Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 16 *ISSN (Printed) : 2579-7271*
Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau *ISSN (Online) : 2579-5406*
Pekanbaru, 23 Oktober 2024

Metode *Robust K-Fold Cross Validation* dengan *Partial Least Square Regression* pada Data *Near Infrared Spectroscopy*

Nuraini Sibuea[*1], Syamsudhuha[2], Arisman Adnan[3]
[1,2,3] Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Email: [1]nuraini.sibuea6881@grad.unri.ac.id, [2]syamsudhuha@lecturer.unri.ac.id,
[3]arisman.adnan@lecturer.unri.ac.id

Abstrak

Penelitian ini mengevaluasi performa model *Partial Least Square Regression* (*PLSR*) dalam kondisi data dengan dan tanpa *outlier*. Penanganan data yang mengandung *outlier* digunakan metode *k-fold cross validation* yang diaplikasikan pada data *Near Infrared Spectroscopy* (*NIRS*) tanah perkebunan kelapa sawit terhadap nitrogen (N) tanah. Sebelum pengolahan data dilakukan terlebih dahulu *pretreatment data* untuk menghilangkan efek hamburan data dengan *Standarlized Normal Variate* (*SNV*). Identifikasi *outlier* dilakukan dengan metode *RBF Kernel PCA* menghasilkan data yang termasuk *outlier* yaitu data ke 7, 8, 92, 93, dan 95. Hasil analisis menunjukkan bahwa keberadaan *outlier* secara signifikan menurunkan performa *PLSR* klasik dengan penurunan nilai R^2 dan peningkatan nilai *RMSE*. Penerapan *k-fold cross validation* pada *PLSR* mampu meningkatkan robustitas model terhadap *outlier* dengan peningkatan nilai R^2 meskipun sedikit peningkatan pada *RMSE*. Disimpulkan bahwa *k-fold cross validation* lebih efektif dalam menangani *dataaset* yang mengandung *outlier* sehingga memberikan prediktabilitas yang lebih stabil dobandingkan *PLSR* klasik.

Kata kunci: *k-fod cross validation, Near Infrared Spectroscopy* (*NIRS*), *Partial Least Square Regression* (*PLSR*), *RBF Kernel PCA*, *Standarlized Normal Variate* (*SNV*).

Abstract

This study evaluates the performance of the Partial Least Square Regression (*PLSR*) *model under conditions with and without outliers. To handle data containing outliers, the k-fold cross validation method was applied to Near Infrared Spectroscopy* (*NIRS*) *data of oil palm plantation soil in relation to soil nitrogen* (*N*). *Before data processing, a data pretreatment was conducted to remove scattering effects using Standarlized Normal Variate* (*SNV*). *Outliers were identified using RBF Kernel PCA method, which identified data point 7, 8, 92, 93, and 95. The analysis result indicate that the presence of outliers significantly degrades the performance of classical PLSR, as evidenced by a decrease in* R^2 *and an increase in RMSE. The application of k-fold cross validation to PLSR was able to enhance the robustness of the model against outliers, resulting in an increase in* R^2 *despite a slight increase in RMSE. It is concluded that k-fold cross validation is more effective in handling dataset containing outliers, providing more stable predictability compared to classical PLSR.*

Keywords*: K-Fold Cross Validation, Near Infrared Spectroscopy* (*NIRS*), *Partial Least Square Regression* (*PLSR*), *RBF Kernel PCA*, *Standarlized Normal Variate* (*SNV*).

1. Pendahuluan

Metode *Partial Least Square Regression* (*PLSR*) adalah teknik analisis data statistik multivariat untuk menangani ruang data berdimensi tinggi pada kumpulan data. *PLSR* berguna untuk mengatasi masalah multikolineritas di ruang data tersebut. Metode ini kurang unggul dalam hal ketahanan terhadap data yang mengandung *outlier*. Penerapan *k-fold cross validation* sangat penting dalam proses evaluasi model agar tetap *robust* meskipun terdapat *outlier* dalam data [1]. Beberapa penelitian, menggunakan *K-fold* dalam penyelesaian *cross validation* [2]. *K-fold cross validation* merupakan bagian metode dari *cross validation* di mana mengulang setiap proses sebanyak *k* kali. Setiap *fold*, *dataset* dibagi menjadi *k* bagian untuk validasi dan *k* – 1 bagian menjadi *subset* pelatihan untuk evaluasi model [3]. Metode ini digunakan dalam data *Near Infrared Spectroscopy* (*NIRS*). *NIRS* adalah metode analisis yang menggunakan infamerah dekat dari spektrum elektromagnetik. Data spektal sebagai hasil dari penyerapan cahaya pada setiap pita panjang gelombang [4]. Spektral menggunakan panjang gelombang antara 350 – 2500 nm [5]. Teknik analisis *NIRS* dapat diaplikasikan dalam industri perkebunan kelapa sawit.

Perkebunan kelapa sawit telah berhasil memperkuat sistem ekonomi di Indonesia [6]. Pengelolaan lahan yang tepat merupakan faktor utama untuk meningkatkan kualitas tanaman dan produksi minyak kelapa sawit. Hal ini dapat diperoleh dengan pemberian pupuk yang memiliki unsur hara yang tinggi. Kelapa sawit membutuhkan unsur hara salah satunya nitrogen (N) untuk meningkatkan nutrisi tanah [7].

Beberapa penelitian sebelumnya, Silalahi, *et al* [8] meneliti tentang data *NIRS* buah kelapa sawit dengan metode *Robust Reliable Weighted Average – Partial Least Square* (*RRWA – PLS*). Metode ini menunjukkan ketahanan yang baik terhadap data yang mengandung *outlier* dengan R^2 sebesar 0.661 (%*ODM*), 0.718 (%*OWM*), 0.747 (%*FFA*) dan *RMSEP* sebesar 3.071 (%*ODM*), 4.185 (%*OWM*), 0.275 (%FFA). Kim, *et al* [9] meneliti tentang data *Visible and Near Infrared* (*VNIR*) pada *Souble Solid Content* (*SSC*) buah jeruk dengan prapemprosesan pemilihan panjang gelombang menggunakan *Competitive Adaptive Reweighted Sampling* (*CARS*). Model prediksi *SSC CARS-PLSR* dengan *outlier* yang dihilangkan menunjukkan nilai R^2 sebesar 0.75 dan *RMSE* sebesar 0.56. Amankwaah, *et al* [10] meneliti tentang pengembangan kurva kalibrasi menggunakan *NIRS* untuk memprediksi kandungan gula pada ubi jalar panggang. Metode yang digunakan adalah *PLSR* yang bertujuan untuk membangun hubungan antara data spektral dengan konsentrasi gula dalam ubi jalar. Hasil penelitian menunjukkan bahwa model *NIRS* yang dikembangkan memiliki tingkat akurasi yang tinggi dalam memprediksi kandungan gula berdasarkan *cross validation*.

Berdasarkan penelitian sebelumnya, peneliti tertarik mengkombinasikan *k-fold cross validation* dengan *PLSR* pada data yang mengandung *outlier*. Kemudian dilihat apakah *robust* atau tidak terhadap adanya *outlier*. Metode ini diaplikasikan pada data *NIRS* tanah perkebunan kelapa sawit yang mengandung unsur hara nitrogen (N). Kemudian evaluasi data dengan menentukan nilai R^2 dan *RMSE* pada setiap model.

2. Metode Penelitian

Penelitian ini menggunakan data skunder yaitu data *NIRS* tanah kelapa sawit dengan panjang gelombang 350 – 2500 nm (variabel ***X***) dan data nitrogen tanah (variabel ***y***). Data *NIRS* tanah perkebunan kelapa sawit yang diolah menggunakan *k-fold cross validation* pada *PLSR* akan tetapi dilakukan *pretreatment* data terlebih dahulu. Kemudian dilihat apakah *robust* terhadap adanya data *outlier* dengan evaluasi model menggunakan R^2 dan *RMSE* pada setiap model.

2.1. *Standarlized Normal Variate* (*SNV*)

Standarlized Normal Variate (*SNV*) digunakan untuk mengatasi intervensi multiplikatif dan aditif dari efek hamburan dan variabelitas untuk partikel dalam spektrum mentah. Koreksi hamburan dengan menstandarkan spektrum menggunakan rata-rata dan standar deviasinya [11]. Misalkan $\boldsymbol{x}_i$ dengan elemen $\boldsymbol{x}_{ij}$ untuk $i, j = 1, 2, \dots, m$. Secara sistematis dinyatakan sebagai berikut

$$\tilde{X}_{ij} = \frac{X_{ij} - mean(X_{ij})}{std(X_{ij})} \quad (1)$$

2.2. *Partial Least Square Regression* (*PLSR*)

Model *PLSR* adalah prosedur iteratif dari metode statistik multivariat. Metode ini digunakan untuk menngubah untuk mengubah *m* variabel prediktor (***X***) yang memiliki masalah multikolineritas menjadi *l* variabel baru tidak berkorelasi disebut komponen. Misalkan persamaan regresi multivariat terdiri variabel prediktor (***X***) dan variabel ***y*** ditulis dalam matriks sebagai berikut

$$y = Xb + e \quad (2)$$

y, ***e*** adalah vektor $n \times 1$, ***X*** adalah matriks $n\ x\ m$ dan **b** adalah vektor $m \times 1$. Umumnya solusi *estimator* ***b*** menggunakan *least square* yaitu

$$\hat{b} = (X^T X)^{-1} X^T y \quad (3)$$

Karena *dataset* mengandung dimensi tinggi yang dengan *m* prediktor maka akan ada jumlah solusi banyak untuk *estimator* ***b***. Jika $X^T X$ adalah *singular* maka tidak memenuhi teorema *fundamental* mengenai *rank* dalam regresi linier klasik. Prosedur khusus untuk kasus ini diperlukan mengekstrak komponen baru dengan memaksimalkan kovarians antara variabel

prediktor (***X***) dan prediktor ***y***. Persamaan (2) digunakan untuk menginisialisasi vektor skor awal *u* dari variabel *y* tunggal. Terdapat hubungan lular untuk prediktor ***X*** yang didefinisikan sebagai berikut:

$$X = VP^T + E \tag{4}$$

P : matriks berukuran $m \times l$ yang terdiri dari vektor *loading* yaitu $\left\{P_g = \frac{(X^T v_g)}{v_g{}^T v_g}\right\}_{g=1}^{l}$

v_g : vektor kolom $n \times 1$ dari skor x_j dalam ***X*** yaitu $\left\{v_g = \frac{(Xw_j)}{w_j{}^T w_j}\right\}_{g=1}^{l}$

w_j : vektor $m \times 1$ dari bobot untuk ***X*** yaitu $\left\{w_j = \frac{(X^T u)}{u^T u}\right\}_{j=1}^{m}$

V : matriks $n \times l$ yang terdiri dari vektor *n* x 1 (v_g)

E : matriks $n \times m$ dari residual hubungan luar variabel prediktor ***X***

Mengikuti prosedur yang sama seperti varibel prediktor ***X***, hubungan luar untuk variabel respon ***y*** sebagai berikut

$$y = u\, q^T + f \tag{5}$$

q : vektor *loading* berukuran $l \times 1$ yaitu $\left\{q_g = \frac{(y^T v_g)}{v_g{}^T v_g}\right\}_{g=1}^{l}$

u : matriks $n \times l$ dari skor blok ***y***

f : vektor $n \times 1$ dari residual variabel respon ***y***

u disebut juga sebagai hubungan linier dalam antara skor blok ***X*** dan skor blok ***y*** yang dihitung sebagai $\left\{u = b_g v_g \; dengan \; b_g = \frac{u^T v_g}{v_g{}^T v_g}\right\}_{g=1}^{l}$ atau ditulis

$$u = Vb_{inner} + g \tag{6}$$

b_{inner} : vektor $l \times 1$ dari koefisien regresi

g : vektor $n \times 1$ dari residual hubungan dalam

Berdasarkan algoritma *Nonlinear Iterative Partial Least Square* (*NIPALS*), hubungan campuran dalam model *PLSR* dapat didefinisikan sebagai berikut

$$y = Xb_{PLSR} + f \tag{7}$$

di mana $b_{PLSR} = W(P^T W)^{-1} a$ adalah vektor koefisien berukuran $m \times 1$, a mewakili vektor koefisien berukuran $l \times 1$ dihitung sebagai $a = V^T y$ dan f menyatakan vektor residual berukuran $n \times 1$ dalam hubungan campuran yang harus diminimalkan. Estimator untuk parameter b_{PLSR} sebagai berikut

$$\hat{b}_{PLSR} = X^T u\,(V^T X X^T u)^{-1} V^T y\,, \hat{b}_{PLSR} \in R^{m\,x\,1}$$

$\hat{b}_{PLSR}$ dinotasikan dengan koefisien regresi dengan matriks berdimensi *m* di *PLSR* [8].

2.3 *K-Fold Cross Validation*

Cross validation untuk meminimumkan masalah dengan *K-fold* yaitu menggunakan bagian data yang tersedia agar sesuai dengan model dan bagian berbeda untuk mengujinya. Misalkan *K* = 5, seperti gambar dibawah ini

ISSN (Printed) : 2579-7271
ISSN (Online) : 2579-5406

Gambar 1. *K-Fold Cross Validation*

Bagian ke-*k* (kolom ketiga) merupakan validasi data dan lainnya adalah data *train* untuk memprediksi suatu model kemudian menghitung kesalahan prediksinya. Ini dilakukan sebanyak *K-fold* dan menggabungkan sebanyak *K* estimasi kesalahan prediksi. Berikan $k: \{1, \dots, N\} \rightarrow \{1, \dots, K\}$ menjadi fungsi pengindeksan yang menunjukkan bagian mana pengamatan *i* dialokasikan dengan secara acak [12]. $\hat{f}^{-k}(x)$ merupakan model prediksi yang dilatih menggunakan seluruh data kecuali bagian ke-*k* sehingga estimasi kesalahan prediksi *cross validation* adalah

$$CV(\hat{f}) = \frac{1}{N} \sum_{i=1}^{N} \quad L\left(y_i, \hat{f}^{-k(i)}(x_i)\right) \tag{8}$$

2.4 Prosedur *K-Fold Cross Validation* dengan *PLSR*

Adapun prosedur *K-Fold Cross Validation* dengan *PLSR* sebagai berikut:

1. Membagi *dataset* menjadi *K subset* yang berukuran sama. Misalkan *dataset* yaitu *D* dibagi menjadi D_1, D_2, ..., D_K
2. Melakukan iterasi setiap *fold i* (1, ..., *K*) dengan langkah berikut
   - Menggunakan D_i sebagai data *test* dan selainnya merupakan data *train*.
   - Melatih model *PLSR* menggunakan data *train*
   - Memprediksi nilai data *test* dari model *PLSR*
   - Mengevaluasi model dengan menghitung nilai *Coefficient of Determination* (R^2) dan *Root Mean Square Error* (*RMSE*) untuk *fold* ke-*i*.
   - Ulangi langkah tersebut untuk setiap *fold i* (1, ..., *K*)
3. Menghitung rata-rata dari R^2 dan *RMSE* semua *fold* untuk mendapatkan performa model secara keseluruhan.

3. Hasil dan Analisa

Near Infrared Spectroscopy (*NIRS*) merupakan teknik analisis yang sangat kuat di bidang penelitian seperti farmasi, pertanian, medis, makanan, dan lainnya. *NIRS* menggunakan inframerah dekat dari spektrum elektromagnetik untuk menciptakan getaran atom molekul dalam tes kimia [5]. Segmen inframerah dekat terletak diantara wilayah terlihat dan inframerah rentang menengah sekitaran 350 – 2500 nm [13]. Sebelum dilakukan pengolahan data, terlebih dahulu *pretreatment* data dengan *Standarlized Normal Variate* untuk menghilangkan efek hamburan dari spektral. Data tersebut sebelumnya tidak diketahui apakah terdapat *outlier* atau tidak dalam *dataset*. Oleh karena itu, metode ini dievaluasi berdasarkan nilai akurasi dengan memprediksi kandungan pupuk nitrogen (N) pada tanah kelapa sawit. Nitrogen (N) tanah salah satu yang dapat mempengaruhi koefisien tanaman, meningkatkan efisiensi penggunaan air dan menyelidiki efek pada pertumbuhan kelapa sawit [14].

3.1. Kandungan Nitrogen (N) Tanah Kelapa Sawit

Penelitian ini menggunkan 100 pengamatan yang terdiri dari 2151 panjang gelombang (350 – 2500 nm) dari spektrum *NIR* tanah perkebunan kelapa sawit. Prosedur *k-fold cross validation* digunakan untuk membagi data sebanyak 5 *fold* yang terdiri dari *data training* untuk melatih model dan *testing* untuk prediksi model. Prediksi yang baik dengan menentukan jumlah komponen optimal dengan *Root Mean Square Error Prediction* (*RMSEP*) rendah untuk mengurangi *overfitting* dan *underfitting*. Setiap *fold* memiliki jumlah komponen optimalnya dan evaluasi metriknya yaitu

ISSN (Printed) : 2579-7271
ISSN (Online) : 2579-5406

Tabel 1. Jumlah Komponen Optimal Setiap *Fold*

Jumlah *Fold*	Jumlah komponen optimal	*RMSE*	R^2
Fold-1	24	0.1842996	0.7207416
Fold-2	13	0.1961883	0.5150748
Fold-3	10	0.1258906	0.7821674
Fold-4	13	0.1892793	0.4695287
Fold-5	6	0.1336715	0.3296541

Tabel 1, dapat dilihat bahwa model dievaluasi *k-fold cross validation* menghasilkan rata-rata nilai *Root Mean Square Error* (*RMSE*) sebesar 0.1658659 yang menunjukkan tingkat kesalahan prediksi model yang relatif rendah. Nilai rata-rata *Coefficient of Determination* (R^2) sebesar 0.5634333 mengindikasikan bahwa model mampu mampu menjelaskan 56.34% variabilitas data yang menunjukkan hubungan cukup erat antara nilai prediksi dan aktual.

3.2. Identifikasi *Outlier* dengan *Radial Basis Function* (*RBF*) *Kernel Principal Component Analysis* (*PCA*)

Mengidentifikasi *outlier* dengan menggunakan *RBF Kernel PCA* di mana fungsinya sebagai berikut

$$K_{ij} = \Phi(x_i).\Phi(x_j)$$

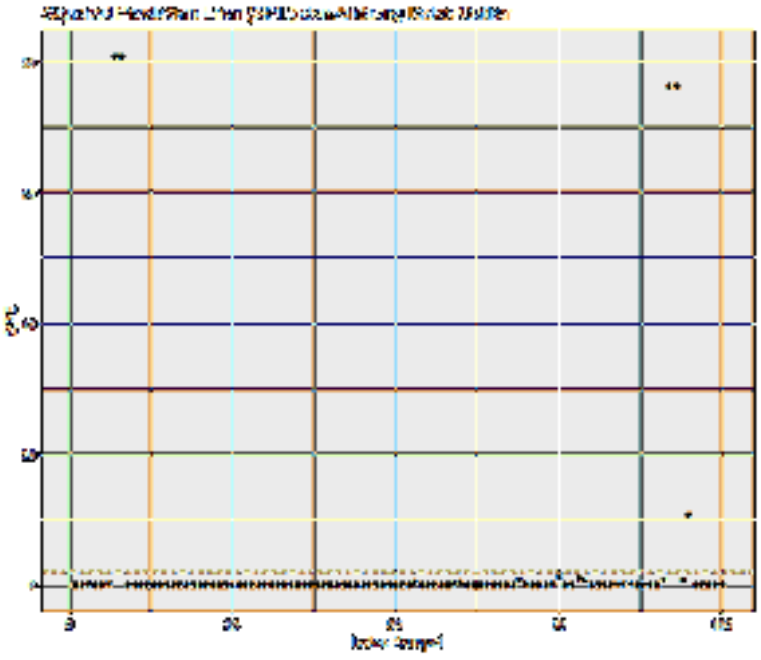
$$= \phi_1(x_i)\,\phi_1(x_j) + \cdots + \phi_\infty(x_i)\,\phi_\infty(x_j)$$

$$= exp\ exp\left(-\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma^2}\right) \quad (8)$$

di mana x_i dan x_j adalah dua sampel data dan σ adalah lebar kernel dari *RBF kernel*. Kemudian hitung nilai *Squared Prediction Error* (*SPE*) sebagai berikut [15]

$$SPE_j = 1 - 2\underline{K_j} + \underline{K} \quad (9)$$

Menggunakan *library* (*kernlab*) pada program *R* dapat memudahkan kita untuk mendeteksi *outlier* ada pada data atau tidak dengan menentukan nilai *SPE*. Jika nilai *SPE* > ambang batas pada kuantil 95% dari data. Berikut grafik data yang terdetiksi *outlier*



Gambar 2. Deteksi *Outlier* dengan *RBF Kernel PCA*

Gambar 2 teridentifikasi bahwa terdapat empat data yang termasuk *outlier* yaitu pada data ke 7, 8, 92, 93 dan 95. Keberadaan *outlier* menandakan bahwa data tersebut memiliki nilai yang secara signifikan berbeda dari pola umum yang ditemukan dalam *dataset*. Adanya *outlier*

ini perlu diperhatikan karena dapat mempengaruhi hasil analisis dan model prediktif yang dibangun. Oleh karena itu, dilakukan penanganan *outlier* dengan cara mengeluarkannya atau diatasi menggunakan metode *k-fold cross validation*.

Tabel 2. Hasil Metode dengan Model *PLSR*

Metode	Nilai R^2		Nilai *RMSE*	
	Ada *Outlier*	Tidak Ada *Outlier*	Ada *Outlier*	Tidak Ada *Outlier*
Partial Least Square Regression (*PLSR*)	0.359	0.833	0.121	0.061
PLSR dengan *K-Fold Cross Validation*	0.563	0.388	0.166	0.141

Berdasarkan hasil Tabel 2, perbandingan performa model *Partial Least Square Regression* (*PLSR*) dengan dan tanpa *K-fold cross validation* dalam dua kondisi adanya *outlier* dan tidak adanya *outlier*. Secara keseluruhan, hasil ini menunjukkan bahwa *PLSR* memberikan hasil terbaik dalam kondisi tidak adanya *outlier* sehingga metode ini tidak mampu dalam menangani *outlier.* Penerapan *K-fold cross validation* mampu tahan (*robust*) terhadap *outlier* dengan R^2 meningkat daripada *PLSR* klasik walaupun nilai *RMSE*nya lebih tinggi namun tidak jauh berbeda. Hal ini menunjukkan *K-fold cross validation* dapat membantu meningkatkan stabilitas model terutama dalam kondisi data yang lebih bervariasi dengan adanya *outlier*.

4. Kesimpulan

Adapun kesimpulan dari penelitian ini adalah sebagai berikut:

1. Adanya *outlier* dalam data secara signifikan mempengaruhi performa model *Partial Least Square Regression* (*PLSR*) klasik dengan penurunan nilai R^2 dan peningkatan nilai *RMSE*. Hal ini menunjukkan bahwa *PLSR* klasik kurang efektif dalam menangani data yang mengandung *outlier*.
2. Penerapan *k-fold cross validation* dalam model *PLSR* menunjukkan peningkatan nilai R^2 pada data yang mengandung *outlier* meskipun nilai *RMSE* sedikit meningkat. Ini menunjukkan bahwa *K-fold cross validation* mampu meningkatkan *robustitas* model terhadap variasi data termasuk adanya *outlier* dengan mempertahankan prediktabilitas yang lebih stabil.

Referensi

[1] D. D. Silalahi, H. Midi, J. Arasan, M. S. Mustafa, and J. P. Caliman. Kernel Partial Diagnostic Robust Potential To Handle High-Dimensional and Irregular Data Space On Near Infrared Spectral Data. *Heliyon*. 2020; 6(1): 03176-03187.
[2] Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak. Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron. *Computational and Mathematical Method in Medicine*; 2020 (2020):10-26.
[3] S. Raschka. *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. University of Wisconsin-Madison. Department of Statistics. Report number:1-13. 2018.
[4] D. D. Silalahi, H. Midi, J. Arasan, M. S. Mustafa, and J. P. Caliman. Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data. *Vib. Spectrosc.* 2018; 97 (11): 55-56.
[5] B. H. Stuart. Infrared Spectroscopy: Fundamentals and Applications. UK: Wiley. 2004.
[6] J. Supriatna, D. Djumarno, A. B. Saluy, and D. Kurniawan. Sustainability Analysis of Smallholder Oil Palm Plantations in Several Provinces in Indonesia. *Sustainability*. 2024; 16(11): 4383-4399.
[7] H. Aleiadeh *et al.* Effect of Co-application of Vetiver Grass Biochar and NPK Fertilizer on the Growth of Oil Palm (Elaeis guineensis Jacq.) Seedlings and Soil Chemical Properties. 2024; 28(1): 26-37.
[8] D. D. Silalahi, H. Midi, J. Arasan, M. S. Mustafa, and J. P. Caliman. Automated fitting process using robust reliable weighted average on near infrared spectral data analysis. *Symmetry (Basel)*. 2020; 12(12): 1-27.
[9] M. J. Kim *et al.* Prediction of Soluble-Solid Content in Citrus Fruit Using Visible–Near-Infrared Hyperspectral Imaging Based on Effective-Wavelength Selection Algorithm. *Sensors*. 2024;

24 (5): 1512-1524.
[10] V. A. Amankwaah *et al.* Development of NIRS calibration curves for sugars in baked sweetpotato. *Journal of the Science of Food and Agriculture*. 2024; 104 (8): 4801–4807.
[11] Y. Jiao, Z. Li, X. Chen, and S. Fei. Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics*, 2020; 34 (11): 3306-3324.
[12] T. Hastie, R. Tibshirani, and J. Friedman. Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction. Second Edition. California: Springer, 2008.
[13] D. A, B. E. W, and Ciurczak. Handbook_of_Near-Infrared_Analysis. Third Edition. Francis: CRC Press. 2008.
[14] R. Sigalingging, Sumono, and O. W. Pratiwi. The effect of NPK fertiliser on oil palm coefficient as a baseline water management during the nursery phase. *IOP Conference Series: Earth and Environmental Science*. 2024; 1302 (1): 012107-012115.
[15] R. Tan, J. R. Ottewill, and N. F. Thornhill. Monitoring statistics and tuning of kernel principal component analysis with radial basis function kernels. *IEEE Access*. 2020; 8(1):198328–198342.