 variabel input sesuai dengan urutan yang dihasilkan dari proses sebelumnya. Proses pencarian model prediksi terbaik digambarkan pada diagram alir pencarian model prediksi terbaik pada **Gambar 2**.

2.3. Proses Perhitungan Persentase Penghematan Biaya Pengujian Laboratorium

Proses terakhir adalah proses perhitungan persentase penghematan biaya pengujian laboratorium. Hasil dari proses sebelumnya adalah subset variabel input yang memiliki performa prediksi terbaik (nilai RMSE dan MAPE terendah). Setiap variabel input memiliki biaya pengujian laboratorium seperti ditampilkan pada tabel daftar biaya pengujian laboratorium yang disajikan pada **Tabel 1**. Dengan menggunakan acuan biaya pada tabel daftar biaya pengujian laboratorium tersebut maka besar persentase penghematan biaya pengujian laboratorium dapat dihitung dengan menggunakan persamaan persentase penghematan biaya pengujian laboratorium.



Gambar 1. Perancangan sistem seleksi variabel input



Gambar 2. Pencarian Model Prediksi Terbaik

Tabel 1. Daftar Biaya Pengujian Laboratorium untuk tiap Variabel Input

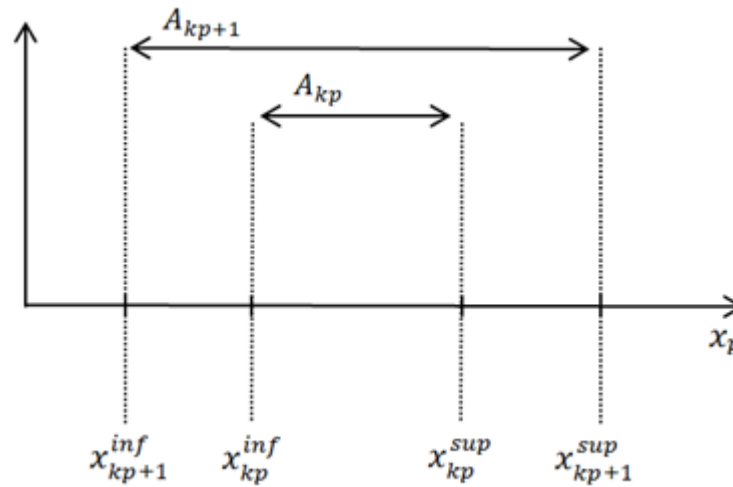
No	Nama Variabel Input	Biaya Pengujian (per sampel)
1	pH H ₂ O	Rp. 30.000,-
2	C-Org	Rp. 40.000,-
3	N Tanah	Rp. 45.000,-
4	Ca-dd	Rp. 105.000,-
5	Mg-dd	Rp. 105.000,-
6	KTK	Rp. 95.000,-
7	Potensial P	Rp. 31.500,-
8	Potensial K	Rp. 31.500,-
9	Tekstur Pasir	Rp. 15.000,-
10	Tekstur Debu	Rp. 15.000,-
11	Tekstur Tanah Liat	Rp. 15.000,-
12	N daun	Rp. 50.000,-
13	P daun	Rp. 50.000,-
14	K daun	Rp. 50.000,-
15	Ca Daun	Rp. 50.000,-
16	Mg daun	Rp. 50.000,-
17	Pupuk N	Rp. 0,-
18	Pupuk P	Rp. 0,-
19	Pupuk K	Rp. 0,-
Total Biaya Uji Per Sampel Lengkap		Rp. 778.000,-

2.4. Perbaikan pada metode IVS CZCL

Pada metode CZCL yang asli belum terdapat definisi persamaan nilai va_p ketika dua kondisi $x_{kp+1}^{inf} \leq x_{kp}^{inf}$ dan $x_{kp+1}^{sup} \geq x_{kp}^{sup}$ terpenuhi secara bersamaan. Kondisi ini mungkin terjadi ketika seluruh A_{kp} merupakan himpunan bagian dari A_{kp+1} , sesuai dengan ilustrasi **Gambar 3**.

Pada penelitian ini diusulkan penambahan sebuah definisi dari va_p ketika kondisi tersebut dipenuhi secara bersamaan dengan cara menambahkan kedua nilai va_p untuk masing-masing kondisi. Persamaan yang dihasilkan disajikan pada persamaan (5).

$$\begin{aligned}
 va_p &= \left(\frac{1}{2} |R(x_k, y_j)| \times (1 + R(x_k, y_j)) \times \left(1 - \frac{|I_{kp}|}{|U_{kp}|} \right) \right) \\
 &+ \left(\frac{1}{2} |R(x_k, y_j)| \times (1 - R(x_k, y_j)) \times \left(1 - \frac{|I_{kp}|}{|U_{kp}|} \right) \right) \\
 &= \left(2 \times |R(x_k, y_j)| \times \left(1 - \frac{1 \times |I_{kp}|}{|U_{kp}|} \right) \right), \text{ jika } x_{kp+1}^{inf} \leq x_{kp}^{inf} \text{ dan } x_{kp+1}^{sup} \geq x_{kp}^{sup} \quad (5)
 \end{aligned}$$



Gambar 3. Ilustrasi kondisi $x_{kp+1}^{inf} \leq x_{kp}^{inf}$ dan $x_{kp+1}^{sup} \geq x_{kp}^{sup}$

Kemudian berdasarkan arahan dari ahli kelapa sawit, akan diimplementasikan penalti terhadap nilai-nilai hasil pengujian yang melebihi batas normal. Penalti ini hanya berlaku untuk kandungan zat kimia dalam tanah dan daun. Bentuk penalti yang diberikan adalah perubahan pada nilai korelasi di Pernyataan 1-4. Pada kondisi penalti akan diimplementasikan Pernyataan 1-4 yang dimodifikasi menjadi sebagai berikut.

Pernyataan 1P : IF x_i is increasing AND y_i is increasing AND $x_i > maxValue_i$ THEN
 $R(x_i, y_i) = 0.5$

Pernyataan 2P : IF x_i is increasing AND y_i is decreasing AND $x_i > maxValue_i$ THEN
 $R(x_i, y_i) = -0.5$

Pernyataan 3P : IF x_i is decreasing AND y_i is increasing AND $x_i < minValue_i$ THEN
 $R(x_i, y_i) = -0.5$

Pernyataan 4P : IF x_i is decreasing AND y_i is decreasing AND $x_i < minValue_i$ THEN
 $R(x_i, y_i) = 0.5$

Dengan $maxValue_i$ adalah batas atas normal dari sebuah variabel i dan $minValue_i$ adalah batas bawah normal dari sebuah variabel i . Kedua nilai batas ini ditentukan oleh pengetahuan dari ahli kelapa sawit.

3. Hasil dan Pembahasan

Pengujian dilakukan dengan menggunakan data input sebanyak 200 data sampel dengan 19 variabel input dan 1 variabel output. Pengujian dilakukan dengan membangun model prediksi dan menghitung nilai RMSE dan MAPE untuk masing-masing model prediksi. Model prediksi dengan nilai RMSE dan MAPE terendah akan dipilih menjadi model terbaik dan kemudian akan dihitung nilai persentase penghematan biaya pengujian laboratoriumnya. Sebagai nilai tolok ukur performa prediksi dari model yang dibuat, dilakukan membangun model prediksi tanpa

melakukan seleksi variabel input (19 variabel digunakan), didapatkan nilai minimum RMSE sebesar 0,709 (MAPE minimum 3,72 %) dan arsitektur jaringan syaraf tiruan 19-13-1.

3.1. Pengujian IVS CZCL dengan kriteria F_k

Untuk pengujian pertama dengan menggunakan seleksi variabel input yang diusulkan (kriteria F_k) didapatkan model-model prediksi dengan nilai RMSE seperti disajikan pada tabel hasil pengujian dengan kriteria F_k pada **Tabel 2**. Dari **Tabel 2** dapat dilihat bahwa nilai RMSE minimum dicapai pada arsitektur jaringan syaraf tiruan 15-1-1 dengan nilai RMSE sebesar 1,372 (MAPE 7,621 %).

Sebagai pembandingan metode yang diusulkan dilakukan pengujian dengan menggunakan metode seleksi variabel input menggunakan nilai *variable importance* pada metode *random forest*. *Random forest* merupakan metode *ensemble learning* yang menggunakan sejumlah *decision tree* yang digunakan untuk membangun model klasifikasi atau regresi [14], [15]. *Random forest* memiliki nilai *variable importance* yang menandakan tingkat pengaruh suatu variabel input terhadap variabel output [16]. Hasil pengujian pencarian model prediksi terbaik dengan menggunakan seleksi variabel input *variable importance* pada *random forest* disampaikan pada tabel hasil pengujian pada **Tabel 3**.

Tabel 2. Hasil Pengujian dengan kriteria F_k

Jumlah Neuron Hidden Layer	Jumlah Variabel Input (Neuron Pada Input Layer)											
	5	6	7	8	9	10	11	12	13	14	15	16
1	4.355	6.523	5.443	6.207	5.133	6.342	6.601	6.109	3.518	1.577	1.372	1.500
2	4.044	7.050	5.872	6.568	5.164	6.440	6.118	6.435	3.129	2.051	1.421	2.007
3	4.465	6.281	5.668	6.121	5.254	5.289	6.457	6.047	3.504	2.837	2.093	1.990
4	5.938	6.939	6.935	5.704	5.079	5.366	7.130	6.622	3.009	2.579	1.590	1.739
5	5.730	7.000	7.263	6.671	5.432	5.536	7.253	6.945	2.147	2.412	1.896	1.807
6	5.140	7.173	7.028	6.440	5.709	7.009	6.946	6.139	3.772	2.060	2.152	2.234
7	5.756	6.336	7.566	6.241	5.360	5.919	7.358	6.015	3.590	2.018	1.999	2.168
8	5.630	7.170	7.038	6.701	5.588	5.349	6.843	6.773	3.998	2.598	1.848	2.252
9		7.195	6.951	6.874	5.343	5.375	6.349	6.146	3.265	2.219	1.853	2.059
10			7.908	6.913	5.588	6.178	8.077	6.428	4.822	2.390	2.281	2.184
11				5.891	5.232	5.444	8.064	5.851	2.626	2.742	2.130	1.940
12					5.855	8.273	7.349	8.398	3.235	4.609	2.783	2.138
13						5.973	7.084	7.328	2.715	6.592	2.697	3.485
14							6.997	6.931	2.470	2.412	2.498	2.353
15								5.349	2.561	2.757	3.200	3.091
16									3.917	2.254	2.171	4.199
17										2.626	5.878	3.155
18											2.373	3.003
19												2.549

Tabel 3. Hasil Pengujian dengan *variable importance* pada *random forest*

Jumlah Neuron Hidden Layer	Jumlah Variabel Input (Neuron Pada Input Layer)											
	5	6	7	8	9	10	11	12	13	14	15	16
1	2.218	3.166	4.696	5.183	4.621	4.523	1.841	1.637	2.776	1.313	1.324	1.956
2	2.117	3.663	4.908	4.53	4.367	4.735	2.741	2.198	2.385	1.876	1.753	2.337
3	2.725	3.616	4.772	4.512	4.676	4.834	3.559	2.288	2.283	1.623	2.204	2.844
4	3.44	4.436	4.946	4.54	4.97	5.324	2.57	4.941	1.841	2.121	2.116	2.599
5	3.706	4.757	4.901	4.7	4.92	5.566	3.522	3.165	1.841	1.505	1.936	3.458
6	2.439	4.551	4.453	4.595	3.801	5.061	2.939	2.556	2.603	2.094	3.338	4
7	2.773	4.052	4.894	5.023	5.089	4.394	3.453	2.754	2.293	1.65	2.066	1.936
8	3.815	4.993	5.423	5.396	5.658	5.592	4.091	2.615	1.906	2.976	2.13	3.146
9		5.012	4.677	5.031	3.832	4.751	3.421	2.911	3.492	2.218	2.586	2.432
10			5.043	5.378	4.866	6.427	2.87	3.24	3.263	2.681	1.795	3.463
11				5.223	5.65	5.857	3.399	3.389	3.006	2.662	1.848	2.598
12					5.488	6.309	2.903	3.855	3.36	4.548	3.38	2.98
13						4.663	3.213	2.666	1.96	6.098	2.472	3.667
14							3.699	2.434	2.943	2.429	1.75	3.129
15								2.548	3.677	3.498	3.453	2.208
16									2.772	2.721	2.385	3.072
17										2.776	6.139	4.048
18											2.407	2.901
19												4.441

Hasil pengujian dengan menggunakan *variable importance* pada *random forest* menunjukkan model prediksi terbaik dicapai pada arsitektur 14-1-1 dengan nilai RMSE 1,313 (MAPE 8,185 %). Kedua metode seleksi variabel input menghasilkan nilai akurasi prediksi yang hampir sama namun masih belum mampu mendekati performa prediksi model prediksi tanpa seleksi variabel input.

3.2. Pengujian IVS CZCL dengan hanya menggunakan kriteria VA_k

Sebagai pembandingan lain digunakan juga kriteria VA_k sebagai dasar metode seleksi variabel input. Hasil pengujian tambahan disajikan pada tabel hasil pengujian dengan kriteria VA_k pada Tabel 4.

Tabel 4. Hasil Pengujian dengan Kriteria VA_k

Jumlah Neuron Hidden	Jumlah Variabel Input											
	5	6	7	8	9	10	11	12	13	14	15	16
1	3.247	3.246	2.811	1.2	1.702	2.675	2.019	1.351	2.456	1.613	2.147	1.455
2	3.811	4.127	3.775	1.291	1.7	2.737	2.377	1.537	2.098	1.888	2.288	1.559
3	4.091	5.075	4.343	0.992	2.126	3.216	2.692	1.789	2.2	1.817	3.091	2.026
4	4.454	4.458	4.579	1.496	2.423	2.772	3.368	2.224	2.666	3.15	4.278	1.938
5	4.295	4.718	4.593	1.542	2.726	3.269	4.362	2.218	2.334	3.048	2.744	1.832
6	4.658	5.754	4.801	1.446	3.144	2.967	3.712	3.636	2.814	4.028	2.824	4.482
7	4.369	5.95	5.475	1.511	2.374	3.224	2.837	2.461	2.746	2.894	3.75	2.256
8	5.154	5.28	5.068	1.587	3.059	3.205	3.193	2.545	2.903	3.479	2.614	1.782
9		6.172	5.869	1.439	3.179	3.496	2.784	2.462	1.893	3.6	3.206	2.388
10			6.638	1.703	3.021	2.955	2.97	3.219	2.594	2.477	4.006	2.715
11				1.527	2.607	3.618	3.223	2.72	2.65	2.984	2.645	2.072

Jumlah Neuron Hidden	Jumlah Variabel Input											
	5	6	7	8	9	10	11	12	13	14	15	16
12					2.209	7.121	3.766	3.564	2.711	2.414	2.881	2.032
13						3.311	3.703	2.658	2.863	3.031	3.149	2.277
14							3.085	2.718	3.314	3.042	3.594	3.743
15								2.597	2.51	3.93	4.323	2.758
16									4.411	3.025	2.643	2.793
17										3.134	3.13	2.033
18											3.74	2.149
19												4.572

Pengujian kriteria VA_k saja menghasilkan model prediksi terbaik dengan 8 variabel input, nilai RMSE sebesar 0,992 (MAPE 4,8 %) dicapai dengan arsitektur jaringan syaraf tiruan 8-3-1. Hasil ini merupakan hasil yang paling mendekati performa prediksi dari model prediksi tanpa seleksi variabel input.

3.3. Pengujian IVS CZCL dengan Penambahan Penalti pada $R(x_i, y_i)$

Kemudian pengujian penambahan penalti pada penentuan nilai $R(x_i, y_i)$ yang akan berpengaruh pada VA_k dan F_k menghasilkan perbandingan pada **Tabel 5**.

Tabel 5. Perbandingan hasil IVS CZCL dengan dan tanpa penalti pada nilai $R(x_i, y_i)$

Ars. F_k	Tanpa Penalti				Dengan Penalti			
	RMSE F_k	Ars. VA_k	RMSE VA_k	Ars. F_k	RMSE F_k	Ars. VA_k	RMSE VA_k	
15-1-1	1,372	8-3-1	0,992	14-4-1	1,177	7-4-1	0,877	

3.4. Perhitungan persentase penghematan biaya pengujian laboratorium

Setelah masing-masing variabel terpilih dan model prediksi terbaik diketahui maka akan dilakukan perhitungan persentase penghematan biaya pengujian laboratorium dengan menggunakan persamaan (3). Hasil dari perhitungan persentase penghematan biaya pengujian laboratorium untuk keenam model prediksi terbaik tersebut ditampilkan pada Tabel 6.

Tabel 6. Persentase Penghematan Biaya Pengujian Laboratorium

Nama Kriteria	Jumlah Variabel Terpilih (variabel)	Persentase Penghematan Biaya Pengujian (%)
Tanpa IVS	19	0
F_k	15	39,85
Random forest	14	37,02
VA_k	8	72,36
F_k dengan penalti	15	39,85
VA_k dengan penalti	7	81,47

Hasil perhitungan menunjukkan bahwa metode seleksi variabel input yang menghasilkan persentase penghematan terbesar adalah seleksi variabel input dengan menggunakan kriteria VA_k dengan penalti iatau kriteria pengetahuan ahli dengan tambahan penggunaan penalti dengan 7 variabel input terpilih dan penghematan biaya pengujian laboratorium mencapai 81,47 %.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian IVS CZCL dan perbandingannya dengan metode IVS CZCL yang dimodifikasi dan metode pembandingan (*random forest*), sistem IVS yang dikembangkan telah dapat melakukan seleksi variabel input dengan metode yang diajukan dengan hasil akurasi model prediksi yang lebih baik dibanding hasil seleksi variabel input dengan metode *random forest*.

Selain itu pengujian tambahan menunjukkan bahwa untuk data variabel input yang dimiliki, kriteria pengetahuan ahli (VA_k) dengan penalti menghasilkan hasil variabel terseleksi lebih sedikit (7 variabel) dengan RMSE sebesar 0.877 (MAPE 3.8%) dan penghematan biaya pengujian laboratorium yang jauh lebih besar yaitu sebesar 81,47 %. Hal ini menunjukkan bahwa dengan memanfaatkan IVS, perkebunan rakyat mampu mengurangi jumlah variabel yang diuji di laboratorium tanpa mengurangi hasil analisis dan prediksi produktivitas perkebunan kelapa sawit yang mereka miliki.

Daftar Pustaka

- [1] Ditjenbun, “Statistik Perkebunan Unggulan Nasional 2019-2021,” *Direktorat Jendral Perkeb. Kementerian. Pertan. Republik Indones.*, pp. 1–88, 2021, [Online]. Available: <https://ditjenbun.pertanian.go.id/template/uploads/2021/04/BUKU-STATISTIK-PERKEBUNAN-2019-2021-OK.pdf>.
- [2] A. Harjoko and U. G. Mada, “Pemrosesan Citra Digital untuk Klasifikasi Mutu Buah Pisang Menggunakan Jaringan Saraf Tiruan,” *IJEIS (Indonesian J. Electron. Instrum. Syst.*, vol. 4, no. 1, pp. 57–68, Apr. 2014, doi: 10.22146/IJEIS.4222.
- [3] H. Herman and A. Harjoko, “Pengenalan Spesies Gulma Berdasarkan Bentuk dan Tekstur Daun Menggunakan Jaringan Syaraf Tiruan,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 9, no. 2, pp. 207–218, Jul. 2015, doi: 10.22146/IJCCS.7549.
- [4] Hermantoro, A. P. Suryotomo, A. I. Uktoro, and R. A. Renjani, “Unmanned Aerial Vehicle Application for Plantation Mapping and Automatic Oil Palm Trees Counting on Oil Palm Plantation Management,” in *International Conference on the Role of Agricultural Engineering for Sustainable Agriculture Production*, 2016, no. December, pp. 47–50.
- [5] H. Hermantoro, “PEMODELAN DAN SIMULASI PRODUKTIVITAS PERKEBUNAN KELAPA SAWIT BERDASARKAN KUALITAS LAHAN DAN IKLIM MENGGUNAKAN JARINGAN SYARAF TIRUAN MODELING AND SIMULATION OF PALM OIL PLANTATION PRODUCTIVITY BASED ON LAND QUALITY AND CLIMATE USING ...,” *Agromet*, vol. 23, no. 1, p. 45, Jun. 2009, doi: 10.29244/j.agromet.23.1.45-51.
- [6] Z. Ismail and A. Khamis, “Neural Network in Modeling Malaysian Oil Palm Yield,” *Am. J. Appl. Sci.*, vol. 8, no. 8, pp. 796–803, 2011.
- [7] Y. Y. Hilal, W. Wan Ishak, A. Yahya, and Z. H. Asha\’ari, “An Artificial Neural Network with Stepwise Method for Modeling and Simulation of Oil Palm Productivity Based on Various Parameters in Sarawak,” *Res. J. Appl. Sci. Eng. Technol.*, vol. 13, no. 9, pp. 730–

740, Nov. 2016, doi: 10.19026/RJASET.13.3347.

- [8] N. D. Kartika, I. W. Astika, and E. Santosa, "Oil Palm Yield Forecasting Based on Weather Variables Using Artificial Neural Network Sustainable in Tropical Crop Production View project MIT Thesis View project NADIA DWI KARTIKA Oil Palm Yield Forecasting Based on Weather Variables Using Artificial Neural Network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 626–633, 2016, doi: 10.11591/ijeecs.v3.i2.pp626-633.
- [9] S. R. Diana and G. Dharma, "Estimation the Amount of Oil Palm Production Using Artificial Neural Network and NDVI SPOT-6 Imagery," 2019. Accessed: May 30, 2021. [Online]. Available: www.ijisrt.com.
- [10] R. Chapman *et al.*, "Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: A proof of concept analysis," *Comput. Electron. Agric.*, vol. 151, pp. 338–348, Aug. 2018, doi: 10.1016/j.compag.2018.06.006.
- [11] T. Chen, C. Zhang, X. Chen, and L. Li, "An Input Variable Selection Method for the Artificial Neural Network of Shear Stiffness of Worsted Fabrics," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 1, no. 5, pp. 287–295, Apr. 2009, doi: 10.1002/SAM.10020.
- [12] R. A. Collazo, L. A. M. Pessôa, L. Bahiense, B. D. B. Pereira, A. F. Dos Reis, and N. S. E Silva, "A COMPARATIVE STUDY BETWEEN ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE FOR ACUTE CORONARY SYNDROME PROGNOSIS," *Pesqui. Operacional*, vol. 36, no. 2, pp. 321–343, May 2016, doi: 10.1590/0101-7438.2016.036.02.0321.
- [13] S. Galelli, G. B. Humphrey, H. R. Maier, A. Castelletti, G. C. Dandy, and M. S. Gibbs, "An evaluation framework for input variable selection algorithms for environmental data-driven models," *Environ. Model. Softw.*, vol. 62, pp. 33–51, Dec. 2014, doi: 10.1016/J.ENVSOFT.2014.08.015.
- [14] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," *Ensemble Mach. Learn.*, pp. 157–175, 2012, doi: 10.1007/978-1-4419-9326-7_5.
- [15] D. Vigneswari, N. K. Kumar, V. Ganesh Raj, A. Gugan, and S. R. Vikash, "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus," in *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, Mar. 2019, pp. 84–87, doi: 10.1109/ICACCS.2019.8728388.
- [16] Y. Liu, Y. Wang, and J. Zhang, "New machine learning algorithm: Random forest," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7473 LNCS, pp. 246–252, 2012, doi: 10.1007/978-3-642-34062-8_32/COVER.