

Feature Selection Model Development on Near-Infrared Spectroscopy Data

Case Study of Beef Freshness Quality Prediction

Ridwan Raafi'udin¹, Y. Aris Purwanto², Imas Sukaesih Sitanggang³, Dewi Apri Astuti⁴

Department of Computer Science, IPB University, Bogor, Indonesia^{1,3}

Mechanical and Biosystem Engineering Department, IPB University, Bogor, Indonesia²

Department of Nutrition and Feed Technology, IPB University, Bogor, Indonesia⁴

Department of Informatics, Universitas Pembangunan Nasional Veteran Jakarta, Indonesia¹

Abstract—This study aims to develop a feature selection model on Near-Infrared Spectroscopy (NIRS) data. The object used is beef with six quality parameters: color, drip loss, pH, storage time, Total Plate Colony (TPC), and water moisture. The prediction model is a Random Forest Regressor (RFR) with default parameters. The feature selection model is carried out by mapping spectroscopic data into line form. The collection of lines is made into one line by finding the mean value. Next, apply the line simplification method based on angle elimination, starting from the smallest angle to the largest. Each iteration will eliminate one corner, reducing one column of data in the corresponding dataset. Then, the predicted value in the form of R2 will be collected, and the highest value will be considered the best feature selection formation. RFR prediction results with R2 values are as follows: color R2= 0.597, drip loss R2=0.891, pH R2=0.797, storage time R2=0.889, TPC R2=0.721, and water moisture R2=0.540. Meanwhile, after applying the feature selection model, the R2 values for all parameters increased to color R2=0.877, drip loss R2=0.943, pH R2=0.904, storage time R2=0.917, TPC R2=0.951, and water moisture R2=0.893. Based on the results of increasing the R2 value of the six parameters, an average value of increasing prediction accuracy of 17.49% can be taken. So, the feature selection method based on line simplification with an angle elimination system can provide very good results.

Keywords—Beef quality prediction; feature selection; machine learning; Random Forest Regressor

I. INTRODUCTION

People have consumed a lot of meat in the last few decades, and consumption has increased in recent years [1]. Beef is an alternative commodity that is widely consumed to meet the need for protein in many countries [2]. However, meat food products are products that rot quickly, especially under certain conditions, which can accelerate microbial growth [3]. Many cases of consumers getting sick are caused by microbes found in beef that are large in number or above standard [4].

The condition of the meat can change quickly, so a method is needed that can determine the current state of beef quality. One fast method for determining meat quality is to use near-infrared (NIR) technology [5]. NIR spectrometers can predict meat quality parameters, including chemical parameters, technological parameters, quality traits, fatty acids, and many mineral contents [6][7][8].

To speed up and simplify the process of determining meat quality, a portable or handheld NIR device can be built that can be taken anywhere [9][10]. The development of a portable NIR device can be applied in industries that require the process of determining the characteristics of meat products [11][12].

Machine learning can be used as a model to predict meat quality parameters [13]. Machine learning is also able to predict meat quality parameters, including color, tenderness, juiciness, and flavor [14]. The random forest (RF) algorithm works well and produces high accuracy in classifying cattle breeds [15]. Random Forest Regressor (RFR) performs well in predicting pH values in beef in real-time in a beef freshness monitoring system [16].

In this study, we propose a feature selection model on spectroscopic data to increase the accuracy of meat quality predictions. The beef quality parameters in question are color, drip loss, pH, storage time, total plate colony (TCP), and water moisture. The algorithm that will be used is RFR, with algorithm parameters by default. Based on the experimental results, the proposed method is able to increase accuracy in predicting the freshness quality of beef compared to results without using feature selection.

II. LITERATURE SURVEY

A. Feature Selection

Feature selection involves exploring algorithms designed to decrease the data's dimensionality, thereby enhancing the performance of machine learning. In datasets with N features and M dimensions, the primary goal is to minimize M to M' , where M' is less than or equal to M [17]. Typically, feature selection often results in improved learning outcomes, such as increased accuracy in learning, reduced computational expenses, and enhanced interpretability of the model.

Recently, experts in fields like computer vision and text mining have introduced numerous feature selection methods. Through both theoretical frameworks and practical experiments, they've demonstrated the effectiveness of their approaches [18]. Feature selection using established line simplification methods such as the Ramer–Douglas–Peucker algorithm and Visvalingam on NIRS data encountered

problems in determining epsilon values, making it difficult to produce optimal datasets.

This study will compare the results of prediction accuracy from the proposed feature selection model with the established method from Scikit-Learn, namely SelectFromModel [19]. Comparison of results in the form of accuracy of value prediction on six meat quality parameters using R-squared values (R²) as part of model evaluation both for evaluation of the SelectFromModel feature selection model and for the proposed feature selection model.

B. Random Forest Regressor

Random forests consist of a collection of tree predictors where each tree relies on a randomly sampled vector, independent and identically distributed across all trees within the forest. As the number of trees in the forest increases, the generalization error for forests gradually converges to a limit. This error is influenced by both the effectiveness of the individual trees in the forest and the level of correlation among them [20].

The advantages of feature selection and its relevance to enhancing the effectiveness and interpretability of machine learning algorithms are well-established. In this context, we focus on incorporating feature selection into a Random Forest setup. We propose a method that integrates hypothesis testing with an estimation of the anticipated impact of an irrelevant feature while constructing a Random Forest [21].

The utilization of R² in this research for model assessment is due to its superior informativeness and reliability compared to SMAPE, along with its absence of interpretational constraints seen in metrics like MSE, RMSE, MAE, and MAPE[22]. The R² value range is easy to understand because the value ranges between 0 and 1 and is indicated by a decimal number. The number 0 indicates poor model performance, and one or close to it indicates good model performance. The formula for calculating the R² value can be seen in Eq. (1) [22].

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}} \\ = 1 - \frac{\sum(y_1 - \hat{y}_1)^2}{\sum(y_1 - \bar{y})^2} \quad (1)$$

C. Near-Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) holds a distinct position within the realm of bioscience and its associated fields, differing in characteristics and potential applications from infrared (IR) or Raman spectroscopy. This type of vibrational spectroscopy uncovers molecular details within a sample by detecting absorption bands arising from overtones and combined excitations [23].

The capacity of near-infrared reflectance spectroscopy (NIRS) to differentiate between Normal and DFD (dark, firm, and dry) beef and forecast quality characteristics within 129 Longissimus thoracic (LT) samples derived from three Spanish pure breeds: Asturiana de los Valles (AV; n = 50), Rubia Gallega (RG; n = 37), and Retinta (RE; n = 42). The outcomes obtained using partial least squares-discriminant analysis (PLS-

DA) demonstrated successful differentiation between Normal and DFD meat samples from AV and RG (with sensitivity exceeding 93% for both and specificity of 100% and 72%, respectively), whereas results for RE and the overall sample sets were less accurate. Soft independent modeling of class analogies (SIMCA) revealed 100% sensitivity for DFD meat across all sample sets (total, AV, RG, and RE) and over 90% specificity for AV, RG, and RE, but notably lower for the total sample set (19.8%). Utilizing NIRS quantitative models via partial least squares regression (PLSR) allowed dependable prediction of color parameters (CIE L*, a*, b*, hue, chroma). The findings from both qualitative and quantitative analyses hold promise for early decision-making in the meat production process to prevent financial losses and food wastage [24].

Growing apprehensions regarding contaminated meat have spurred the industry to explore novel, non-invasive techniques for swift and precise assessment of meat quality. The primary chromophores in meat (such as myoglobin, oxy-myoglobin, fat, water, and collagen) exhibit similar absorption patterns within the visible to near-infrared (NIR) spectral range. Consequently, variations in the structure and composition of meat can result in proportional disparities in light absorption [25].

III. MATERIAL

A. Sample Preparation

This study used fresh beef objects obtained from traditional markets in Bogor City, West Java, Indonesia, and then brought to the laboratory using an ice box, as shown in Fig. 1. The part of the carcass used is tenderloin. The main sample used weighs 1 kg, as shown in Fig. 2.

The next step is to cut it into eight pieces of large samples weighing 17+2 grams and eight pieces of small samples of 3+1 grams, as shown in Fig. 3. All samples were placed on the laboratory table in Petri dishes and supported by wire gauze, as shown in Fig. 4. Large one of samples were used for testing the parameters of color, pH, WHC, water content, and NIR, while small one of samples were used for TPC measurements. The total samples used were eight samples per day and repeated for ten days to obtain 80 samples.



Fig. 1. Sample in the icebox.



Fig. 2. 1 Kg tenderloin.



Fig. 3. Small pieces of sample.



Fig. 4. The samples are placed on a petri dish with a wire mesh base.

B. Data Acquisition

This study uses six beef freshness parameters, where each parameter uses a different tool. The six freshness parameters include color, drip loss, pH, storage time, TPC, and water moisture. The tool used to retrieve color data is Chromameter, as shown in Fig. 5.

The drip loss parameter value is obtained by measuring the weight of the sample between the initial and final times. The measurement interval is one hour with a total time span of seven hours so that eight drip loss data are produced starting from the 0th hour to the 7th hour. The tool for measuring sample weight is a laboratory scale, as shown in Fig. 6.



Fig. 5. Measurement results using a chromameter.



Fig. 6. Weight measurement process.

TPC parameter values were obtained from other laboratories and measured professionally by a third party. The samples sent are the same pieces used for measuring other parameters. Examples of samples are shown in Fig. 7.



Fig. 7. Sample for TPC measurement.



Fig. 8. pH measurement process.

The pH value parameter of the sample was measured using an electronic pH meter, as shown in Fig. 8. The storage time parameter value is obtained by simply storing the sample on the table from the 0th hour to the 7th hour with a break every hour. Then, the water moisture value is carried out using the drying or thermogravimetric method in an oven at a temperature of 105 degrees Celsius for 16 hours. Laboratory testing data collection activities were carried out for ten days, resulting in a dataset for 80 data rows, as shown in Table I.

TABLE I. DATA EXAMPLE FROM LABORATORY

(A)

Code			WHC			
Day	Hour	Code	W0	Wt	ΔW	%drip loss
4	0	d4h0	16.67	16.67	0.00	0%
4	1	d4h1	18.13	17.48	0.65	4%
4	2	d4h2	18.02	16.93	1.09	6%
4	3	d4h3	16.01	14.46	1.55	10%
4	4	d4h4	18.06	16.06	2.00	11%
4	5	d4h5	17.52	15.19	2.33	13%
4	6	d4h6	18.41	15.34	3.07	17%
4	7	d4h7	17.52	14.37	3.15	18%

(B)

Color			pH	Water Moisture (WM)			
L	a	b		Wdish	Wsample	Wt	%WM
33,05	16,23	1,42	5,51	25,47	3,45	26,32	75,36%
33,92	17,51	1,28	5,46	29,60	3,93	30,50	77,10%
36,46	16,79	0,08	5,55	28,46	3,84	29,41	75,26%
33,29	16,04	3,25	5,50	28,81	4,00	29,82	74,75%
32,47	15,68	4,94	5,55	30,16	3,92	31,19	73,72%
29,50	11,18	3,00	5,50	29,13	3,94	30,19	73,10%
31,62	12,67	3,23	5,48	25,12	3,67	26,16	71,66%
28,29	10,88	5,43	5,59	27,97	3,68	28,97	72,83%

W_0 . initial weight; W_t . new weight; ΔW . weight changes; L ; lightness; a. red/green value; b. yellow/blue value; W_{dish} , the weight of the empty dish; W_{sample} , the weight of the current sample; W_t , weight changes.

C. NIRS Data and Dataset

NIRS data is obtained by scanning samples using a NIRS sensor. The sensor used is the NeoSpectra Development Kit [26] as shown in Fig. 9 and Fig. 10. For the wireless data acquisition process, use a notebook unit equipped with requesting software that is available from the sensor manufacturer as shown in Fig. 11.



Fig. 9. NeoSpectra development kit.



Fig. 10. NeoSpectra development kit with case and sample.

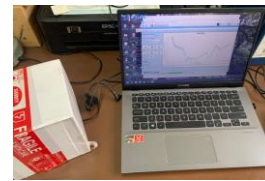


Fig. 11. Acquisition using NeoSpectraKit in progress covered by a box.

Measurement results in a spreadsheet file. The example of the averaged data can be seen in Table II. The original data of NIR Spectroscopy plotted as a graphic can be seen in Fig. 12. The total NIRS data collected is 720 data rows and 136 columns according to wavelength value from the sensor and visualized. as shown in Fig. 12. The wavelength used is between 1346.61 - 2556.24 nanometers (nm).

TABLE II. EXAMPLE OF NIRS DATA

Hour	Wavelength (nm)				
	2556.24	2539.35	...	1351.35	1346.61
0	2.27	2.37	...	1.41	1.35
1	1.96	2.08	...	1.72	1.95
2	1.78	1.86	...	1.62	1.53
3	2.12	2.22	...	1.89	1.88
4	2.11	2.28	...	1.85	1.77
5	1.71	1.86	...	2.01	2.11
6	2.13	2.30	...	2.86	2.55
7	2.81	2.91	...	3.79	4.01

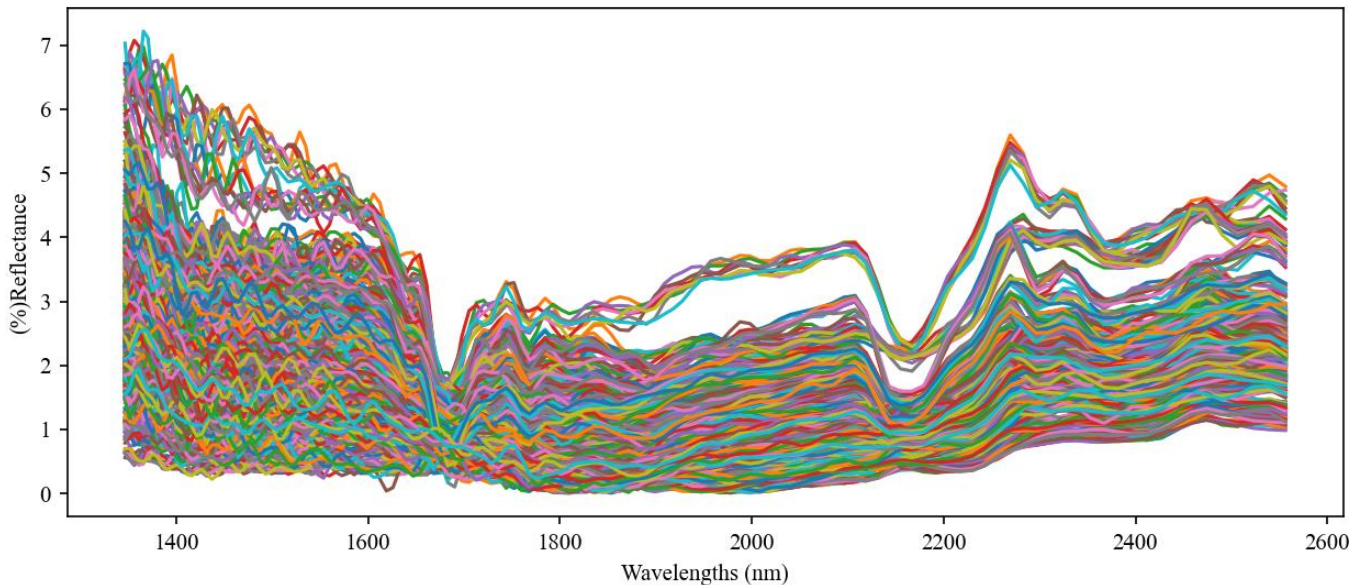


Fig. 12. Plot of original data spectrum.

IV. PROPOSED FEATURE SELECTION MODEL

The proposed feature selection model aims to reduce the amount of data based on columns in the dataset, which by default has 136 columns. This feature selection concept is based on the line simplification model by reducing straight or slightly curved lines. The main stages of the feature selection process can be seen in Fig. 13.

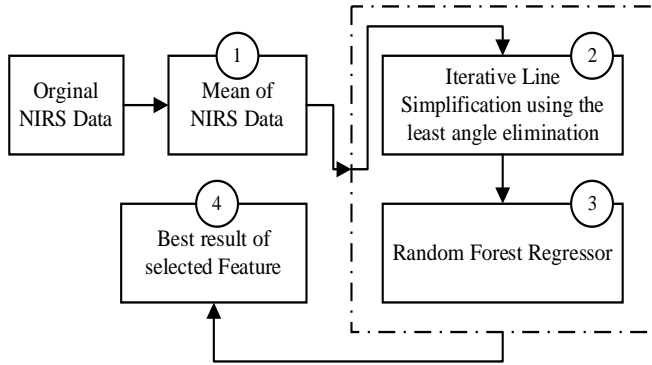


Fig. 13. Flow process of proposed selection feature.

The feature selection process goes through four stages as follows :

A. Mean and Single Data Line

In this process, 1 line is produced, which will represent all spectrum data. as shown in Fig. 14. Then, it is divided into one separate piece represented by one different color, as shown in Fig. 15. Then pair them with adjacent lines. as shown in Fig. 15. Then calculate the angle values of two adjacent lines as illustrated by Fig. 16.

B. Iterative Line Simplification

The process of calculating all angles along a line produces 134 angle values. Then, the angle value data is sorted starting from the smallest value. Each iteration will eliminate one corner with the smallest value. The eliminated corners will correspond to the columns that will be eliminated as well. At this stage. a set of data columns is stored with the storage index in a sequence of iterations. The elimination stages are depicted in the diagram Fig. 17.

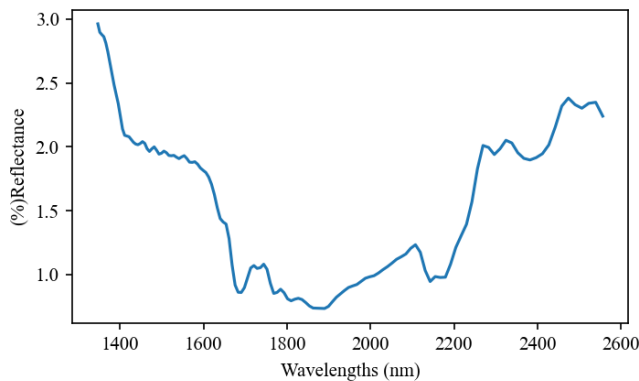


Fig. 14. Mean of all spectrums.

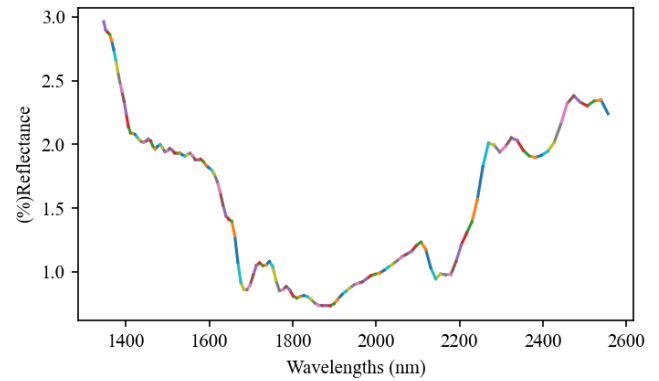


Fig. 15. Separated spectrum into single data line.

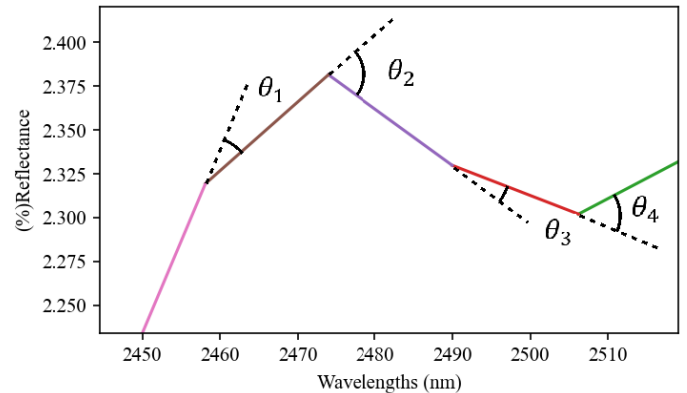


Fig. 16. Illustration of calculating the angle between two adjacent lines.

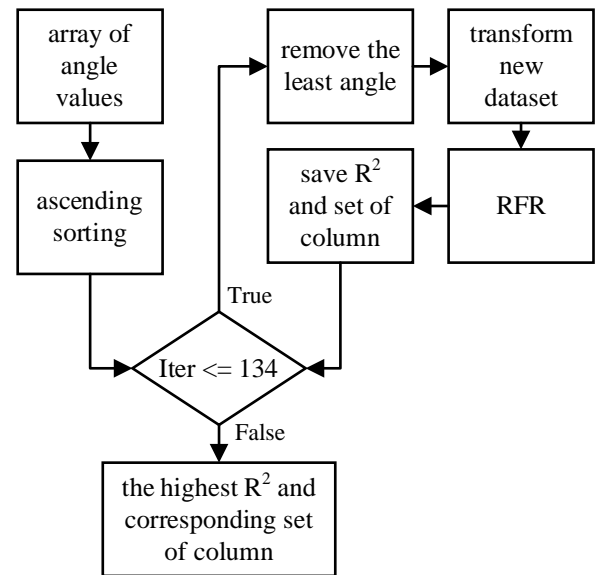


Fig. 17. Iterative angle and column elimination.

C. Random Forest Regressor

Each iteration produces a new dataset with a reduction of 1 column of data. The new dataset will then enter the machine learning process to produce an R^2 value at each iteration. At this stage, the RFR parameters used are the default settings, as shown in Table III. Data splitting for the learning and testing process is 70% to 30%.

TABLE III. DEFAULT PARAMETERS OF RFR [19]

Parameter	Data type	Default value
n_estimators		100
criterion		squared_error
max_depth		None
min_samples_split		2
min_samples_leaf		1
min_weight_fraction_leaf	float	0.0
max_features	int or float	1.0
max_leaf_nodes	int	None
min_impurity_decrease	float	0.0
bootstrap	bool	True
oob_score	bool or callable	False
n_jobs	int	None
random_state	int	None
verbose	int	0
warm_start	bool	False
ccp_alpha	non-negative float	0.0
max_samples	int or float	None
monotonic_cst	array-like of int of shape (n_features)	None

The results of R^2 at each iteration will be stored in an array. Then, it will find the highest R^2 value and location in the array to be used as the best column index set.

D. The best result of the selected feature

The final stage is to determine the largest value from the collection of R^2 that has been accommodated in the array. To determine the highest R^2 value in this study, use the numpy library with the numpy.max() command [27]. To find out the iteration position of the highest R^2 value, also use the numpy library with the command numpy.argmax() [28]. This command will display the data index of the highest R^2 value.

To determine the best set of columns is to store the value of R^2 in each iteration. Then, find the highest value of the array to determine the index value. The index value is used to retrieve the set of columns.

V. RESULTS

A. Results from the Original Dataset and the Proposed Model

Of the six meat quality parameters, namely color, drip loss, pH, storage time, TCP, and water moisture, feature selection and machine learning models are applied alternately. Then, we will compare the prediction results using the original dataset with the dataset that has gone through the feature selection stage. The results of the comparison of R^2 values can be seen in IV.

Based on the R^2 value in Table IV, it can be seen that the line simplification-based feature selection model with the corner elimination method has succeeded in increasing the performance of the RFR algorithm in predicting all beef quality parameters. The increase in the R^2 value for all parameters can

be said to be satisfactory, with the smallest increase being in the predicted storage time parameter, which is only 0.028.

TABLE IV. R^2 SCORE USING THE ORIGINAL DATASET AND USING FEATURE SELECTION

Beef Quality Parameters	R^2 score		
	Original dataset	With the proposed feature selection	increment
color	0.597448	0.876992	0.279544
drip loss	0.891545	0.942723	0.051178
pH	0.796903	0.904225	0.107322
storage time	0.889070	0.916581	0.027511
TPC	0.720895	0.951388	0.230493
water moisture	0.539780	0.893200	0.353420

The results were very good, and the highest increase in the R^2 value was in the prediction of the water moisture parameter, namely 0.353. From all the increases in R^2 values, the average value can be taken to be 0.1749 or 17.49%.

B. Results from the Original Dataset and SelectFromModel

As a comparison of performance results in this study, the feature selection model from Scikit-Learn, namely SelectFromModel, was also tested. The experimental results of implementing the SelectFromModel library can be seen in Table V.

TABLE V. R^2 SCORE OF USING ORIGINAL DATASET AND SELECTFROMMODEL

Beef Quality Parameters	R^2 score		
	Original dataset	SelectFromModel	increment
color	0.597448	0.835490	0.238042
drip loss	0.891545	0.917345	0.025800
pH	0.796903	0.846478	0.049575
storage time	0.889070	0.933890	0.044820
TPC	0.720895	0.928947	0.208052
water moisture	0.539780	0.823419	0.283639

Based on the R^2 value in Table V, it can be seen that feature selection by the SelectFromModel library can also improve the performance of the RFR algorithm. The smallest improvement was in the prediction of the drip loss parameter, namely 0.026, while the biggest improvement was in the prediction of the water moisture parameter, namely 0.2836. With the increase in the R^2 value in the prediction of all parameters, it can be said that feature selection using the SelectFromModel library works very well. Of all the increases in R^2 , the average value can be taken to be 0.1417 or 14.17%.

VI. DISCUSSION

A. R^2 Score Comparison between the Proposed Model and SelectFromModel

In this stage, the results of the R^2 values from the proposed feature selection model and the SelectFromModel feature selection model are compared. The comparison results also contain the number of features selected based on their highest R^2 value, which can be seen in Table VI.

TABLE VI. COMPARISON OF R² SCORES BETWEEN THE PROPOSED MODEL AND SELECTFROMMODEL

Beef Quality Parameters	Proposed Model		SelectFromModel	
	R ²	n feature	R ²	n feature
color	0.876992	84	0.835490	40
drip loss	0.942723	117	0.917345	20
pH	0.904225	57	0.846478	26
storage time	0.916581	71	0.933890	18
TPC	0.951388	56	0.928947	20
water moisture	0.893200	77	0.823419	31

Based on Table VI. it can be seen that the R² values of the proposed feature selection mode are higher than the SelectFromModel results, except for the R² value in the storage time parameter prediction. So overall, the average increase in the R² value of the proposed model is 17.49% higher than the

average R² value of SelectFromModel. which produces 14.17%.

B. Overview of Selected Features

The result of the feature selection model is a set of features in the form of data columns; in this study, the column names are wavelength values in nanometer (nm) units. The number of features produced by the proposed model and SelectFromModel is definitely less than the number of columns in the original dataset, so this feature selection also leads to a reduction in data dimensionality. For the differences in the number and features selected, a visualization was created for all parameters, which can be seen in Fig. 18 to Fig. 23.

Each color represents one feature selection model. The red color represents the mean of all NIRS data. The green color represents the mean of the data columns selected by the proposed model, while the blue represents the results from the SelectFromModel library.

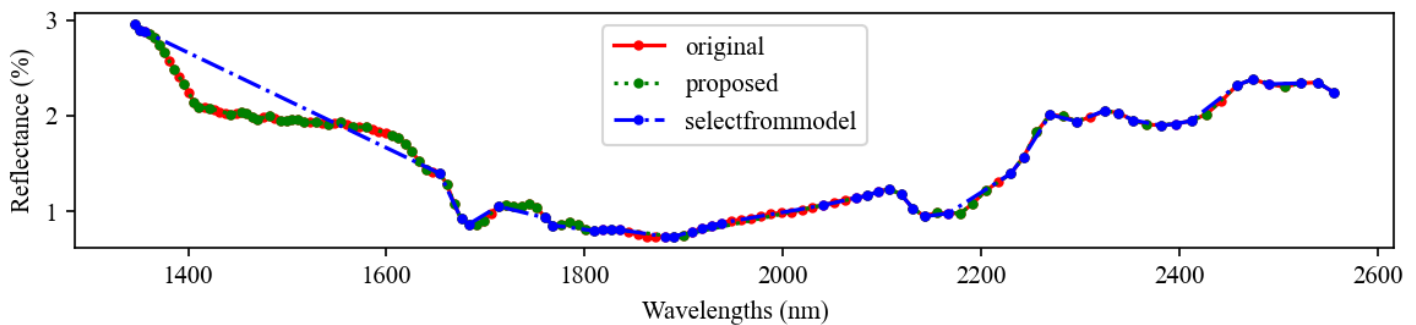


Fig. 18. Spectrums overlay for the color parameter.

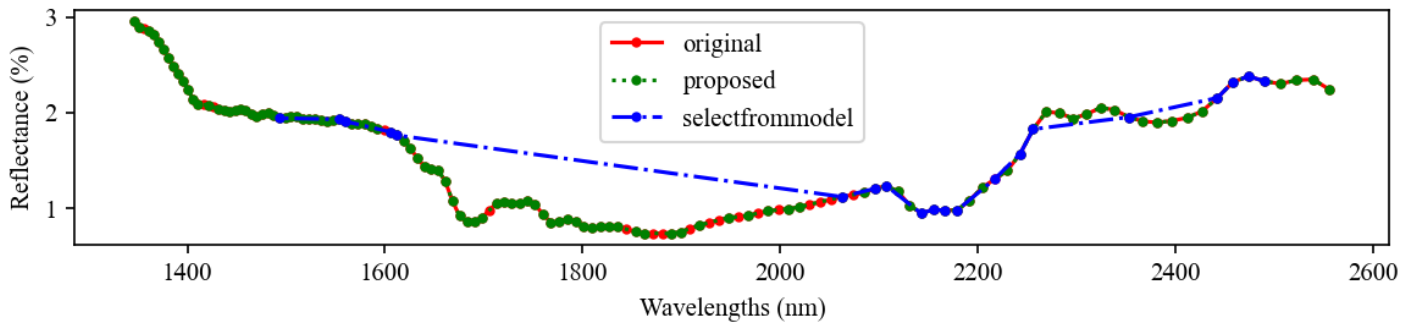


Fig. 19. Spectrum overlay for drip loss parameter.

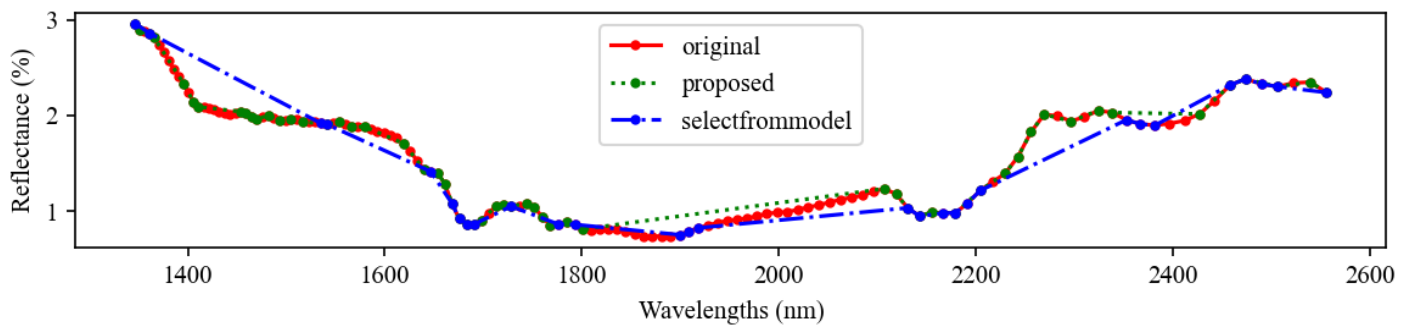


Fig. 20. Spectrums overlay for pH parameter.

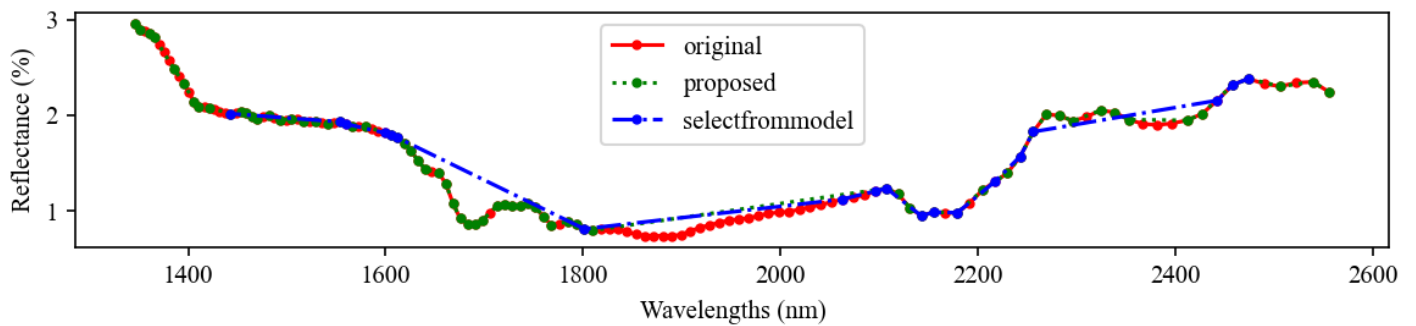


Fig. 21. Spectrums overlay for storage time parameter.

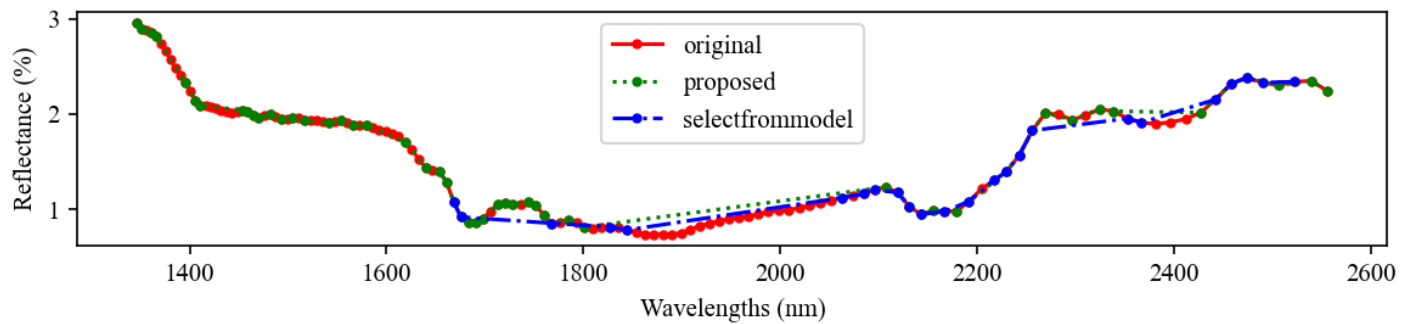


Fig. 22. Spectrums overlay for TPC parameter.

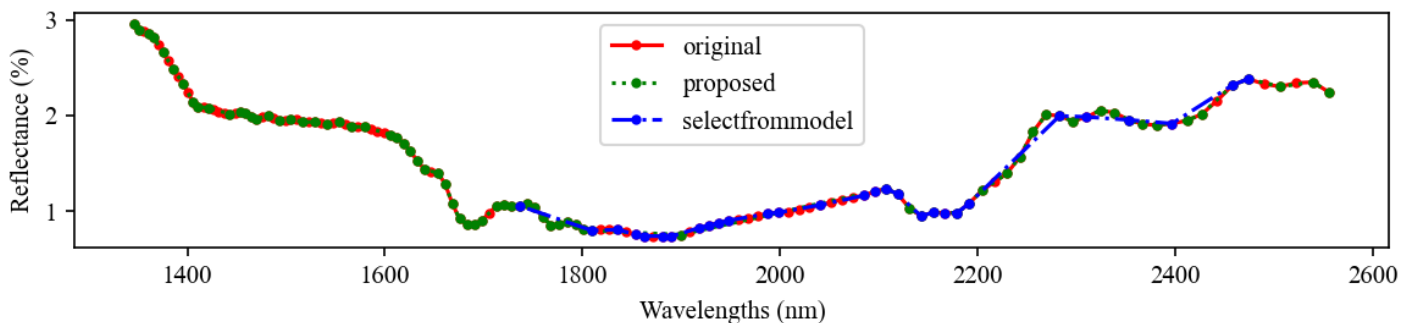


Fig. 23. Spectrums overlay for water moisture parameter.

VII. CONCLUSION AND FUTURE WORK

This study has produced a feature selection model based on line simplification by eliminating angles in the average spectrum from beef NIRS data. The result of increasing RFR performance after using the proposed feature selection model is 17.49%. This result is higher than the average increase in R2 value produced by the SelectFromModel library of 14.17%. Apart from being able to increase prediction accuracy, this model can also reduce data dimensions, where fewer data will require a shorter time in the machine learning process.

For further work and development, the proposed feature selection model can be applied to deep learning algorithms. It can also be combined with RFR by applying hyperparameter tuning. The combination with hyperparameter tuning may require a longer time to find the solution set for the highest accuracy. However, RFR with hyperparameter tuning produces better accuracy compared to RFR with default parameters.

REFERENCES

- [1] M. Lima, R. Costa, I. Rodrigues, J. Lameiras, and G. Botelho, "A Narrative Review of Alternative Protein Sources: Highlights on Meat, Fish, Egg and Dairy Analogues," *Foods*, vol. 11, no. 14, p. 2053, Jul. 2022, doi: 10.3390/foods11142053.
- [2] M. Molfetta et al., "Protein Sources Alternative to Meat: State of the Art and Involvement of Fermentation," *Foods*, vol. 11, no. 14, p. 2065, Jul. 2022, doi: 10.3390/foods11142065.
- [3] B. Fletcher et al., "Advances in meat spoilage detection: A short focus on rapid methods and technologies," *CyTA - J. Food*, vol. 16, no. 1, pp. 1037–1044, Jan. 2018, doi: 10.1080/19476337.2018.1525432.
- [4] V. Tesson, M. Federighi, E. Cummins, J. de Oliveira Mota, S. Guillou, and G. Boué, "A Systematic Review of Beef Meat Quantitative Microbial Risk Assessment Models," *Int. J. Environ. Res. Public Health*, vol. 17, no. 3, p. 688, Jan. 2020, doi: 10.3390/ijerph17030688.
- [5] W. Barragán-Hernández, L. Mahecha-Ledesma, J. Angulo-Arizala, and M. Olivera-Angel, "Near-Infrared Spectroscopy as a Beef Quality Tool to Predict Consumer Acceptance," *Foods*, vol. 9, no. 8, p. 984, Jul. 2020, doi: 10.3390/foods9080984.
- [6] A. Goi, J.-F. Hocquette, E. Pellattiero, and M. De Marchi, "Handheld near-infrared spectrometer allows on-line prediction of beef quality traits," *Meat Sci.*, vol. 184, p. 108694, Feb. 2022, doi: 10.1016/j.meatsci.2021.108694.

- [7] A. Sahar et al., "Online Prediction of Physico-Chemical Quality Attributes of Beef Using Visible—Near-Infrared Spectroscopy and Chemometrics," *Foods*, vol. 8, no. 11, p. 525, Oct. 2019, doi: 10.3390/foods8110525.
- [8] N. Patel, H. Toledo-Alvarado, A. Cecchinato, and G. Bittante, "Predicting the Content of 20 Minerals in Beef by Different Portable Near-Infrared (NIR) Spectrometers," *Foods*, vol. 9, no. 10, p. 1389, Oct. 2020, doi: 10.3390/foods9101389.
- [9] S. Savoia et al., "Prediction of meat quality traits in the abattoir using portable and handheld near-infrared spectrometers," *Meat Sci.*, vol. 161, p. 108017, Mar. 2020, doi: 10.1016/j.meatsci.2019.108017.
- [10] S. Savoia, A. Albera, A. Brugiapaglia, L. Di Stasio, A. Cecchinato, and G. Bittante, "Prediction of meat quality traits in the abattoir using portable near-infrared spectrometers: heritability of predicted traits and genetic correlations with laboratory-measured traits," *J. Anim. Sci. Biotechnol.*, vol. 12, no. 1, p. 29, Dec. 2021, doi: 10.1186/s40104-021-00555-5.
- [11] I. M. N. Perez, L. J. P. Cruz-Tirado, A. T. Badaró, M. M. de Oliveira, and D. F. Barbin, "Present and future of portable/handheld near-infrared spectroscopy in chicken meat industry," *NIR news*, vol. 30, no. 5–6, pp. 26–29, Aug. 2019, doi: 10.1177/0960336019861476.
- [12] M. Simoni, A. Goi, M. De Marchi, and F. Righi, "The use of visible/near-infrared spectroscopy to predict fibre fractions, fibre-bound nitrogen and total-tract apparent nutrients digestibility in beef cattle diets and faeces," *Ital. J. Anim. Sci.*, vol. 20, no. 1, pp. 814–825, Jan. 2021, doi: 10.1080/1828051X.2021.1924884.
- [13] C. N. Sánchez, M. T. Orvañanos-Guerrero, J. Domínguez-Soberanes, and Y. M. Álvarez-Cisneros, "Analysis of beef quality according to color changes using computer vision and white-box machine learning techniques," *Heliyon*, vol. 9, no. 7, p. e17976, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17976.
- [14] T. Qiao, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall, "Quantitative Prediction of Beef Quality Using Visible and NIR Spectroscopy with Large Data Samples Under Industry Conditions," *J. Appl. Spectrosc.*, vol. 82, no. 1, pp. 137–144, Mar. 2015, doi: 10.1007/s10812-015-0076-1.
- [15] R. Kasarda, N. Moravčíková, G. Mészáros, M. Simčič, and D. Zaborski, "Classification of cattle breeds based on the random forest approach," *Livest. Sci.*, vol. 267, p. 105143, Jan. 2023, doi: 10.1016/j.livsci.2022.105143.
- [16] Y. Lin, J. Ma, D.-W. Sun, J.-H. Cheng, and Q. Wang, "A pH-Responsive colourimetric sensor array based on machine learning for real-time monitoring of beef freshness," *Food Control*, vol. 150, p. 109729, Aug. 2023, doi: 10.1016/j.foodcont.2023.109729.
- [17] H. Liu, "Feature Selection," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 402–406. doi: 10.1007/978-0-387-30164-8_306.
- [18] J. Miao and L. Niu, "A Survey on Feature Selection," *Procedia Comput. Sci.*, vol. 91, pp. 919–926, 2016, doi: 10.1016/j.procs.2016.07.111.
- [19] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011, [Online]. Available: <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [20] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.
- [21] J. Rogers and S. Gunn, "Identifying Feature Relevance Using a Random Forest BT - Subspace, Latent Structure and Feature Selection," 2006, pp. 173–184.
- [22] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.
- [23] K. B. Beć, J. Grabska, and C. W. Huck, "Near-Infrared Spectroscopy in Bio-Applications," *Molecules*, vol. 25, no. 12, p. 2948, Jun. 2020, doi: 10.3390/molecules25122948.
- [24] D. Tejerina, M. Oliván, S. García-Torres, D. Franco, and V. Sierra, "Use of Near-Infrared Spectroscopy to Discriminate DFD Beef and Predict Meat Quality Traits in Autochthonous Breeds," *Foods*, vol. 11, no. 20, p. 3274, Oct. 2022, doi: 10.3390/foods11203274.
- [25] M. Peyvasteh, A. Popov, A. Bykov, and I. Meglinski, "Meat freshness revealed by visible to near-infrared spectroscopy and principal component analysis," *J. Phys. Commun.*, vol. 4, no. 9, p. 095011, Sep. 2020, doi: 10.1088/2399-6528/abb322.
- [26] "Si-Ware Systems announces NeoSpectra Micro Development Kits | NeoSpectra." <https://www.si-ware.com/featured-resources/si-ware-systems-announces-neospectra-micro-development-kits> (accessed Jun. 19, 2023).
- [27] "numpy.max — NumPy v1.26 Manual." <https://numpy.org/doc/stable/reference/generated/numpy.max.html> (accessed Nov. 16, 2023).
- [28] "numpy.argmax — NumPy v1.26 Manual." <https://numpy.org/doc/stable/reference/generated/numpy.argmax.html> (accessed Nov. 16, 2023).