Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1157

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

The Random Forest Regression for Price Prediction and Profitability Analysis in a Mushroom Sales Enterprise System

Ulya Ilhami Arsyah[1]*, Dian Eka Putra[1], Defni[1], Ikhsan[1], Ainil Mardiah[1]
[1]Information Technology, Politeknik Negeri Padang, Padang, Indonesia
Corresponding Author: ulya@pnp.ac.id

Article Information

Article history:

No. 1157
Rec. July 02, 2026
Rev. October 31, 2026
Acc. November 06, 2026
Pub. November 14, 2026
Page. 134–150

Keywords:

- Random Forest Regression
- Price prediction
- Profitability analysis
- Enterprise system
- Machine learning
- XGBoost
- K-fold cross-validation

ABSTRACT

Mushroom processing enterprises managed by community groups in Limau Manis Village, Pauh District, frequently fail to survive because producers lack a reliable basis for setting selling prices and estimating profit. This study developed an enterprise system that applies Random Forest Regression to predict the selling price of processed mushroom products and to analyze profitability. The dataset consisted of 2,193 daily sales records of three processed products (chips, nuggets, and floss) from an industry partner during 2023–2024. The base features (product type, weight, production cost, packaging, and season) were expanded with lagged market-demand indicators, including sales revenue, and with holiday, Ramadan/Eid, and calendar variables. Random Forest was benchmarked against XGBoost, linear regression, and ridge regression using MAE, RMSE, and R^2 under 5-fold cross-validation, with hyperparameters tuned by grid search. Tuning improved Random Forest (test MAE Rp 2,129, RMSE Rp 2,448, R^2 0.641; cross-validated R^2 0.677 ± 0.023) and reduced its train–test R^2 gap from 0.329 to 0.208. XGBoost (R^2 0.647) and the linear models (R^2 0.661–0.662) performed comparably, and the expanded features did not reduce residual variance. Further analysis showed that prices follow a cost-plus pattern: production cost explains 69% of price variation, while the remaining margin is set by producers independently of demand and calendar conditions. The model was deployed as a Streamlit web application that produces price predictions, profit estimates, and profit visualizations to support data-driven pricing decisions.

How to Cite:

Arsyah, U. I., & et al. (2026). Random Forest Regression for Price Prediction and Profitability Analysis in a Mushroom Sales Enterprise System. Jurnal Teknologi Informasi Dan Pendidikan, 20(1), 134–150. https://doi.org/10.24036/jtip.v20i1.1157

This open-access article is distributed under the Creative Commons Attribution-ShareAlike 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2023 by Jurnal Teknologi Informasi dan Pendidikan.

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1157

CC BY SA

1. INTRODUCTION

According to figures released by the Central Statistics Agency (BPS), the volume of mushroom production in 2022 fell to 63.15 tons [1]. The same agency's figures also suggest that cultivating and processing mushrooms may serve as a supplementary livelihood for households. Such potential encourages families to raise their earnings, particularly through mushroom farming and the processing of mushroom-based products. Because many households in the Limau Manis area own sizeable yards, the local community is well placed to develop such a venture. A single community-run mushroom business operates in Limau Manis Village, Pauh District. Based on BPS data, the average source of income for the community in the Limau Manis area is farming or employment in government and non-government agencies.

These economic and geographic circumstances help residents of Limau Manis pursue extra income, and one route toward that goal is opening up new job opportunities. Among the ventures favored by the Limau Manis community are mushroom cultivation and processing. Growing mushrooms does not depend on a particular season, requires only a small plot of land, and yields both fresh mushrooms and processed goods that residents can consume directly. This certainly attracts many people interested in mushroom cultivation.

Set up by the Limau Manis sub-district, the village children's forum serves, among other roles, to mentor the local community in strengthening its economic capacity and unlocking economic potential. Working with both firms and public institutions, the forum has run numerous joint efforts to lift the local economy, one of which supports the mushroom growing and processing community, recorded in its register as 77 mushroom groups. Out of those 77 groups, merely 65% have managed to endure and progress in production as well as processing, while the remainder have shut down and halted their operations. Several causes underlie this, one of them being that members of these groups have limited understanding of how to promote their goods and set an appropriate selling price.

Producers frequently struggle to fix ideal selling prices and to obtain the greatest possible profit. Because prices shift with production expenses, buyer demand, seasonal cycles, and weather, the task of pricing becomes complicated [2], [3]. Adopting technologies like Artificial Intelligence (AI) and machine learning offers a way to forecast prices reliably and to examine profitability [4].

Several previous studies have demonstrated the effectiveness of machine learning, particularly the Random Forest algorithm, for price prediction in the agricultural and commodity domains. As an ensemble technique, Random Forest constructs many decision trees from bootstrap resamples and combines their predictions; this design limits overfitting and lets the model capture non-linear patterns [5], [6]. In a review of Random Forest and Decision Tree regression for forecasting crop prices, Rakhra et al. found that ensemble methods reliably surpass models based on a single tree [7]. Testing a range of machine learning methods for predicting vegetable prices across seventeen Odisha markets in India, Paul et al. showed that these methods perform better than traditional time-series models [8]. Chen et al. built an automated system for predicting agricultural commodity prices and observed that tree-based ensembles kept their accuracy stable even on varied market data [9]. In the Indonesian context, Ibrahim applied Random Forest Regression to forecast shallot prices to support price stabilization [10], Muchtar and Afiyati showed that Random Forest outperformed linear regression for premium rice price prediction in West Java [11], and Pratama et al. reported similar findings for shallot prices in Samarinda [12]. The same method has likewise been used effectively, with strong accuracy, to estimate property prices [13], [14]. Recent studies on ensemble forecasting of agricultural commodity prices reaffirm that Random Forest is still among the most precise and dependable baselines, trailing only gradient-boosting methods [15].

While Random Forest has been examined extensively for price prediction, the majority of earlier studies end at the modeling phase and concentrate on unprocessed agricultural goods (rice, shallots, crops) or on property. Moreover, many small-scale price-prediction studies evaluate a single model on a single train–test split, without benchmarking it against alternative learners or validating its generalization through cross-validation [11], [12]. Studies that (i) target processed micro-scale products (processed mushroom products such as chips, nuggets, and floss), (ii) enrich the price model with calendar and market-demand information, (iii) evaluate it rigorously against linear and gradient-boosting baselines under k-fold cross-validation, and (iv) integrate the model into a ready-to-use enterprise application that couples price prediction with profitability analysis are still very limited. This constitutes the research gap and the novelty of this study.

This study selected Random Forest Regression because it copes well with complex, non-linear data [5], [7]. This research aims to develop an AI-based enterprise system to assist mushroom entrepreneurs in better decision-making, specifically by (1) building a Random Forest Regression model to predict the selling price of processed mushroom products, enriched with holiday/calendar and market-demand features; (2) benchmarking the model against XGBoost, linear regression, and ridge regression using MAE, RMSE, and R^2 under k-fold cross-validation with hyperparameter tuning; and (3) deploying the model in a web application equipped with profit analysis to support pricing decisions for mushroom business groups in Limau Manis Village.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

2. RESEARCH METHOD

There are several stages that constitute the methodology of this research. These stages serve as a guide in carrying out the research activities, adapted from the stages of machine-learning system development [4], [9]: requirements, design, evaluation, deployment, and analyzing. The stages of the research methodology are presented in Figure 1.

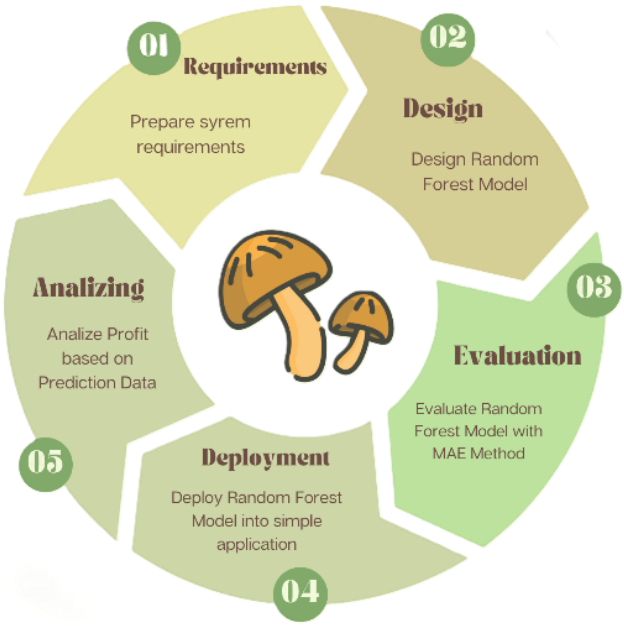


Figure 1. Research methodology

2.1. Requirements

This study begins by examining what the system needs. Here the requirements are described, covering the data employed, the way it is gathered, and how it is subsequently processed. The following explains the requirements stage.

Data Collection. The dataset for this work comprises sales records of processed mushroom goods from 2023 through 2024. It was compiled by interviewing industry partners and by logging each day's sales transactions directly. The dataset covers three types of processed products (mushroom chips, mushroom nuggets, and mushroom floss) with eight attributes: date, type of product, weight of product (grams), cost of production (Rp), packaging type, season, selling price (Rp), and number of units sold, totaling 2,193 transaction records (three products × 731 days). Table 1 presents an example of the dataset used in this study.

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1157

Table 1. Example of the dataset

Date	Type of Product	Weight (g)	Cost of Production (Rp)	Packaging	Season	Selling Price (Rp)	Units Sold
2023-01-01	Chips	250	3,860	Plastic	Rainy	11,086	81
2023-01-01	Nuggets	500	14,284	Plastic	Rainy	22,549	84
2023-01-01	Floss	200	14,636	Box	Rainy	19,555	161
2023-01-02	Chips	200	4,685	Plastic	Rainy	7,454	97
2023-01-02	Nuggets	250	9,949	Box	Rainy	17,260	197
2023-01-02	Floss	500	4,184	Box	Rainy	9,569	31
2023-01-03	Chips	500	14,096	Plastic	Rainy	16,570	68
2023-01-03	Nuggets	150	5,047	Box	Rainy	9,794	24
2023-01-03	Floss	200	4,899	Plastic	Rainy	8,166	73
2023-01-04	Chips	100	14,394	Plastic	Rainy	20,284	144

Data Preparation. The preprocessing stage cleans the data before it is used in the training phase. This study includes several preprocessing activities, including data cleaning, encoding of categorical features, and feature selection. There is no labeling process, as the goal of the study is to predict price, not quality.

Dataset cleaning. The cleaning step guarantees that the dataset contains no missing entries and that every record follows a uniform format. Entries with blank or invalid content were checked so that all features held valid values prior to training.

Feature encoding and selection. The categorical attributes, namely product type, packaging, and season, were transformed into numeric form through one-hot encoding to make them usable by the model. The raw date string was not used directly as an input; instead, it was used to derive the calendar features described below and to order the records of each product chronologically when computing lagged demand indicators. The base feature set consisted of product type, product weight, production cost, packaging type, and season, with the selling price as the prediction target. After the first record of each product was removed (its lagged values are undefined), 2,190 records remained. These were divided randomly into a training set (80%, 1,752 records) and an independent test set (20%, 438 records) with random_state = 42; the training set was used for cross-validation and hyperparameter tuning, and the test set only for the final evaluation.

Feature expansion. To reduce the variance left unexplained by the base features, two groups of derived features were added (Table 2). First, market-demand indicators were computed for each product from the partner's records: the previous day's units sold, the 7-day and 30-day moving averages of units sold, a demand-trend ratio (7-day over 30-day average), and the 7-day moving average of daily sales revenue (selling price × units sold). All demand indicators were shifted by one day so that only information available before the transaction date is used, which prevents target leakage. Second, holiday and calendar variables were constructed for every date in 2023–2024 from the official joint ministerial decrees on national holidays and collective leave [20]: national holiday or collective leave, Ramadan, a window of ±7 days around Eid al-Fitr, weekend, and month. The partner did

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

not record promotions or raw-material purchase prices, so a promotion flag could not be built, and input-cost variation is represented by the production cost per package, which already embeds the cost of raw materials. The expanded feature set therefore contains 15 input variables (5 base and 10 derived).

Table 2. Base and expanded feature sets

Group	Features	Source / derivation
Base	Product type, packaging, season (one-hot); product weight; production cost	Partner sales records
Market demand	Units sold on the previous day; 7-day and 30-day moving averages of units sold; demand trend (7-day / 30-day average); 7-day moving average of sales revenue (all lagged by one day)	Derived from partner records
Holiday and calendar	National holiday or collective leave; Ramadan; ±7 days around Eid al-Fitr; weekend; month	Joint ministerial decrees 2023–2024 [20]

2.2. Design

For the model, this study adopts Random Forest Regression, an ensemble approach that merges the outputs of numerous decision trees grown on bootstrap resamples while choosing features at random for every split [5], [6]. The baseline configuration uses n_estimators = 100 with fully grown trees, and random_state = 42 to guarantee reproducibility [16]. To position Random Forest objectively, three comparison models were trained on the same features and data split: ordinary least-squares linear regression, ridge regression with L2 regularization [22], and XGBoost, a regularized gradient-boosting tree ensemble [23]. Numerical features were standardized for the two linear models, whereas the tree-based models used unscaled inputs. The hyperparameters of Random Forest, XGBoost, and ridge regression were tuned by an exhaustive grid search over the values listed in Table 3 [24]; the combination with the lowest mean MAE in 5-fold cross-validation on the training portion was selected and refitted on the whole training set. All models were implemented with the scikit-learn [17] and xgboost [23] libraries, and prediction is executed with the predict function.

Table 3. Hyperparameter search space and selected values (grid search, 5-fold cross-validation)

Model	Hyperparameter	Values searched	Selected
Random Forest	n_estimators	100, 300, 500	500
	max_depth	None, 10, 20	20
	min_samples_leaf	1, 2, 5	5
	max_features	sqrt, 0.5, 1.0	0.5
XGBoost	n_estimators	200, 500	200
	max_depth	3, 5, 7	3
	learning_rate	0.03, 0.1	0.03
	subsample	0.8, 1.0	0.8
	colsample_bytree	0.8, 1.0	1.0
Ridge regression	alpha (α)	0.01, 0.1, 1, 10, 100	10

2.3. Evaluation

In the Random Forest model evaluation stage, new values or data are entered to test whether the Random Forest model can predict prices. The data entered include product weight, production costs, product type, packaging, and season, while the demand and calendar features are taken from the latest sales records and the prediction date. The results of this stage are predicted selling prices and estimated profits based on the newly input data.

Model performance was measured using three metrics. MAE, given in (1), captures the mean absolute gap, in rupiah, between predicted and actual prices; RMSE, given in (2), penalizes large errors more strongly; and R^2, given in (3), reflects the proportion of selling-price variation that the model accounts for [18], [19].

$$MAE = (1/n)\ \Sigma\ |y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{[\ (1/n)\ \Sigma\ (y_i - \hat{y}_i)^2\ ]} \tag{2}$$

$$R^2 = 1 - [\ \Sigma\ (y_i - \hat{y}_i)^2 / \Sigma\ (y_i - \bar{y})^2\ ] \tag{3}$$

where y_i is the actual selling price, $\hat{y}_i$ is the predicted selling price, $\bar{y}$ is the mean of the actual selling prices, and n is the number of test data.

Cross-validation. Generalization was assessed with k-fold cross-validation [25]. The training data were divided into k = 5 shuffled folds (random_state = 42); each model was trained on four folds and validated on the remaining one, and the procedure was repeated so that every fold served once for validation. Results are reported as the mean and standard deviation of MAE, RMSE, and R^2 across the five folds. Because the records form a time series, a forward-chaining split (TimeSeriesSplit, five splits), in which each validation fold lies strictly after its training data, was also applied to the tuned Random Forest as a robustness check [26]. Overfitting was examined by comparing the R^2 obtained on the training set with the R^2 on the test set; a large gap indicates that a model memorizes the training data rather than generalizing. Finally, all models were evaluated once on the held-out 20% test set.

The estimated profit is calculated from the prediction results as in (4), where the profit is the difference between income and production cost:

$$profit = income - cost \tag{4}$$

income = predicted price × amount of production, and cost = cost of production × amount of production.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

2.4. Deployment

During this phase, the trained Random Forest model is embedded in a lightweight web application built with the Streamlit library [27]. Streamlit was preferred because it wraps a machine-learning model into an interactive interface rapidly, one that business owners can use without any coding skills. The application receives new feature inputs (product type, weight, production cost, packaging, season, and production quantity), fills the demand and calendar features automatically from the latest sales records and the prediction date, and displays the predicted selling price, profit estimation, analysis graphs, and a downloadable prediction result document.

2.5. Analyzing

The analysis stage is used to design a profit analysis from sales data. The profit analysis used in this study consists of profit per transaction and profit over time. In addition to profit analysis, this study also uses actual versus predicted price analysis. The analysis is displayed in the form of data distribution graphs.

3. RESULTS AND DISCUSSION

3.1. Results

There are five ways to display the results of this study, namely the accuracy of the baseline Random Forest model, the comparative benchmark with cross-validation and hyperparameter tuning, the prediction of selling prices and profits, the profit analysis and the comparison of predicted and actual prices, and the prediction result documents that can be downloaded from the application.

The first result of this study is the performance of the baseline Random Forest Regression model (five base features, single 80:20 split) on the test data, measured using MAE and R^2. A summary of the model configuration and evaluation results is presented in Table 4.

Table 4. Random Forest Regression model configuration and evaluation results

Parameter / Metric	Value
Number of estimators (n_estimators)	100
Random state	42
Train : test ratio	80 : 20
MAE	Rp 2,031
R^2 Score	0.63
MAE relative to average selling price (±Rp 14,500)	±14%

For the baseline model, the MAE came out to Rp 2,031, whereas the R^2 Score reached 0.63, equivalent to 63%. A lower MAE and a higher R^2 both signal that the Random Forest Regression model predicts new selling prices more precisely [18]. An MAE of Rp 2,031 implies that predictions differ from the true price by about Rp 2,031 on average — close to 14% of the dataset's mean selling price — a margin still usable as a pricing guide for micro-scale enterprises. Meanwhile, an R^2 of 0.63 shows that the five base features account for 63% of the variation in selling price, leaving the other 37% to influences outside the model, possibly buyer demand, holidays, or input prices. This motivated the feature expansion, tuning, and benchmarking reported next.

The second result is the comparative benchmark that positions Random Forest against alternative learners, all trained on the expanded feature set. Table 5 reports the 5-fold cross-validated performance on the training set (mean ± standard deviation) and the performance on the held-out test set.

Table 5. Comparative performance under 5-fold cross-validation and on the hold-out test set (15 features)

Model	CV MAE (Rp)	CV RMSE (Rp)	CV R^2	Test MAE (Rp)	Test RMSE (Rp)	Test R^2	Train R^2	Train–test gap
Linear regression	1,974 ± 66	2,291 ± 57	0.691 ± 0.020	2,078	2,378	0.661	0.699	0.038
Ridge regression (α = 10)	1,973 ± 66	2,290 ± 57	0.691 ± 0.020	2,077	2,377	0.662	0.699	0.038
Random Forest (baseline, 100 trees)	2,025 ± 104	2,393 ± 93	0.662 ± 0.029	2,172	2,500	0.626	0.954	0.329
Random Forest (tuned)	1,997 ± 87	2,342 ± 78	0.677 ± 0.023	2,129	2,448	0.641	0.849	0.208
XGBoost (tuned)	1,991 ± 79	2,322 ± 65	0.682 ± 0.023	2,114	2,429	0.647	0.750	0.103

Note: CV = 5-fold cross-validation on the training set (n = 1,752); test = 20% hold-out set (n = 438). Mean selling price = Rp 15,020.

Hyperparameter tuning improved Random Forest consistently: compared with the untuned baseline, the tuned model lowered the test MAE from Rp 2,172 to Rp 2,129 and the RMSE from Rp 2,500 to Rp 2,448, and raised R^2 from 0.626 to 0.641. The MAE of Rp 2,129 corresponds to 14.2% of the mean selling price. Tuning also addressed overfitting. The untuned forest with fully grown trees reached a training R^2 of 0.954 but only 0.626 on the test set (gap 0.329), whereas constraining the trees (min_samples_leaf = 5, max_depth = 20, max_features = 0.5) reduced the gap to 0.208 and lowered the fold-to-fold variability (MAE ± Rp 87 versus ± Rp 104; R^2 ± 0.023 versus ± 0.029). The forward-chaining TimeSeriesSplit produced comparable results for the tuned Random Forest (MAE Rp 2,021, RMSE Rp 2,363, R^2 0.674), showing that performance is maintained when the model predicts later periods.

Among the tree ensembles, tuned XGBoost was slightly more accurate than tuned Random Forest (test MAE Rp 2,114 versus Rp 2,129; R^2 0.647 versus 0.641) and overfitted

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

less (gap 0.103). However, the linear baselines performed as well as or slightly better than both ensembles: ridge regression obtained the lowest test MAE (Rp 2,077) and the highest R^2 (0.662), with a train–test gap of only 0.038. The differences in cross-validated MAE among all tuned models (at most Rp 24) are much smaller than the fold-to-fold standard deviations (Rp 66–87), so the four models should be regarded as statistically comparable rather than one being clearly superior.

The contribution of the expanded features was evaluated by retraining the tuned Random Forest on different feature sets (Table 6).

Table 6. Effect of feature expansion on the tuned Random Forest

Feature set	No. of features	CV R^2	Test MAE (Rp)	Test RMSE (Rp)	Test R^2	Unexplained variance ($1 - R^2$)
Base features	5	0.674	2,083	2,416	0.651	34.9%
Base + market demand	10	0.675	2,125	2,444	0.642	35.8%
Base + holiday/calendar	10	0.673	2,077	2,397	0.656	34.4%
All features (expanded)	15	0.677	2,129	2,448	0.641	35.9%

Contrary to expectation, the expanded features did not reduce residual variance: the unexplained variance remained between 34.4% and 35.9% for every feature set, and the differences (at most 0.015 in R^2) lie within the cross-validation standard deviation. The permutation importance in Figure 5 explains this result. Production cost is the only influential feature (mean decrease in R^2 of 1.27), whereas all demand, holiday, and calendar features, as well as product type, weight, packaging, and season, have an importance close to zero (below 0.005).

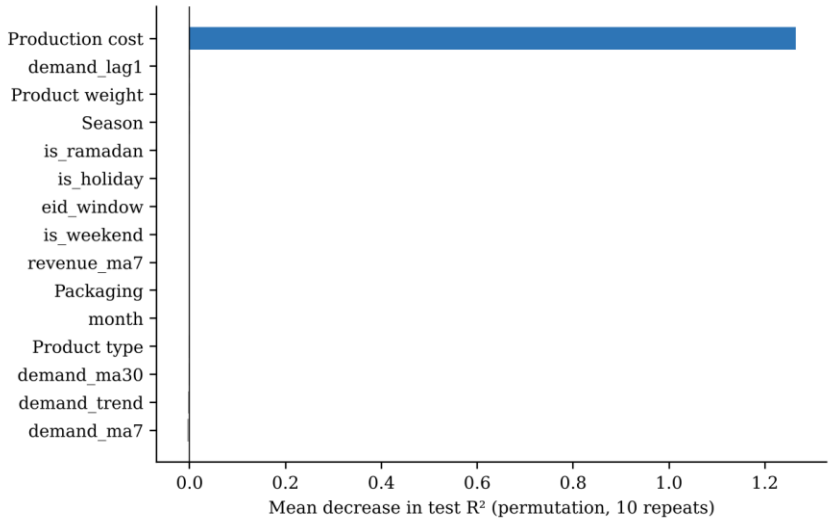


Figure 2. Permutation feature importance of the tuned Random Forest (test set)

A closer inspection of the data confirms this pattern. The gross margin per package (selling price minus production cost) averages Rp 5,957 and is spread almost uniformly between Rp 2,001 and Rp 9,994. It is practically uncorrelated with every observed variable ($|r| \leq 0.035$ for weight, all demand indicators, holidays, Ramadan, and weekends) and hardly differs by product (Rp 5,851–6,040) or season (Rp 5,932–5,979). The selling price therefore follows a cost-plus rule in which producers add a margin that does not respond to the measured market conditions. Production cost alone explains 69.1% of the price variance, which is effectively the upper bound attainable with the available information, and all models in Table 5 operate close to this bound (cross-validated R^2 0.662–0.691). This also explains why the linear models perform as well as the ensembles: the dominant relationship is linear, with each additional rupiah of production cost raising the predicted price by approximately one rupiah.

The third result is the prediction of the selling price and estimated profit using new data. The features entered as new data are product type, product weight, production cost per package, packaging type, season, and production quantity. In this study, the new data used were product type: mushroom chips, product weight 200 grams, production cost per package Rp 8,000, plastic packaging, rainy season, and a production quantity of 100 packages, on a regular (non-holiday) weekday with the demand indicators set to their latest observed values. From these data, the tuned Random Forest predicts a selling price of Rp 14,243 per package, and the estimated profit is Rp 624,300, obtained from (Rp 14,243 − Rp 8,000) × 100 packages, in accordance with (4).

The fourth result is a graph analyzing the actual and predicted prices, as well as graphs analyzing profit per transaction and profit over time. The first graph is a comparative analysis of actual and predicted prices on the test set, shown in Figure 2. The plot displays how actual and predicted selling prices are distributed and reveals how closely predicted values sit next to the real ones; the greater the distance from the diagonal, the poorer the prediction. In Figure 2, the points follow the diagonal reference line but with a roughly constant vertical spread of about ±Rp 4,000 across the whole price range; low prices tend to be slightly over-predicted and high prices slightly under-predicted. This band corresponds to the producer-set margin described above and is consistent with the obtained R^2 of 0.641.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

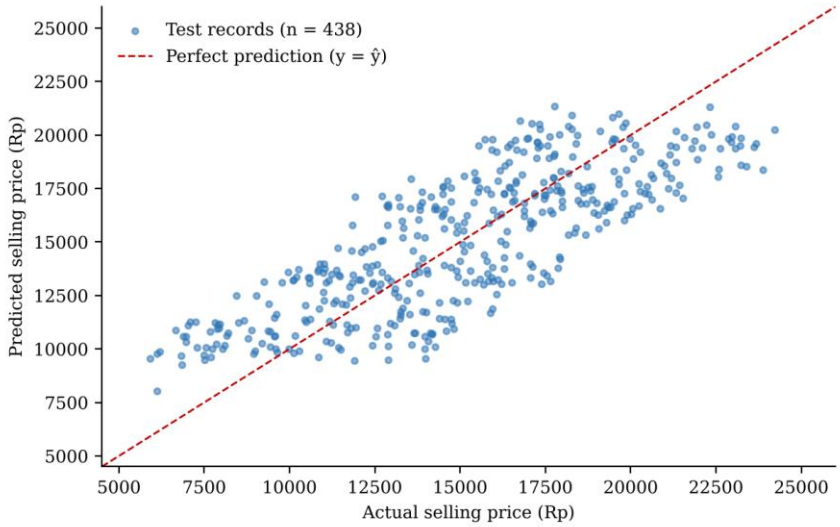


Figure 3. Comparison of actual and predicted prices (tuned Random Forest, test set, n = 438)

The remaining deviations in Figure 2 reflect the subjective component of prices set by producers, which cannot be captured by observable features. Consistent with Probst et al. [16], limiting tree depth and the minimum leaf size controlled variance, while adding more trees gave diminishing returns. Because the hyperparameters have already been optimized and the models operate close to the information bound, further gains are more likely to come from additional information, such as competitor prices or the producer's pricing rationale, than from further tuning.

The next graph is the profit analysis graph. The profit analysis graphs are derived from the 2023–2024 sales data, not from forecasts, and profit is computed per package as the selling price minus the production cost. Figure 3 shows the profit per transaction for each processed mushroom product as stacked bars, so the height of each bar (up to about Rp 29,000) is the combined daily margin of the three products. The individual products contribute similar margins: on average Rp 6,040 for mushroom chips, Rp 5,979 for mushroom nuggets, and Rp 5,851 for mushroom floss per package. Hence no product clearly dominates in margin per package, and product prioritization should also consider sales volume.

Jurnal Teknologi Informasi dan Pendidikan
Volume 20, No. 1, March 2027
https://doi.org/10.24036/jtip.v20i1.1157

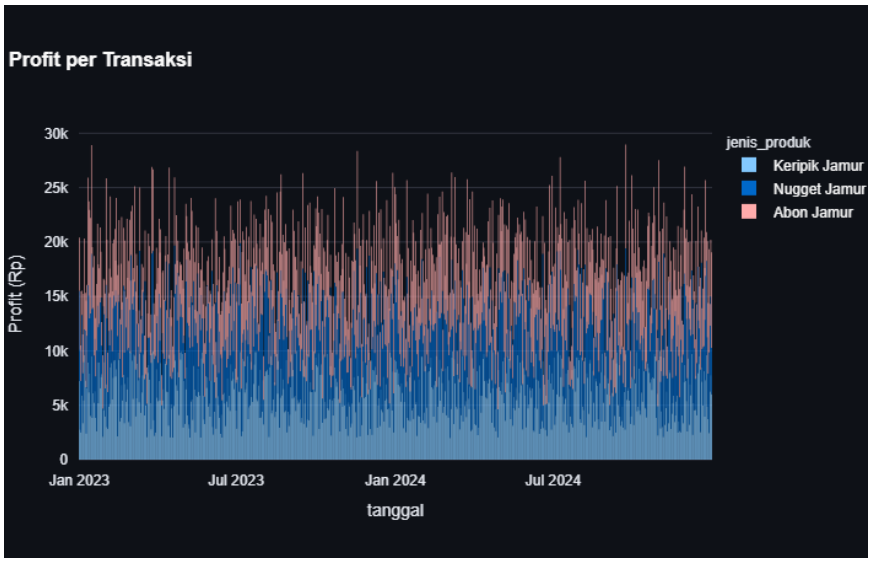


Figure 4. Profit per transaction for each processed mushroom product

The difference between the profit-over-time graph and the per-transaction graph is that the profit-over-time graph does not show the profit for each product, but it shows the development of profit in more detail. The profit-over-time graph is presented in Figure 4. Figure 4 shows that the combined daily margin of the three products fluctuates between Rp 7,760 and Rp 28,964 (mean Rp 17,870) without a downward trend, and the total annual margin was nearly identical in 2023 (Rp 6.50 million) and 2024 (Rp 6.56 million), indicating that the partner's mushroom processing business is relatively stable across rainy and dry seasons.

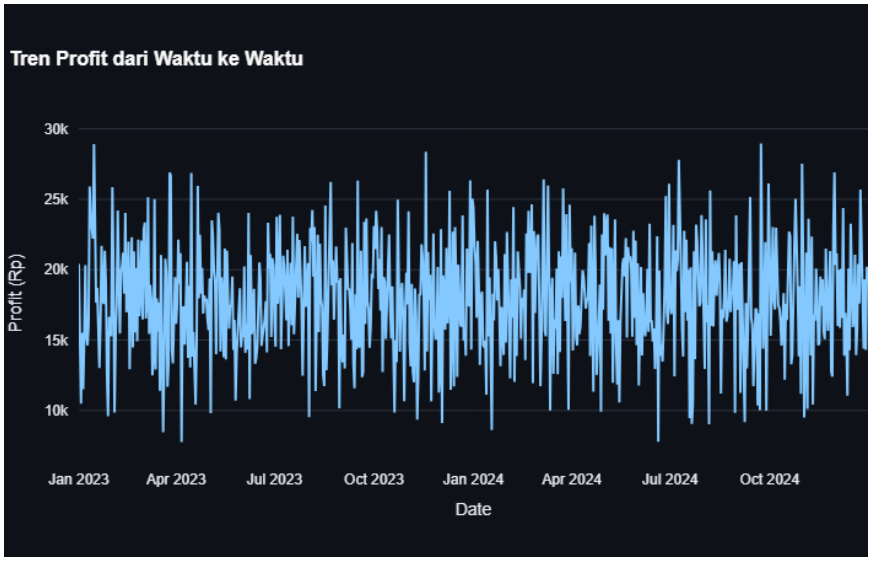


Figure 5. Profit trend over time during 2023–2024

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

3.2. Discussion

The tuned Random Forest (R^2 = 0.641, relative error 14.2%) performs in a range comparable to earlier work on price prediction: Ibrahim obtained a MAPE of 8.71–10.31% when forecasting shallot prices with an optimized Random Forest [10], although the data of this study come from a micro-scale business whose prices are partly set subjectively by the producers. Unlike Muchtar and Afiyati [11] and Pratama et al. [12], who found Random Forest superior to linear regression for rice and shallot prices, the benchmark in Table 5 shows that Random Forest, XGBoost, and the linear baselines are statistically comparable on this dataset. The difference lies in the price-formation mechanism: commodity market prices are shaped by non-linear interactions of supply, demand, and weather, whereas the prices of processed products from a single micro-enterprise are set by a cost-plus rule. Lestari et al. [15] reported R^2 values of up to 0.82 for XGBoost enriched with macroeconomic and climatic features; the present results show that such gains require features that actually drive the price. The unchanged R^2 after feature expansion (0.651 with the base features versus 0.641 with all features) indicates that calendar and demand information does not explain the variance left by the five base features, because the producer's margin does not respond to these conditions.

Relative to earlier research [7], [8], [10]–[14], the principal contribution here is embedding the prediction model within an enterprise application that mushroom business groups can use directly, uniting price forecasting, profit estimation, and profitability visualization within one tool, with accuracy verified through cross-validation and a benchmark against XGBoost and linear models. The benchmark also yields a practical insight: because the models are statistically comparable, the application can use either the tuned Random Forest or the more interpretable ridge regression, and producers can read the prediction as "production cost plus a typical margin of about Rp 6,000 per package". In this way, the research output is not only a model with measured accuracy, but an operational decision-support instrument for the 77 mushroom groups assisted by the Limau Manis village forum.

A number of limitations apply to this study. To begin with, the data originate from just one industry partner across a two-year span, so the model might not transfer well to other areas or market situations. Second, promotion records, raw-material purchase prices, and competitor prices were not available; a raw-material index such as the consumer price index of Padang City [21] and a promotion log should be collected in future work, although their effect is expected to be limited as long as the producer's margin is set independently of market conditions. Third, the demand indicators rely on past sales of the partner itself, so they reflect realized rather than latent demand. Finally, because about 31% of the price variance stems from the producer's discretionary margin, future work should record the reasons behind each pricing decision or model the margin directly, and evaluate additional learners such as support vector regression.

4. CONCLUSION

In line with the aim set out in the introduction, this study built an enterprise system that predicts the selling price of processed mushroom products and evaluates profitability by means of Random Forest Regression. The baseline model, built with 100 estimators and five base features, achieved an MAE of Rp 2,031 (±14% of the average selling price) and an R^2 of 0.63. With the expanded features and grid-search tuning under 5-fold cross-validation, Random Forest achieved a cross-validated MAE of Rp 1,997 and R^2 of 0.677 ± 0.023, and on the hold-out test set an MAE of Rp 2,129, an RMSE of Rp 2,448, and an R^2 of 0.641; tuning reduced the train–test R^2 gap from 0.329 to 0.208, and the forward-chaining validation (R^2 0.674) confirmed that the model generalizes to later periods. In the benchmark, XGBoost (R^2 0.647) and the linear baselines (R^2 0.661–0.662) performed comparably to Random Forest, and the market-demand and holiday features did not reduce residual variance. Both findings are explained by the cost-plus pricing of the partner: production cost explains 69% of price variation, while the remaining margin (Rp 2,001–9,994 per package) is set by producers independently of the observed conditions. The model was deployed as a Streamlit-based web application that produces price predictions (e.g., Rp 14,243 per package of mushroom chips with an estimated profit of Rp 624,300 for 100 packages) and visual analyses of profit per transaction and profit over time, which show similar margins for the three products and a stable business profit. With this system, mushroom business groups in Limau Manis Village can determine selling prices in a more objective, data-driven manner. Future work should collect promotion, raw-material, and competitor-price data and extend the dataset to several business partners so that the model generalizes better.

REFERENCES

[1] Badan Pusat Statistik, Statistik Hortikultura 2022. Jakarta, Indonesia: BPS RI, 2023.

[2] M. Fafchamps, "Cash crop production, food price volatility, and rural market integration in the third world," Am. J. Agric. Econ., vol. 74, no. 1, pp. 90–99, 1992.

[3] M. K. Mohanty, P. K. G. Thakurta, and S. Kar, "Agricultural commodity price prediction model: A machine learning framework," Neural Comput. Appl., vol. 35, pp. 15109–15128, 2023, doi: 10.1007/s00521-023-08528-7.

[4] K. Srivastava, R. Sharma, and P. Singh, "House price prediction using machine learning," in Proc. 3rd Int. Conf. Advances in Computing, Communication Control and Networking (ICAC3N), 2021, pp. 203–206, doi: 10.1109/ICAC3N53548.2021.9725552.

[5] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[6] T. K. Ho, "Random decision forests," in Proc. 3rd Int. Conf. Document Analysis and Recognition, Montreal, QC, Canada, 1995, pp. 278–282, doi: 10.1109/ICDAR.1995.598994.

[7] M. Rakhra et al., "Crop price prediction using random forest and decision tree regression: A review," Mater. Today Proc., 2021, doi: 10.1016/j.matpr.2021.03.261.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

[8] R. K. Paul et al., "Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India," PLOS ONE, vol. 17, no. 7, e0270553, 2022, doi: 10.1371/journal.pone.0270553.

[9] Z. Chen, H. S. Goh, K. L. Sin, K. Lim, N. K. H. Chung, and X. Y. Liew, "Automated agriculture commodity price prediction system with machine learning techniques," Adv. Sci. Technol. Eng. Syst. J., vol. 6, no. 4, pp. 376–384, 2021.

[10] M. N. R. Ibrahim, "Price forecasting of shallots using the machine learning approach of random forest regression supporting price stabilization," J. Keteknikan Pertanian, vol. 13, no. 3, pp. 449–461, 2025.

[11] I. R. Muchtar and A. Afiyati, "Comparison of linear regression and random forest algorithms for premium rice price prediction (case study: West Java)," J. Indones. Sos. Teknol., vol. 5, no. 7, pp. 3122–3132, 2024, doi: 10.59141/jist.v5i7.1184.

[12] M. A. Pratama, M. Munawaroh, and W. J. Pranoto, "Perbandingan performa algoritma linear regresi dan random forest untuk prediksi harga bawang merah di Kota Samarinda," Tektonik, vol. 3, no. 2, p. 172, 2024, doi: 10.62017/tektonik.

[13] E. S. Lestari and I. Astuti, "Penerapan random forest regression untuk memprediksi harga jual rumah dan cosine similarity untuk rekomendasi rumah pada Provinsi Jawa Barat," J. Ilmiah FIFO, vol. 14, no. 2, p. 131, 2022, doi: 10.22441/fifo.2022.v14i2.003.

[14] A. N. Rais, Warjiyono, I. Alfarobi, S. Wira Hadi, and W. Kurniawan, "Analisa prediksi harga jual rumah menggunakan algoritma random forest machine learning," J. Riset Sist. Inf. Teknol. Inf. (JURSISTEKNI), vol. 6, no. 2, pp. 416–423, 2024.

[15] D. R. Lestari, E. A. S. Bangun, F. L. Gaol, and T. Matsuo, "Machine learning-based forecasting of agricultural commodity prices using ensemble models," J. Hum. Earth Future, vol. 6, 2025.

[16] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," WIREs Data Min. Knowl. Discov., vol. 9, no. 3, e1301, 2019, doi: 10.1002/widm.1301.

[17] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[18] 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, e623, 2021, doi: 10.7717/peerj-cs.623.

[19] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not," Geosci. Model Dev., vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.

[20] Menteri Agama, Menteri Ketenagakerjaan, and Menteri PANRB Republik Indonesia, Keputusan Bersama tentang Hari Libur Nasional dan Cuti Bersama Tahun 2023 (beserta perubahannya) dan Tahun 2024 (No. 855, 3, 4 Tahun 2023). Jakarta, Indonesia, 2022–2023.

[21] Badan Pusat Statistik Kota Padang, "Perkembangan Indeks Harga Konsumen Kota Padang, Januari–Desember 2024," Berita Resmi Statistik, Padang, Indonesia, 2024.

[22] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970, doi: 10.1080/00401706.1970.10488634.

[23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[24] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281–305, 2012.

[25] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. 14th Int. Joint Conf. Artificial Intelligence (IJCAI), 1995, pp. 1137–1143.

[26] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Inf. Sci., vol. 191, pp. 192–213, 2012, doi: 10.1016/j.ins.2011.12.028.

[27] Streamlit Inc., "Streamlit: A faster way to build and share data apps," 2024. [Online]. Available: https://streamlit.io.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id