

Machine Learning Applications for Predicting Brewery By-Products in the Context of Circular Economy

Hoang Khanh Van , Lam Hong Thanh *



Use your smartphone to scan this QR code and download this article

University of Economics and Law,
Faculty of Information Systems, Vietnam
National University, Ho Chi Minh City
700000, Vietnam

Correspondence

Lam Hong Thanh, University of Economics and Law, Faculty of Information Systems, Vietnam National University, Ho Chi Minh City 700000, Vietnam

Email: thanhhlh@uel.edu.vn

History

- Received: 09/01/2025
- Revised: 27/10/2025
- Accepted: 09/04/2026
- Published Online: 05/05/2026

DOI: <https://doi.org/10.32508/vnuhcmjebvl.v10i2.1549>



Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



ABSTRACT

It can be seen that the brewing industry generates significant by-products such as spent grains, hops, and yeast, which are often underutilized or disposed of inefficiently despite their potential value. To align with the principles of the Circular Economy (CE) principles, which requires manufacturers to keep resources in use for as long as possible, breweries must transform by-products into value-added resources, thereby contributing to sustainable development by reducing environmental impacts and resource depletion. While brewing factories recognize the importance of waste management and by-product reuse, there is a lack of comprehensive studies on predicting by-product quantities. This gap hinders efforts to optimize resource usage and reduce waste disposal.

To address this challenge, this paper aims to optimize waste management and enhance the reuse of brewing by-products, contributing to the sustainability goals of the CE. Machine learning (ML) models are applied to predict by-product quantities. The models' performance is then evaluated using regression metrics, with Gradient Boosting emerging as the most accurate, while. At the same time, the dashboard is employed to visualize the predicted and actual by-product data. The study uses real-world data, covering multiple production variables over a five-year period, ensuring both five years, ensuring industrial relevance and empirical validity. The research not only assesses prediction accuracy but also assesses prediction accuracy and interprets how predictive insights can be integrated into daily brewery operations through dashboard analytics, enabling data-driven decision-making.

The resulting insights enable breweries to forecast waste volumes, adjust production, optimize operations and reduce inefficiencies. Furthermore, this approach provides a scalable framework for other food and beverage manufacturers seeking to apply predictive analytics for sustainability purposes. The study also highlights the potential for data-driven models to improve resource management in the brewing industry, fostering sustainable practices and advancing the goals of the CE.

Key words: Machine learning, brewery by-products, Circular Economy, Gradient Boosting Regression, sustainability, brewing industry, by-product reuse, spent grains, spent hops, spent yeast

INTRODUCTION

Results reveal that brewing operations generate large amounts of by-products, including spent grain, hops and yeast, which are often underutilized despite their potential value¹. For instance, every 100 litres of beer production yields roughly 20 kg of spent grain plus 0.3 kg each of spent yeast and hops¹. In the linear economic model, characterized by the “take, make, dispose” paradigm, most breweries use these solid residues as low-value animal feed or waste them, leading to crucial environmental management issues¹. Instead, CE principles emphasize converting such outputs into resources to reduce environmental impact¹. Yet most valorization research to date has focused on post-production uses, such as animal feed and bioenergy; however, there remains a lack of studies forecasting waste volumes in advance^{2,3}. For this

reason, recent industry reports note that breweries are beginning to adopt sensors and data analytics to optimize production and resource use⁴. Nonetheless, there remain few data-driven tools for anticipating the quantities of brewing by-products that could inform waste-reduction and reuse strategies.

As a result, this study addresses the mentioned gap by developing an ML-based predictive framework that estimates by-product quantities from brewery production parameters and embeds them in an operational dashboard for adaptive decision-making, facilitating their reuse across industries. Unlike prior works, this approach does not only evaluates model accuracy but also investigates how these models can be strategically integrated into brewery decision-making, using an operational dashboard to inform waste management and reuse strategies.

Cite this article : H K V, L H T. Machine Learning Applications for Predicting Brewery By-Products in the Context of Circular Economy. VNUHCM J. Econ. Bus. Law 2026; 10(2):6498-6507.

Methodologically, this paper employs a quantitative and data-driven approach that combines aggregated industry data to train and validate multiple regressors, including Decision Tree, Random Forest, k-Nearest Neighbors (kNN) and Gradient Boosting. These models are evaluated using error metrics (Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-Squared (R^2), while an integration scheme is proposed to integrate predictive insights into brewery operations. The results show that Gradient Boosting Regression provided the best accuracy, though all models yielded modest R^2 (near near-zero or slightly negative) results due to data variability. The insights from this work demonstrate the feasibility of ML for anticipating waste streams and enabling breweries to adjust production plans accordingly, advancing CE practices in the industry. The dashboard prototype also demonstrates how real-time inputs drive model predictions and reuse recommendations.

The paper's contribution is threefold by mentioning a novel predictive tool tailored to brewery by-products, along with a concrete integration strategy via dashboard deployment, and also interpreting empirical insights into model performance and practical implications for CE adoption.

LITERATURE REVIEW:

2.1 Background theories:

Beer production involves multiple stages - malting, mashing, lautering, boiling, fermentation, maturation, and packaging - each producing by-products such as brewer's spent grain (BSG), hops, and yeast. In particular, BSG, which consists predominantly of grain husks, endosperm fragments and pericarp, accounts for approximately 85% of total brewery residues, with yields averaging 19 - 20 kg per 100 L of beer⁵. It is rich in fiber, proteins and other nutrients, making it a promising candidate for reuse in food, feed and industrial applications^{6,7}. In contrast, spent hops, containing around 85% of the hop input not retained in the final product, present reuse challenges due to bitterness⁸. Surplus yeast, while less abundant (approximately 10% of by-products), is rich in proteins, amino acids and bioactive compounds, enabling niche reuse opportunities despite compositional variability⁹.

Existing research has explored numerous valorization pathways for these by-products, including incorporating BSG into bakery products³, producing biogas through anaerobic digestion¹⁰, and extracting antioxidants from hops². In turn, these approaches are

primarily reactive, focusing on post-production reuse rather than forecasting waste volumes.

Moreover, studies in the brewing industry highlight the importance of aligning by-product management with CE principles¹¹ to encourage closing material loops, but few have focused on developing data-driven tools to estimate by-product quantities in advance. As a result, understanding the forecast of by-product generation before the production process enables breweries to match supply with reuse pathways, reducing disposal and improving resource efficiency. This relationship between process variables and waste outputs forms the theoretical basis for applying ML models to estimate by-product volumes and guide reuse decisions consistent with CE objectives.

2.2. Related works:

Machine learning has been increasingly applied in the food and beverage sector for crop yield prediction, quality control, and food waste management^{12,13}. Within brewing, digital technologies support process optimization, sensory monitoring, and energy efficiency¹.

With regard to by-products, previous research remains centered on valorization pathways rather than predictive modeling. Specifically, Amoriello and Ciccioritti¹ reviewed the nutritional properties of BSG, while Belardi et al³ demonstrated its incorporation into bakery products. Similarly, Pehlivani et al² examined anaerobic digestion of brewery residues for biogas generation, and Knez et al¹³ investigated the recovery of antioxidants from spent hops. These studies demonstrate the potential of brewery by-products but do not address how their quantities can be forecasted in advance to facilitate reuse planning.

In parallel, ML-based forecasting models have been successfully applied in related fields such as predicting agricultural by-products¹⁴, food waste generation¹⁵ and biogas yields¹⁶, suggesting strong potential for brewery applications.

As a result, this study extends the literature by introducing a supervised regression framework linking production parameters with by-product outputs. In this way, the paper connects valorization-focused research with predictive analytics, thereby contributing methodological innovation and generating practical insights to support CE adoption in the brewing industry.

MATERIALS - METHODS:

3.1 Data:

The dataset utilized for this study was derived from internal production records provided by Heineken Vietnam, spanning the period from 2020 to 2024. Initially, the data consisted of 9,440 raw observations

collected from 717 breweries. This dataset captures comprehensive details about the brewing process, ingredient usage, by-product generation and brewery-specific characteristics. After data cleaning, the study retained 501 valid observations from brewery records. Subsequently, the infinite variables and missing value data were removed, resulting in a final set of 500 valid data. The five-year collection period was deliberately chosen to reflect contemporary brewing practices and technological advancements, while also capturing seasonal variability. This temporal scope enhances the robustness of the dataset by accounting for fluctuations and anomalies, thereby providing a solid basis for predictive modeling to advance the objectives of the CE in the brewing industry.

The confidential dataset was obtained from Heineken Vietnam’s internal production management system and is confidential. The company formally granted permission for permitted the dataset to be used in this research. Aggregated summary statistics and anonymized features used in model development are presented in the Appendix (Tables 1 - 2 and Figures 2 - 3). Access to the primary data may be considered by Heineken Vietnam Heineken Vietnam may consider access to the primary data upon reasonable request and in accordance with corporate data-sharing policies.

3.2. Research pipeline:

Figure 1 illustrates the overall data analysis process, which is a systematic and structured approach to prepare, analyze and model the dataset to address the prediction of brewing by-products and their optimal reuse.

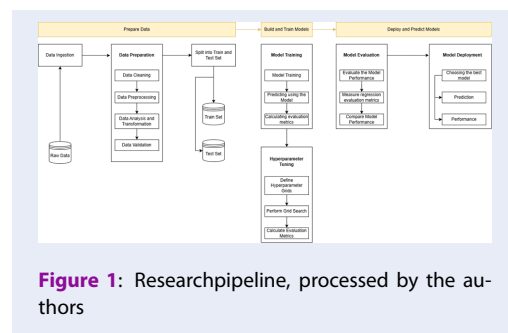


Figure 1: Research pipeline, processed by the authors

Specifically, after the stage of data ingestion, where raw data from brewery operations, including production volumes and by-product quantities, are collected and stored in a structured format, the pipeline begins with data cleaning and preprocessing, where missing values were addressed using mean imputation for numerical variables and mode imputation for categorical variables. Outliers were identified and capped

using interquartile range (IQR) methods to minimize their influence on model performance. Additionally, as Table 1 illustrates, numerical features, such as by-product volumes and production metrics, were normalized to ensure uniform scaling across variables. Also, as detailed in Table 2, categorical variables, including brewery size, seasonality and country, were encoded using one-hot encoding to enhance model compatibility.

Table 2: Statistics of categorical variables, processed by the authors

Categorical Variables	count	unique	top	freq
Seasonality	500	4	Summer	140
Brewery Size	500	3	Medium	187
Country	500	219	Korea	9

Following data preprocessing, exploratory data analysis (EDA) was conducted to uncover patterns and relationships within the dataset. Univariate analysis provided descriptive statistics for key variables, such as by-product quantities and production volumes, while multivariate analysis examined correlations between features. For instance, a strong positive correlation (0.68), as revealed in Figure 2, was observed between production volume and spent grains, highlighting their interdependence.

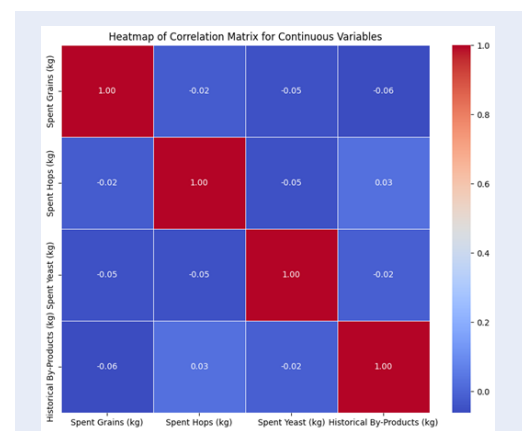


Figure 2: Correlation matrix of continuous variables, processed by the authors

Once the data is prepared, it is then split into training and testing subsets in an 80:20 ratio to ensure robust model training while preserving sufficient data for unbiased evaluation. This division supports the

Table 1: Statistics of key descriptive statistics for numerical variables, processed by the authors

Numerical Variables	count	mean	std	min	25%	50%	75%	max
Amount of Hops Used (kg)	500.000	27.826	13.137	5.220	16.740	28.255	39.468	49.970
Amount of Malt Used (kg)	500.000	563.781	252.815	100.640	350.470	575.525	774.668	999.950
Amount of Yeast Used (kg)	500.000	10.390	5.560	1.020	5.663	10.135	15.068	19.930
Water Volume (liters)	500.00	5495.431	2578.017	1001.610	3355.715	5396.825	7722.795	9995.790
Fermentation Time (days)	500.000	9.618	2.847	5.000	7.000	10.000	12.000	14.000
Temperature)	500.000	21.564	1.960	18.000	19.900	21.700	23.100	25.000
Spent Grains (kg)	500.000	449.436	201.918	80.490	285.730	454.180	621.445	798.610
Spent Hops (kg)	500.000	13.696	6.697	2.060	8.080	13.495	19.735	24.990
Spent Yeast (kg)	500.000	8.867	3.381	3.030	5.895	8.935	11.643	14.980
Wastewater Volume (liters)	500.000	1108.020	514.920	200.730	653.135	1081.555	1550.860	1992.960
Production Volume (liters)	500.000	2672.560	1304.430	504.020	1531.388	2636.455	3798.453	4997.850
Number of Batches	500.000	2.450	1.104	1.000	1.000	2.000	3.000	4.000
Market Demand	500.000	0.976	0.272	0.500	0.750	0.975	1.183	1.500

development of predictive models capable of generalizing unseen data. During the model training phase, four ML models - Decision Trees, Random Forest, kNN and Gradient Boosting Regression - were subsequently trained to predict by-product quantities. Hyperparameter tuning is then performed to optimize the predictive models, enhancing their accuracy and performance, with parameters such as max depth, learning rate, and the number of neighbors refined using grid search techniques.

Finally, the model deployment phase focuses on integrating the chosen predictive model into brewery operations via the dashboard, which is utilized to forecast by-product quantities, thereby providing actionable insights to support breweries in planning reuse strategies effectively and facilitating circular resource management. This deployment not only facilitates waste reduction but also supports the broader adoption of CE principles by optimizing resource utilization and reducing environmental impact.

3.3. Predictive modelling:

Four supervised regression algorithms were implemented to generate foresight into brewery by-product generation, four supervised regression algorithms were implemented: Decision Tree Regressor, Random Forest Regressor, kNN Regressor, and Gradient

Boosting Regressor. These models were selected for their robustness in handling tabular datasets, ability to capture nonlinear relationships, and prior success in similar food and waste prediction tasks¹⁶⁻¹⁸.

3.3.1 Hyperparameter tuning:

Hyperparameter optimization was conducted using a grid search procedure combined with five-fold cross-validation to ensure fair comparison among models. For tree-based models (Decision Tree and Random Forest), parameters such as maximum depth, number of estimators, and minimum samples per split were adjusted. For kNN, the number of neighbors and distance metrics were varied. For Gradient Boosting, tuning focused on learning rate, maximum depth, and number of boosting iterations. The parameter sets yielding the lowest cross-validated error were selected for final model evaluation.

3.3.2 Baseline comparison:

To contextualize model performance, results were compared against two baselines: (i) a naive mean predictor, which estimates by-product quantities using historical averages; and (ii) a multivariate linear regression model, representing a conventional statistical approach. This benchmarking strategy allows assessment of whether advanced ML regressors

provide significant performance gains beyond traditional methods.

RESULTS:

4.1. Regression analysis:

Table 3 presents a comparative analysis of four ML algorithms for prediction tasks: Decision Tree, Random Forest, kNN and Gradient Boosting. The performance metrics of four widely used regressors were evaluated based on MAE, MSE, RMSE, and R^2 metrics. Specifically, MAE measures the average magnitude of errors between predicted and actual values, providing a straightforward interpretation of prediction accuracy. MSE is a measure that calculates the average squared differences between predicted and actual values, penalizing larger errors more heavily. RMSE provides an error metric in the same units as the target variable, making it easier to interpret. The final measure, R^2 , quantifies the proportion of variance in the dependent variable, indicating overall goodness of fit. These metrics are calculated for a regression model evaluated on a dataset with 500 instances, offering practical implications in the strengths and limitations of each classifier’s predictive capabilities.

Table 3: Comparative performance of ML models before hyperparameter tuning, processed by the authors

Model	MAE	MSE	RMSE	R2 Score
Decision Tree	228.84	75948.12	275.59	-1.11
Random Forest	177.90	44737.73	211.51	-0.25
Gradient Boosting	199.48	58094.60	241.03	-0.62
kNN	192.38	54588.41	233.64	-0.52

To facilitate interpretation of these results, this study defines acceptance and non-acceptance thresholds for MAE, MSE, and RMSE. Following prior ML performance assessments^{16,19}, lower metric values indicate superior predictive performance, whereas higher values signify weaker accuracy. Considering the mean target value for spent grains (≈ 449 kg) and the operational tolerance observed in brewery processes, models with $MAE \leq 200$ and $RMSE \leq 250$ (approximately 40 - 55 percent of the mean) are deemed acceptable for operational forecasting of by-product quantities. Values exceeding these thresholds are considered non-acceptable for reliable practical application.

These cut-offs were chosen to align with industry variability and reflect conservative operational standards. Based on these criteria, Random Forest (MAE = 177.90; RMSE = 211.51) and Gradient Boosting (MAE = 199.48; RMSE = 241.03) achieve acceptable prediction accuracy, while Decision Tree and kNN exceed the limits and therefore show limited suitability before hyperparameter tuning.

The paper revealed significant differences in the levels of predictive accuracy among four models before hyperparameter tuning. Specifically, Random Forest demonstrated the lowest MAE of 177.90, indicating that its predictions, on average, deviated less from actual values compared to the other models. Also, its MSE of 44737.73 and RMSE of 211.51 further underscored its precision, as lower MSE and RMSE values reflect a reduced level of prediction errors. However, its R^2 score of -0.25 suggested that while Random Forest handled absolute deviations well, it struggled to explain the variance in by-product generation effectively. kNN followed closely, achieving an MAE of 192.38 and an RMSE of 233.64, but its higher error metrics, accounting for -0.52, indicated a slightly less accurate performance before tuning. In comparison, Gradient Boosting and Decision Tree demonstrated significantly lower performance, characterized by higher MAE and RMSE values, as well as and markedly negative R^2 scores (-0.62 and -1.11, respectively). These results suggest a limited ability to capture the variability and underlying patterns in the data. As a result, the findings shown in this research are crucial as they emphasize the need for hyperparameter optimization to improve the models’ reliability and applicability within the context of brewery operations.

4.2. Model comparison after the tuning process:

From Table 4, the findings indicate shows that all models exhibited improved accuracy and robustness after optimization via the tuning method. The most notable model involved the Gradient Boosting Regression, which emerged as the most effective paradigm, with a reduced MAE of 159.22, MSE of 37034.11, and RMSE of 192.44, signifying its ability to make highly accurate predictions with minimal deviation from actual values. Its R^2 score of -0.03, although slightly negative, highlighted its ability to closely approximate the variance in by-product generation compared to other models. The results also highlight that for breweries, Gradient Boosting Regression is particularly suited for capturing complex, non-linear relationships within the dataset, which are critical for accurate forecasting in diverse operational scenarios.

Table 4: Table 4. Performance of ML models after hyperparameter tuning, processed by the authors

Model	MAE	MSE	RMSE	R2 Score
Decision Tree	184.18	49954.8	223.51	-0.39
Random Forest	164.72	40264.6	200.66	-0.12
Gradient Boosting	159.22	37034.1	192.44	-0.03
k-Nearest Neighbors (kNN)	179.91	49259.2	221.94	-0.37

Random Forest also showed substantial improvements, with an MAE of 164.72 and RMSE of 200.66, confirming its robustness and reliability as a predictive tool. However, its slightly higher error rates (-0.12) compared to Gradient Boosting suggest that while it remains a strong candidate, it is marginally less precise for this specific application. In contrast, kNN and Decision Trees, while benefiting from optimization, exhibited higher error rates, with MAE values of 179.91 and 184.18, respectively, and relatively poor R^2 scores (-0.37 and -0.39). Also, although kNN initially ranked second in by-product prediction accuracy before tuning, its performance was hindered by the dataset's complexity. Similarly, Decision Trees appeared prone to overfitting, further limiting their predictive accuracy. These results suggest that both algorithms may be less suitable for precise by-product prediction tasks without further adjustments or refinements. Based on the insights and evaluations, it can be concluded that Gradient Boosting Regression emerges as the most suitable model for predicting by-product quantities in breweries. This conclusion also underscores the model's superior performance compared to its counterparts, while also shedding light on the relative strengths and limitations of the other models.

4.3. Insight analytics and implications:

Figure 3 demonstrates a detailed visualization of waste generation, reuse rates and disposal patterns, generating decision-support evidence in brewery by-product management, thereby enhancing CE practices.

In particular, ML models demonstrate exceptional accuracy, with error rates for predicting spent grains, hops, and yeast at 0.42%, 0.41%, and 0.82% respectively, ensuring reliable forecasting of by-product quantities. Reuse rates were also impressively high, highlighting strong potential for resource recovery and environmental sustainability. Additionally, seasonal analysis reveals significant variations in by-product disposal rates, with Autumn exhibiting the

highest disposal rate at 27.74%, while Winter shows the lowest at 20.36%. These variations suggest that production cycles and demand fluctuations substantially influence waste generation and disposal. Such insights are crucial for breweries to address seasonal challenges by scaling reuse capacities or collaborating with stakeholders to find innovative uses for surplus by-products during high-disposal periods. Another notable insight pertains to monthly trends in by-product generation, which show peaks in May, October, August, and July, suggesting the importance of enhanced storage and reuse strategies during these times, while lower volumes in January and February offer opportunities for recalibrating production schedules. The dashboard further provides additional actionable insights tailored to specific beerhouses, enabling more targeted decision-making and operational improvements. Collectively, these insights emphasize the value of leveraging predictive modeling to optimize by-product management, address seasonal and monthly fluctuations, and align operations with CE principles for enhanced environmental and economic sustainability.

DISCUSSION:

The findings of this study underscore the effectiveness of advanced ML techniques in optimizing by-product prediction and supporting CE practices in breweries. Specifically, among the four ML models evaluated, Gradient Boosting Regression emerged as the most effective model (MAE = 159.22, RMSE = 192.44), confirming its suitability for capturing complex, non-linear patterns in brewery data. Although its slightly negative R^2 score of -0.03 suggests limited explanatory power, the model nonetheless provides robust predictions. Dashboard analytics also revealed reuse rates exceeding 95% for spent grains, hops, and yeast, illustrating how predictive tools can enhance CE-oriented resource management.

The findings also provide brewing manufacturers with several key insights. Initially, the predictive model demonstrated high accuracy in anticipating the quantity of brewing by-products. This aligns with previous studies that emphasize the importance of predictive analytics in anticipating resource flows to enhance decision-making in waste management²⁰. Moreover, key residues such as spent grains, hops, and yeast can be repurposed for animal feed, bioenergy, or food additives - aligning with findings from Osorio et al¹⁹, who noted the nutritional and functional value of such by-products. These reuse opportunities not only mitigate environmental impacts but also create additional economic value.

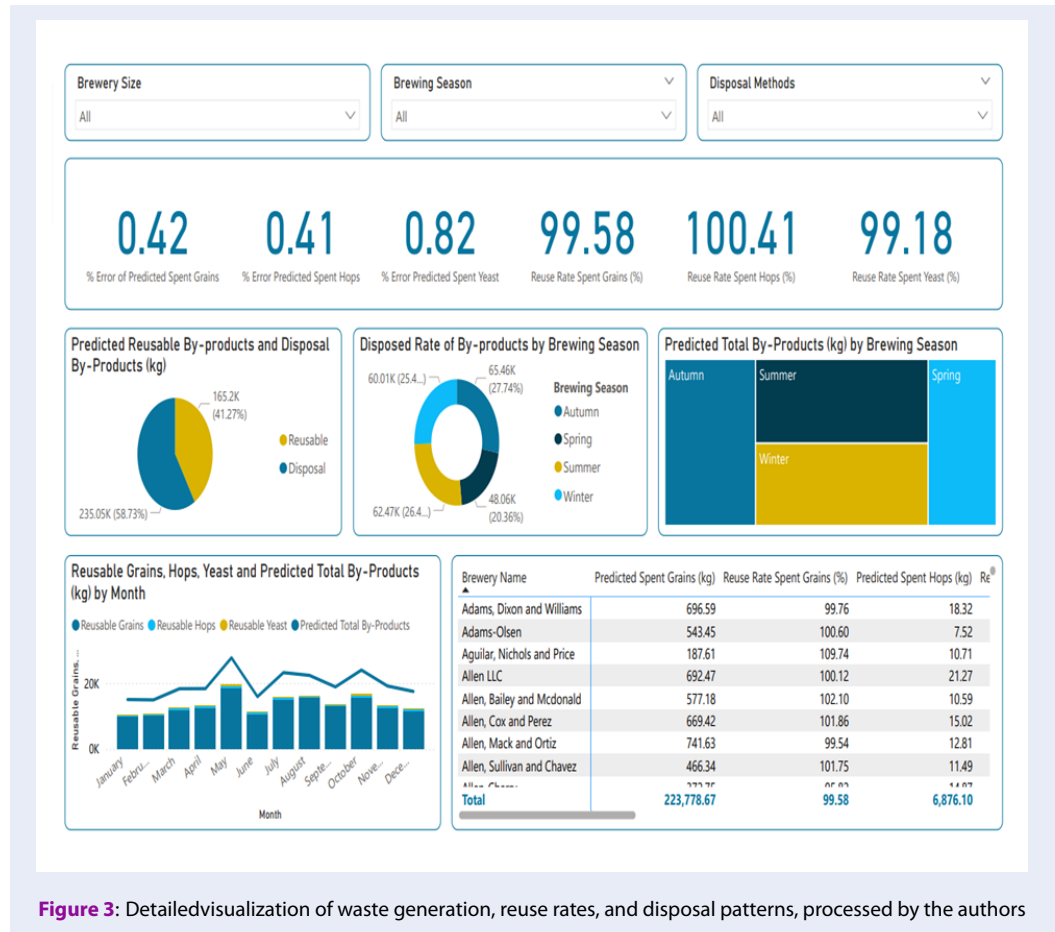


Figure 3: Detailed visualization of waste generation, reuse rates, and disposal patterns, processed by the authors

Moreover, from an operational perspective, integrating predictive models requires investments in staff training and data systems. Although these initiatives entail upfront costs, their long-term benefits justify adoption, as Dey et al²¹ highlight that implementing CE practices through improved resource efficiency can reduce disposal costs by up to 40%, confirming the economic viability of predictive analytics in sustainability transitions.

However, the study also identified challenges. The relatively small dataset may not fully represent the variability of large-scale brewing operations. Model performance might also be influenced by differences in by-product reporting across breweries, which could introduce bias. Additionally, the study did not explore in depth the specific reuse pathways for each by-product. Future work should expand the dataset and examine applications such as bioenergy and feed production to enhance the models' generalizability and practical value.

In conclusion, this study contributes to the literature only by demonstrating the potential of ML-based pre-

dition to improve sustainability in the brewing industry. The results validate the use of Gradient Boosting for forecasting by-product quantities and show how predictive insights can guide operational and environmental decision-making, reinforcing the goals of the CE.

CONCLUSIONS:

This research demonstrates the applicability of predictive analytics to strengthen CE practices in the brewing industry by forecasting by-product generation and reuse opportunities. Among the evaluated approaches, Gradient Boosting Regression delivered the highest predictive accuracy, confirming its methodological suitability for modeling non-linear and complex production dynamics. The integration of predictive models with dashboard analytics further illustrates how data-driven tools can transform brewery operations from reactive waste management to proactive resource planning.

Furthermore, the research makes two key contributions. Academically, it extends the literature on by-

product valorization by introducing ML - specifically Gradient Boosting - as a novel methodological perspective for managing food-industry residues. Practically, it provides breweries with a framework for anticipating by-product availability, aligning reuse strategies with seasonal fluctuations, and supporting operational efficiency and sustainability goals.

Nonetheless, the study is limited by the scope and size of the dataset, as well as its focus on a subset of by-products. Future research should validate the framework with larger and more diverse datasets, explore advanced modeling architectures, and extend predictive analytics to cross-industry applications of by-product reuse. Such efforts would enhance model robustness, offer broader market insights, and accelerate the transition toward data-driven CE solutions. In conclusion, the findings highlight the dual value of predictive modeling: advancing theoretical understanding of resource forecasting while delivering actionable tools for sustainable brewery management.

COMPETING INTERESTS

The authors declare that they have no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Hoang Khanh Van: conceptualized the research idea, collected and analyzed the data, developed the methodology, and drafted the full manuscript.

Lam Hong Thanh: served as the corresponding author, providing research guidance, methodological advice, and critical feedback on the manuscript structure and content.

ABBREVIATIONS:

CE: Circular Economy
 CO₂ : Carbon Dioxide
 °C: degree Celsius
 kNN: K-Nearest Neighbors
 ML: Machine learning
 pH: potential of hydrogen
 R²: R-Squared, Coefficient of Determination
 KTTH: Kinh tế Tuần hoàn
 MSE: Mean Squared Error
 RMSE: Root Mean Squared Error
 MAE: Mean Absolute Error
 BSG: brewer's spent grain
 IQR: interquartile range
 EDA: exploratory data analysis

A. APPENDIX

Appendix A. Figure 1. Research pipeline

A stepwise framework for data preparation, model training and evaluation to optimize brewery by-product predictions through predictive modeling.

Appendix B. Table 1: Statistics of key descriptive statistics for numerical variables

Summary of descriptive statistics for numerical variables, such as by-product volumes (spent grains, spent hops, and spent yeast) and production metrics (water volume, wastewater volume, production volume), including measures such as mean, median, standard deviation, and range, providing insights into data distribution and variability.

Appendix C. Table 2: Statistics of categorical variables

Summary of descriptive statistics for categorical variables, including frequency counts, offering insights into the distribution of categories within the dataset.

Appendix D. Figure 2: Correlation matrix of continuous variables

Correlation coefficients among continuous variables like spent grains, spent hops, spent yeast, and historical by-products. The color gradient reflects correlation strength, with red indicating strong positive relationships (1.0) and blue weak or negative correlations (-1.0). This heatmap shows the minimal interdependence between variables for by-product generation.

Appendix E. Table 3: Comparative performance of ML models before hyperparameter tuning

Comparative accuracy of ML models prior to hyperparameter tuning. Best performers are indicated across MAE, MSE, RMSE and R², with results underscoring both predictive potential and limitations in forecasting brewery by-products

Appendix F. Table 4: Performance of ML models after hyperparameter tuning

Accuracy and robustness of ML models after hyperparameter tuning, evaluated using MAE, MSE, RMSE and R². Gradient Boosting Regression outperformed other models in forecasting brewery by-products.

Appendix G. Figure 3: Detailed visualization of waste generation, reuse rates, and disposal patterns

The achieved prediction errors of <1% for spent grains, hops and yeast indicate the robustness of the applied models, reinforcing the importance of integrating such approaches into adaptive reuse strategies and circular resource management frameworks.

REFERENCES

1. Amoriello T, Ciccoritti R. Sustainability: Recovery and Reuse of Brewing-Derived By-Products. *Sustainability* (Basel). 2021;13(4):2355. Available from: <https://10.3390/su13042355>.
2. Pehlivi K, Elshani A, Loshi I, Zhara A. Water and Waste Assessment, Implications for Sustainable Production in Brewery Birra Peja. *Ecological Engineering & Environmental Technology*. 2023; Available from: <https://10.12912/27197050/165790>.
3. Belardi I, De Francesco G, Alfeo V, Bravi E, Sileoni V, Marconi O, et al. Advances in the valorization of brewing by-products. *Food Chem*. 2025;465(Pt 1). Available from: <https://10.1016/j.foodchem.2024.141882>.
4. Group WMFT. Watson-Marlow Fluid Technology Group. Better business in brewing: Global trends 2020-2025; 2019. Available from: <https://www.wmfts.com/globalassets/literature/tl-better-business-in-brewing-global-trends2020-2025.pdf>.
5. Barbu MC, Montecuccoli Z, Förg J, Barbeck U, Klímek P, Petutschnigg A, et al. Potential of Brewer's Spent Grain as a Potential Replacement of Wood in pMDI, UF or MUF Bonded Particleboard. *Polymers* (Basel). 2021;13(3):319. Available from: <https://10.3390/polym13030319>.
6. Soceanu A, Dobrin S, Popescu V, Buzatu A, Sirbu A. Sustainable Strategies for the Recovery and Valorization of Brewery By-Products - A Multidisciplinary Approach. *Sustainability* (Basel). 2024;16(1):220. Available from: <https://10.3390/su16010220>.
7. Danciu CA, Tulbure A, Stanciu MA, Antonie I, Capatana C, Zerbe MV, et al. Overview of the Sustainable Valorization of Using Waste and By-Products in Grain Processing. *Foods*. 2023;12(20):3770. Available from: <https://10.3390/foods12203770>.
8. Kao TH. Health Potential for Beer Brewing Byproducts. London: IntechOpen; 2018. Available from: <https://10.5772/intechopen.76126>.
9. Rachwałk, Waško A, Gustaw K, Polak-Berecka M. Utilization of brewery wastes in food industry. *PeerJ*. 2020;8. Available from: <https://10.7717/peerj.9427>.
10. Nettesheim P, Burggräf P, Steinberg F. Applications of machine learning in the brewing process: a systematic review. *Discov Artif Intell*. 2024;4(1):80. Available from: <https://10.1007/s44163-024-00177-6>.
11. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. *Comput Electron Agric*. 2018;147:70-90. Available from: <https://10.1016/j.compag.2018.02.016>.
12. Garcia-Garcia G, Woolley E, Rahimifard S, Colwill J, White R, Needham L. A Methodology for Sustainable Management of Food Waste. *Waste Biomass Valoriz*. 2017;8(6):2209-27. Available from: <https://10.1007/s12649-016-9720-0>.
13. H MK, Španinger E, IJ K, Ž K, U B. Hop compounds: extraction techniques, chemical analyses, antioxidative, antimicrobial, and anticarcinogenic effects. *Nutrients*. 2019;11(2):257. Available from: <https://10.3390/nu11020257>.
14. Rodrigues M, Migueis V, Freitas S, Machado T. Machine learning models for short-term demand forecasting in food catering services: a solution to reduce food waste. *J Clean Prod*. 2024;435. Available from: <https://10.1016/j.jclepro.2023.140265>.
15. Salehi R, Yuan Q, Chairapat S. Development of data-driven models to predict biogas production from spent mushroom compost. *Agriculture*. 2022;12(8):1090. Available from: <https://10.3390/agriculture12081090>.
16. Singh D, Tembhare M, Pundalik K, Dikshit AK, Kumar S. Machine learning based prediction of biogas generation from municipal solid waste: A data-driven approach. *Process Saf Environ Prot*. 2024;192:93-103. Available from: <https://10.1016/j.psep.2024.10.037>.
17. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? *Adv Neural Inf Process Syst*. 2022;35:507-20. Available from: <https://10.48550/arXiv.2207.08815>.
18. Yang Y, Zhao C, Lu H. Leveraging machine learning for food waste reduction: an analysis of predictive models. *Applied and Computational Engineering*. 2024;112(1):154-60. Available from: <https://10.54254/2755-2721/2025.18115>.
19. Osorio LL, Flórez-López E, Grande-Tovar CD. The Potential of Selected Agri-Food Loss and Waste to Contribute to a Circular Economy: Applications in the Food, Cosmetic and Pharmaceutical Industries. *Molecules*. 2021;26(2):515. Available from: <https://10.3390/molecules26020515>.
20. Awan U, Shamim S, Khan Z, Zia NU, Shariq SM, Khan MN. Big data analytics capability and decision-making: the role of data-driven insight on circular economy performance. *Technol Forecast Soc Change*. 2021;168. Available from: <https://10.1016/j.techfore.2021.120766>.
21. Dey PK, Malesios C, Chowdhury S, Saha K, Budhwar P, De D. Adoption of circular economy practices in small and medium-sized enterprises: evidence from Europe. *Int J Prod Econ*. 2022;248. Available from: <https://10.1016/j.ijpe.2022.108496>.

ỨNG DỤNG HỌC MÁY ĐỂ DỰ ĐOÁN PHỤ PHẨM SẢN XUẤT BIA TRONG BỐI CẢNH NỀN KINH TẾ TUẦN HOÀN

Hoàng Khánh Vân , Lâm Hồng Thanh *



Use your smartphone to scan this QR code and download this article

Trường Đại học Kinh tế - Luật, Khoa Hệ thống Thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh, TP. Hồ Chí Minh 700000, Việt Nam

Liên hệ

Lâm Hồng Thanh, Trường Đại học Kinh tế - Luật, Khoa Hệ thống Thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh, TP. Hồ Chí Minh 700000, Việt Nam

Email: thanhh@uel.edu.vn

Lịch sử

- Ngày nhận: 09/01/2025
- Ngày sửa đổi: 27/10/2025
- Ngày chấp nhận: 09/04/2026
- Ngày đăng: 05/05/2026

DOI : <https://doi.org/10.32508/vnuhcmjeb.v10i2.1549>



Bản quyền

© ĐHQG Tp.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



TÓM TẮT

Ngành công nghiệp sản xuất bia tạo ra một lượng lớn các loại phụ phẩm có giá trị tiềm năng như bã mạch nha, hoa bia và men bia, vốn chưa được tận dụng hiệu quả. Trong bối cảnh của nền Kinh tế Tuần hoàn (KTTH), trong đó yêu cầu các nhà sản xuất duy trì việc sử dụng tài nguyên trong thời gian dài, các nhà máy bia được khuyến khích chuyển đổi phụ phẩm thành nguồn tài nguyên có giá trị gia tăng, góp phần phát triển bền vững thông qua việc giảm thiểu tác động môi trường và cạn kiệt tài nguyên. Tuy nhiên, mặc dù các nhà máy bia đã nhận thức được tầm quan trọng của hoạt động quản lý chất thải và tái sử dụng phụ phẩm, hiện nay vẫn còn thiếu các nghiên cứu toàn diện về việc dự đoán định lượng phụ phẩm phát sinh. Khoảng trống này trở thành rào cản đối với mục tiêu tối ưu hóa sử dụng tài nguyên và giảm thiểu chất thải.

Để giải quyết thách thức trên, nghiên cứu này hướng đến mục tiêu tối ưu hóa quản lý chất thải và nâng cao khả năng tái sử dụng phụ phẩm ngành bia, góp phần đạt được các mục tiêu bền vững của KTTH. Cụ thể, các mô hình học máy được áp dụng nhằm dự đoán khối lượng phụ phẩm. Hiệu suất của các mô hình được đánh giá bằng các chỉ số hồi quy, trong đó Gradient Boosting cho kết quả chính xác nhất. Đồng thời, bảng điều khiển được sử dụng để trực quan hóa dữ liệu dự đoán và dữ liệu thực tế. Nghiên cứu sử dụng dữ liệu thực tế trong giai đoạn 5 năm, bao gồm nhiều biến sản xuất, đảm bảo tính thực tiễn và giá trị thực nghiệm. Nghiên cứu không chỉ đánh giá độ chính xác của mô hình mà còn phân tích cách tích hợp các kết quả dự đoán vào hoạt động vận hành hàng ngày của nhà máy thông qua bảng điều khiển phân tích, hỗ trợ ra quyết định dựa trên dữ liệu.

Các kết quả thu được giúp nhà máy bia dự báo khối lượng chất thải, điều chỉnh sản xuất, tối ưu hóa hoạt động và giảm thiểu lãng phí. Bên cạnh đó, hướng tiếp cận này cung cấp một khung ứng dụng có thể mở rộng cho các doanh nghiệp thực phẩm và đồ uống khác muốn áp dụng phân tích dự đoán cho mục tiêu phát triển bền vững. Nghiên cứu cũng nhấn mạnh tiềm năng của các mô hình dựa trên dữ liệu trong việc cải thiện quản lý tài nguyên của ngành bia, thúc đẩy quá trình phát triển bền vững và hiện thực hóa các mục tiêu của KTTH.

Từ khóa: Học máy, Phụ phẩm nhà máy bia, Kinh tế tuần hoàn, Hồi quy Gradient Boosting, Bền vững, Ngành sản xuất bia, Tái sử dụng phụ phẩm, Bã malt, Bã hoa bia, Bã men

Trích dẫn bài báo này: H K V, L H T. ỨNG DỤNG HỌC MÁY ĐỂ DỰ ĐOÁN PHỤ PHẨM SẢN XUẤT BIA TRONG BỐI CẢNH NỀN KINH TẾ TUẦN HOÀN. *VNUHCM J. Econ. Bus. Law.* 2026; 10(2):6498-6507.