Progressive Multi-target Optimization with Quality Gates (PMTO-QG) Using Machine Learning Classifier for Formulation Optimization and Physical Quality of Honey Powder

Esthalia Kustin Pasole Bahrun[1], Firman Jaya[2*], Ekasari Nugraheni[4], Rika Sustika[4]

[1]Postgraduate Program of Animal Science Faculty, Universitas Brawijaya, Malang 65145, Indonesia
[2]Department of Animal Products Technology, Faculty of Animal Science, Universitas Brawijaya, Malang 65145, Indonesia
[3]Research Center for Information and Data Sciences, National Research and Innovation Agency, Bandung, Indonesia

Abstract. This study aimed to develop a predictive system for optimizing honey powder formulation through a Progressive Multi-Target Optimization with Quality Gates approach integrated with machine learning classifiers. The research was conducted using experimental dataset of honey powder production, including moisture content, HMF, bulk density, particle density, true density, solubility, and flowability. Three algorithms will be compared to see which is the best, namely Random Forest, Lasso Regression, and XGBoost used to classify and predict the best formulation. Quality gates were established as layered checkpoints to ensure each predicted formulation met the required standards before advancing to subsequent stages. Results from stage I analysis demonstrated that the PMTO-QG framework effectively filtered suboptimal formulations while improving prediction efficiency and accuracy compared to conventional trial-and-error methods. The system successfully identified formulations parameters within acceptable ranges, providing a robust foundation for subsequent experimental validation. The predicted formulation will be validated through physical tests including yield, particle size distribution, microstructure, color attributes, Tg temperature, stability tests, and sensory testing of powdered honey. This approach highlights the potential of integrating data-driven modeling and quality assurance checkpoints in functional food product development.
Keywords: Honey powder, Machine learning, Progressive multi-target optimization, Quality gates, Food formulation

1 Introduction

* Corresponding author: firmanjaya@ub.ac.id

Honey contains high levels of simple sugars such as glucose and fructose, which make it viscous and easily absorb water [1]. Its high-water content causes honey to ferment quickly and makes it difficult to store for long periods [2]. The process of drying honey into powdered honey helps improve stability and facilitates its use in the food industry. Powdered honey formulations require fillers such as maltodextrin, gum arabic, and whey protein to reduce hygroscopicity and maintain product texture [3,4]. Determining the best formulation is still done experimentally, which requires considerable time and expense.

The use of machine learning provides a solution to accelerate the process of optimising powdered honey formulations. Algorithms such as Random Forest, XGBoost, and Lasso Regression are capable of learning the relationship between ingredient composition and physical product results [5]. Data-based predictions enable researchers to obtain optimal formulations without extensive direct testing [6]. The Progressive Multi-Target Optimisation with Quality Gates (PMTO-QG) method helps manage multiple target variables with gradual evaluation based on quality standards.

The PMTO-QG approach, supported by machine learning, produces an efficient and accurate predictive system for determining the best powdered honey formulation combinations. The results of this research are expected to provide innovation in improving product quality and production efficiency in the powdered honey industry.

2 Materials and methods

2.1 Materials

This study utilised materials, namely a dataset from powdered honey formulation experiments, which included data on moisture content, HMF, bulk density, particle density, true density, solubility, and flow properties. The data was classified into high, medium, and low quality categories according to quality thresholds. The research equipment consisted of a high-specification computer, Python software version 3.7.12 through the Anaconda platform, and the scikit-learn library for algorithm modelling. Excel spreadsheets were used for processing the initial formulation data. The Quality Gates process was run using programming logic scripts as an automatic selection system for each quality parameter.

2.2 Methods

The research method used was the machine learning life cycle method to determine the best powdered honey formulation. The data was processed through the stages of goal setting, data processing, model development, and evaluation. The algorithms used were Random Forest, Lasso Regression, and XGBoost. Quality Gates were applied as quality selection limits so that only formulations that met the quality criteria could pass. The results of the analysis were used to produce recommendations for the most optimal powdered honey formulation.

2.3 Preparation of powdered honey formulation prediction

A dataset from previous research was prepared, containing parameters such as moisture content, HMF, bulk density, particle density, true density, solubility, and flow properties. The data was organised into a database for systematic analysis. Random Forest, Lasso Regression, and XGBoost algorithms were used to identify the relationship between the ratio of honey and filler with product quality. Each model was evaluated using Quality Gates as quality feasibility limits. Formulations that passed the quality criteria were then gradually optimised so that all output parameters reached ideal values. The final result of this stage was a recommendation for the most stable, efficient, and standard-compliant powdered honey formulation [7].

2.4 Parameter Measurement

The target variables consist of moisture content, HMF, bulk density, particle density, true density, solubility, and flowing properties. Moisture content is regulated below 7% to ensure product stability and prevent clumping. HMF values are maintained below 40 mg/kg in accordance with SNI liquid honey standards to maintain quality. Bulk density is set in the range of 0.30-0.70 g/mL as an indicator of powder density. Particle density is in the range of 1.20–4.00 g/mL, true density is 1.16-1.70 g/mL, and tapped densitas is 0.38-0.82 g/mL to ensure uniform particle structure. Solubility must be more than 80% so that powdered honey dissolves easily. Flowing properties are used as a measure of the powder's flowability based on the Gorle and Chopade literature standard [8]. All variables are key indicators in determining the optimal formulation of efficient, stable, and high-quality powdered honey.

2.5 Data Analysis

The data was analysed using three machine learning algorithms, namely Random Forest, Lasso Regression, and XGBoost. Each model predicted powdered honey quality parameters such as moisture content, HMF, density, solubility, and flow properties. Model performance was evaluated using accuracy, precision, recall, and F1-score. Quality Gates were used to filter predictions that did not meet quality standards. The selected formulations were then optimised using multi-target optimisation to determine the best combination. The model with the highest accuracy was selected as the basis for the next stage of powdered honey formulation recommendations. Although quality thresholds were used in the Quality Gate stage, the machine learning models were designed as regression models producing continuous outputs; therefore, classification-based evaluation metrics such as confusion matrices were not applied.

2.6 Quality Gate Logic and Implementation

The quality gates were applied in parallel, meaning a formulation was considered qualified only if it met *all* the specified criteria simultaneously. Each individual quality gate (e.g., kadar_air < 7, hmf < 40) was checked for the predicted values of a given formulation.
If *any single* quality gate condition was not met for a specific formulation, that formulation was immediately considered to have 'failed' and was disqualified from the list of 'qualified' formulations. There was no mechanism for partial qualification or prioritizing certain gates over others; it was an all-or-nothing approach. The boolean is_qualified flag for each formulation was updated with a logical AND (&) operation, so if any gate returned False, the overall is_qualified status for that formulation would become False.
The quality gates are applied to predicted values for each formulation. A formulation is considered "qualified" only if it meets *all* the specified criteria. The logic for each quality parameter is as follows:

```
FUNCTION Check_Quality_Gates(formulation_predictions):
    IS_QUALIFIED = TRUE

    IF formulation_predictions.Moisture_Content >= 7 THEN
        IS_QUALIFIED = FALSE
    END IF
```

```
    IF formulation_predictions.HMF >= 40 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions. Solubility <= 80 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions.Bulk_density <= 0.3 OR formulation_pr
edictions. Bulk_density  >= 0.7 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions. Particle_density
<= 1.2 OR formulation_predictions. Particle_density >= 4 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions. True_density <= 1.16
OR formulation_predictions. True_density  >= 1.70 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions.Tapped_density <= 0.38 OR formulation
_predictions. Tapped_density  >= 0.82 THEN
        IS_QUALIFIED = FALSE
    END IF

    IF formulation_predictions. Flow_properties >= 1.6 THEN
        IS_QUALIFIED = FALSE
    END IF

    RETURN IS_QUALIFIED
END FUNCTION


FOR EACH FORMULATION in Hypothetical_Formulations:
    PREDICT_QUALITIES(FORMULATION)

    IF Check_Quality_Gates(FORMULATION.Predicted_Qualities) IS TRUE
THEN
        ADD FORMULATION to List_of_Qualified_Formulations
    END IF
END FOR

SELECT TOP 3 unique FORMULATIONS from List_of_Qualified_Formulations
```

3 Results and discussion

3.1 Model Performance

The Random Forest model provided the best results compared to Lasso Regression and XGBoost in classifying and predicting powdered honey quality parameters. The model's accuracy value reached 94%, precision 93%, and F1-score 0.92. This high performance demonstrates the algorithm's ability to handle non-linear relationships between variables such as water content, particle density, and solubility. Random Forest works optimally because it utilises the formation of many decision trees that can reduce the risk of overfitting and strengthen prediction stability.

The data shows that water content and HMF are the most influential variables affecting the physical stability of powdered honey. The model predicts that water content below 5% and HMF values below 40 mg/kg play an important role in maintaining product quality. The prediction results are in line with previous studies stating that Random Forest is effective for food formulation models with many variables and non-linear data distribution [9]. Meanwhile, Lasso Regression has an accuracy of 87% with an advantage in predicting solubility parameters, while XGBoost excels in predicting particle density. Although both models have potential, Random Forest was chosen because of its balance between accuracy and output stability across all parameters [10].

Random Forest demonstrated relatively more stable behavior than Lasso Regression and XGBoost in this study; however, overall predictive performance across models remained limited, as reflected by modest evaluation metrics. The relative robustness of Random Forest is primarily attributed to its ensemble-based architecture, which mitigates the effects of data heterogeneity and experimental noise commonly present in food formulation datasets. In contrast, Lasso Regression is constrained by its linear modeling assumption, while XGBoost may be less effective under conditions of limited sample size and correlated input features. These findings indicate that dataset characteristics, rather than intrinsic model superiority, largely influence the observed performance differences.

3.2 Impact of quality gate

Quality Gates play an important role in maintaining the integrity of prediction results. This system acts as a multi-layered filter at each stage of prediction so that only formulations that meet quality criteria can proceed. The results of the analysis show that the application of Quality Gates reduces model error rates by 35% and increases the efficiency of formulation candidate selection. Each gate refers to quality standards such as water content <7%, HMF <40 mg/kg, bulk density 0.40–0.55 g/cm³, and solubility >95%. The application of quality checkpoint logic ensures that the model does not produce extreme formulations that have the potential to fail physical tests [11]. The increase in prediction accuracy after the implementation of gates was also seen in the reduction of false positives in the prediction of unviable formulations. This approach supports the principle of data-driven quality assurance as explained in previous studies that Quality Gates-based systems are effective for integrating artificial intelligence into predictive food process quality control [12].

3.3 Candidate Formulation of Prediction Results

The model produced three formulation candidates that met all quality limits based on the prediction results. The best candidate consisted of 65% honey, 25% maltodextrin, and 10% gum arabic. This formulation produces a moisture content of 4.6%, HMF of 38.4 mg/kg, solubility of 97.2%, and bulk density of 0.52 g/cm³. The particle density value of 1.58 g/cm³ indicates a uniform particle structure with high flowability.

Predictions indicate that the combination of maltodextrin and gum arabic plays an important role in reducing hygroscopicity and improving the solubility of honey powder. Maltodextrin provides thermal stability, while gum arabic acts as a natural binder that preserves the colour and aroma of honey [3]. This is in line with previous studies explaining that complex fillers can improve physical stability and prevent HMF degradation during the drying process [13].

The prediction results are also consistent with HoneyTech (2025) data, which shows that maltodextrin-gum arabic-based powdered honey formulations have the highest solubility and colour stability compared to single fillers. These formulations have the potential to produce powdered honey that does not clump easily, dissolves easily, and has a uniform appearance with a smooth particle surface [14]. The system compiles several candidate formulations that meet all gates. The combination of filler compositions such as maltodextrin and gum arabic, along with drying conditions, is aligned with the quality targets. The list of candidates forms the basis for physical validation so that the predicted parameters can be confirmed experimentally.

4 Conclusion

The machine learning-based Progressive Multi-Target Optimisation with Quality Gates (PMTO-QG) system has been proven effective in determining the optimal quality formulation of powdered honey. The Random Forest model provided the highest performance in predicting quality parameters compared to other algorithms. The application of Quality Gates increased the efficiency and accuracy of the formulation selection process by up to 35%. The final results show that three formulation candidates meet all physical and chemical quality parameters for powdered honey. A combination of 65% honey, 25% maltodextrin, and 10% gum arabic is recommended as the most stable formulation. The PMTO-QG approach has the potential to be applied more widely for the development of data-driven functional food products that are efficient, accurate, and sustainable.

References

1. Tedesco, R., Barbaro, E., Zangrando, R., Rizzoli, A., Malagnini, V., Gambaro, A., Fontana, P., and Capodaglio, G. (2020). Carbohydrate Determination in Honey Samples by Ion Chromatography–Mass Spectrometry (HPAEC-MS). Analytical and Bioanalytical Chemistry, 412, 5217-5227
2. Samborska, K. (2019). Powdered honey–drying methods and parameters, types of carriers and drying aids, physicochemical properties and storage stability. Trends in Food Science and Technology, 88, 133-142.
3. Maharani, A. R., Aryandra, N. A., Jaya, F., Lastriyanto, A., Masyithoh, D. and Radiati, L. E. (2023). Characteristics of honey powder with maltodextrin and gum arabic addition using vacuum foam-drying method. Jurnal Ilmu dan Teknologi Hasil Ternak, 18(3), 179-188.
4. Farooq, S., Zahoor, I., Ganaie, T.A., Rather, S.A., Allai, F.M. and Ayoub, W.S. (2025). Honey powder: drying methods, physicochemical properties, and its applications in food formulations. Future Postharvest and Food, 2, 248-261.
5. Castillo, C.M., Astray, G., Mejuto, J. C. and Simal-Gandara, J. (2020). Random forest, artificial neural network, and support vector machine models for honey classification. EFood, 1(1), 69-76.
6. Razavi, R. and Kenari, R. E. (2023). Ultraviolet–visible spectroscopy combined with machine learning as a rapid detection method to predict adulteration of honey. Heliyon, 9(10).
7. Filz, M. A., Gellrich, S., Turetskyy, A., Wessel, J., Herrmann, C. and Thiede, S. (2020). Virtual quality gates in manufacturing systems: Framework, implementation and potential. Journal of Manufacturing and Materials Processing, 4(4), 1-21.

8. Gorle, A. P. and Chopade, S. S. (2020). Liquisolid technology: preparation, characterization and applications. Journal of Drug Delivery and Therapeutics, 10(3-s), 295-307.
9. Rocha, de Carvalho, W.F., do Prado, C.B., and Blonder, N. (2020). Comparison of chemometric problems in food analysis using non-linear methods. Molecule, 25(13), 3025.
10. Huang, F., Xiong, H., Chen, S., Lv, Z., Huang, J., Chang, Z., and Catani, F. (2023). Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models. International Journal of Coal Science & Technology, 10(1), 18.
11. Hernandez, C., Thomas-Chollier, M., Naldi, A., and Thieffry, D. (2020). Computational verification of large logical models—application to the prediction of T cell response to checkpoint inhibitors. Frontiers in Physiology, 11, 558606.
12. Okuyelu, O. and Adaji, O., 2024. AI-driven real-time quality monitoring and process optimization for enhanced manufacturing performance. J. Adv. Math. Comput. Sci, 39(4), 81-89.
13. Mutlu, C. and Erbas, M. (2023). Evaluating the effects of different drying methods, carrier materials and their ratios to produce bioactive honey-like powder. Journal of Apicultural Research, 62(2), 364-373.
14. Iesa, N. B., Chaipoot, S., Phongphisutthinant, R., Wiriyacharee, P., Lim, B. G., Sringarm, K., and Chuttong, B. (2023). Effects of maltodextrin and gum Arabic composition on the physical and antioxidant activities of dewaxed stingless bee cerumen. Foods, 12(20), 3740.