proceedings

MDPI

Proceeding Paper

Application of Machine Learning Algorithms to Predict Composting Process Performance †

Vassilis Lyberatos [1,*] and Gerasimos Lyberatos [2,3]

[1] School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechneiou 9, Zografou, 15780 Athens, Greece
[2] School of Chemical Engineering, National Technical University of Athens, Iroon Polytechneiou 9, Zografou, 15780 Athens, Greece; lyberatos@chemeng.ntua.gr
[3] Institute of Chemical Engineering Sciences, Stadiou Str., Platani, 26504 Patras, Greece
* Correspondence: vaslyb@ails.ece.ntua.gr
† Presented at the 1st SUSTENS Meeting, 4–5 June 2025; Available online: https://www.sustenshub.com/welcome/.

Abstract

Four machine learning models (Decision Tree Regressor, Linear Regression, XGBoost Regression, K-Neighbors Regressor) were developed to predict the outcomes of a composting process based on key input parameters, including Ambient Temperature, mixture composition, and initial feedstock volume. The models were trained on data from 88 composting batches, monitoring temperature evolution, and compost yield. Performance evaluation demonstrated high accuracy in predicting compost maturity, process duration, and final product quantity. These predictive models could optimize composting operations by enabling real-time adjustments, improving efficiency, and enhancing resource management in sustainable waste processing.

Keywords: food waste; composting; machine learning

check for updates

Academic Editors: Nikolaos Diangelakis and Leonidas Matsakas

Published: 16 July 2025

Citation: Lyberatos, V.; Lyberatos, G. Application of Machine Learning Algorithms to Predict Composting Process Performance. *Proceedings* **2025**, *121*, 3. https://doi.org/10.3390/proceedings2025121003

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, the application of artificial intelligence (AI) has emerged as a promising approach to modeling composting processes. Despite this potential, there is a limited body of research focused on using machine learning (ML) to predict the stability and performance of composting systems. Most ML efforts in this domain have centered on process optimization, handling missing data, detecting anomalies, and managing complex variables.

ML is particularly adept at processing complex datasets, predicting nonlinear relationships, and addressing data gaps. These capabilities make it well-suited for overcoming the methodological challenges inherent in modeling processes like composting biowaste [1]. Several recent studies have demonstrated the effectiveness of ML techniques in composting applications, such as predicting CO_2 emissions [2], improving compost quality [3], monitoring moisture levels [4], and classifying compost maturity [5]. A critical review [6] highlighted both the advantages and limitations of various ML and AI algorithms applied to composting, underscoring their significance for optimizing this essential bioprocess.

In a recent work [7], alternative machine learning models were developed in order to describe a novel composting process, carried out in batches. The work presents an interdisciplinary framework designed to help policymakers, planners, and relevant stakeholders assess the potential of decentralized food waste composting systems, fostering the advancement of sustainable and effective waste management practices. The objective of

the present work is to evaluate and compare the performance of four ML models based on a larger number of composting batches.

2. Methodology

2.1. Experimental Setup

The dataset comprises 88 distinct composting processes derived from various batches, with features such as Biowaste Feed, Pruning Feed, Recycled Compost Feed, Sawdust Feed, Leaf Feed, Ambient Temperature, Mean Temperature, Max Temperature, Duration, Compost Crude, and Compost Net. For the modeling process, we selected Biowaste Feed, Pruning Feed, Recycled Compost Feed, Sawdust Feed, Leaf Feed, and Ambient Temperature as input features, while the remaining variables were treated as outputs. Figure 1 presents key statistics and feature correlations. The dataset was divided into training and testing subsets, with 80 batches used for training and 8 batches for testing. Model performance was assessed using evaluation metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) to ensure a comprehensive performance evaluation. Additionally, we employed permutation-based feature importance to identify the key features driving the model's predictions. All experiments were conducted using Python 3.10 libraries, including Matplotlib 3.5.0 and scikit-learn 1.0.1.

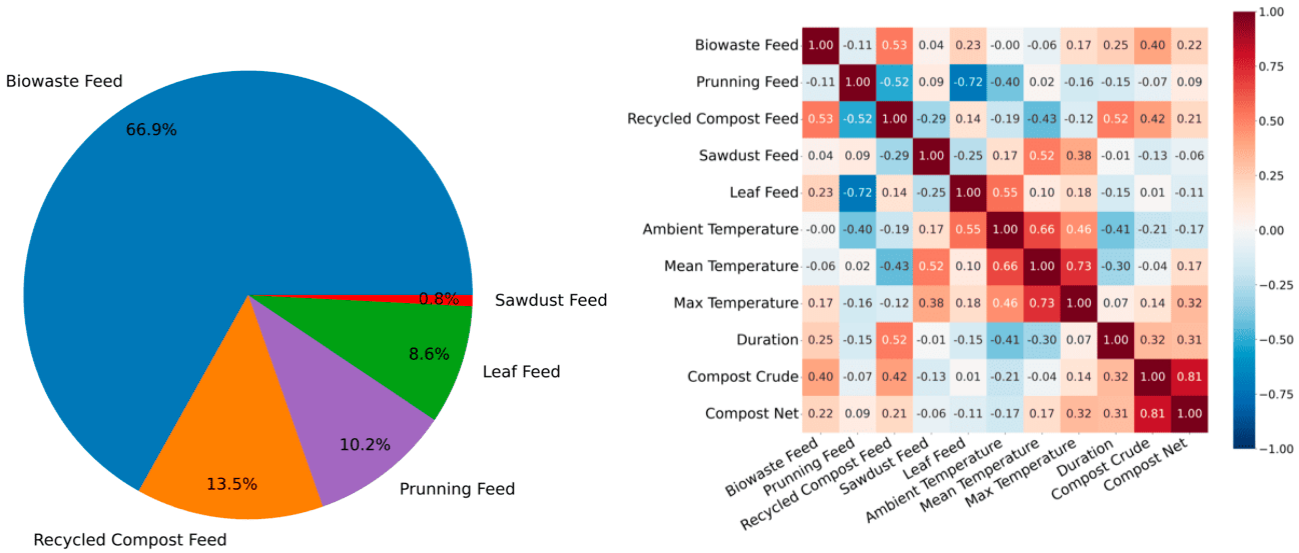


Figure 1. Average percentage of feed types (**left**). Correlation matrix between the features (**right**).

2.2. Machine Learning Modeling for Compost

Four regression models were implemented to predict composting process outcomes based on feed composition and Ambient Temperature. The first model, Decision Tree Regressor, is a nonlinear model that splits data into branches, making it highly interpretable and capable of capturing complex relationships between features [8]. The second model, XGBoost Regression, is an ensemble technique that utilizes gradient boosting for enhanced performance and scalability, often outperforming other models on large datasets [9]. The third model, Linear Regression, offers a simple yet effective approach by modeling the relationship between dependent and independent variables as a straight line, making it easy to interpret and widely used for its efficiency [10]. Finally, the K-Neighbors Regressor is a non-parametric method that predicts the target variable based on the average of the nearest neighbors [11].

3. Results and Discussion

Based on the results presented in Table 1, the K-Neighbors Regressor outperforms the other models, achieving the lowest error on the test set. The Linear Regressor also performs well, yielding good results. In contrast, the Decision Tree Regressor exhibits the poorest performance, with the highest error according to our metrics. The K-Neighbors Regressor's strong performance is further highlighted in Figure 2 (right), where the predicted values closely align with the actual values. Additionally, Figure 2 (left) shows that the Sawdust Feed has the greatest impact on our model's performance, while the Leaf Feed plays the least significant role.

Table 1. Metrics for each of the four developed models.

	MSE		MAE	
	Test Set	**5-Fold**	**Test Set**	**5-Fold**
Decision Tree Regressor	683.386	705.108 ± 153.128	15.485	15.641 ± 1.281
XGBoost Regression	304.487	512.636 ± 183.908	9.733	12.643 ± 1.727
Linear Regression	282.981	413.277 ± 112.680	11.218	12.263 ± 1.278
K-Neighbors Regressor	120.841	418.641 ± 94.392	7.943	12.365 ± 1.321

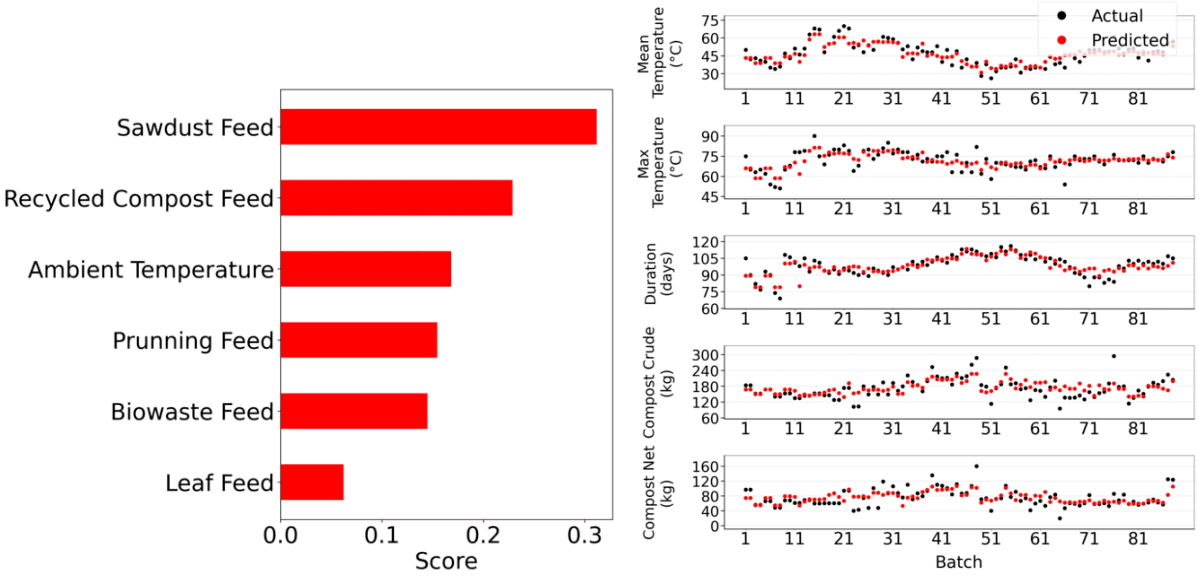


Figure 2. Permutation feature importance analysis for the K-Neighbors model (**left**) and the comparison of actual and predicted values for all outputs across 88 composting batches (**right**).

These performance differences can be attributed to the dataset's characteristics. The K-Neighbors Regressor likely excels due to the small but structured nature of the data, where similar input parameters correlate with similar output values. Linear Regression's effectiveness suggests underlying linear relationships, while Decision Trees' poor performance likely stems from overfitting on the limited training data. The modest sample size appears to favor less complex algorithms like K-Neighbors over sophisticated models such as XGBoost, which generally require larger datasets for optimal performance.

Future research directions should focus on collecting more diverse data across different operational conditions and seasons, and experiment with modeling techniques that could exploit the sequential nature of the composting process, such as recurrent neural networks or time series analysis methods, to enhance the models' predictive capabilities and validate their performance across a broader range of applications.

Author Contributions: V.L.: formal analysis, data curation, visualization, investigation, and writing. G.L.: supervision, conceptualization, and review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was carried out with the financial assistance of the European Union under the ENI CBC Mediterranean Sea Basin Programme, SIRCLES "Supporting Circular Economy Opportunities for Employment and Social Inclusion" (Project Number: B_A.3.1_0157_SIRCLES).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available in the repository https://github.com/vaslyb/compost, accessed on 1 July 2025.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Manley, K.; Nyelele, C.; Egoh, B.N. A review of machine learning and big data applications in addressing ecosystem service research gaps. *Ecosyst. Serv.* **2022**, *57*, 101478. [CrossRef]
2. Li, Y.; Li, S.; Sun, X.; Hao, D. Prediction of carbon dioxide pro- duction from green waste composting and identification of critical factors using machine learning algorithms. *Bioresour. Technol.* **2022**, *360*, 127587. [CrossRef]
3. Yılmaz, E.C.; Aydın Temel, F.; Cagcag Yolcu, O.; Turan, N.G. Modeling and optimization of process parameters in co-compost- ing of tea waste and food waste: Radial basis function neural net- works and genetic algorithm. *Bioresour. Technol.* **2022**, *363*, 127910. [CrossRef] [PubMed]
4. Moncks, P.C.; Corrêa, É.K.; Guidoni, L.L.; Moncks, R.B.; Cor- rêa, L.B.; Lucia, T., Jr.; Araujo, R.M.; Yamin, A.C.; Marques, F.S. Moisture content monitoring in industrial-scale composting systems using low-cost sensor-based machine learning techniques. *Bioresour. Technol.* **2022**, *359*, 127456. [CrossRef] [PubMed]
5. Kujawa, S.; Mazurkiewicz, J.; Czekała, W. Using convolutional neural networks to classify the maturity of compost based on sewage sludge and rapeseed straw. *J. Clean. Prod.* **2020**, *258*, 120814. [CrossRef]
6. Temel, F.A.; Yolcu, O.C.; Turan, N.G. Artificial intelligence and machine learning approaches in composting process: A review. *Bioresour. Technol.* **2023**, *370*, 128539. [CrossRef] [PubMed]
7. Lytras, C.; Lyberatos, V.; Lytras, G.; Papadopoulou, K.; Vlysidis, A.; Lyberatos, G. Development of a Model Composting Process for Food Waste in an Island Community and Use of Machine Learning Models to Predict its Performance. *Waste Biomass Valor.* **2024**, *16*, 683–700. [CrossRef]
8. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall: New York, NY, USA, 1986. [CrossRef]
9. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD' 16), San Francisco, CA, USA, 13–14 August 2016; pp. 785–794. [CrossRef]
10. Seber, G.A.F.; Lee, A.J. *Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
11. Cover, T.M.; Hart, P.E. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.