Human Interaction and Emerging Technologies (IHIET-AI), Vol. 201, 2026, 238–250
https://doi.org/10.54941/ahfe1007177

Taste Matters: Machine Learning Models for Context-Aware Recipe Prediction

Michael Müller[1,2], David Kraus[2], Moritz Zink[2], and Eric Sax[2]

[1]WMF Business Unit Consumer GmbH, WMF Platz 1, Geislingen, 73312, Germany
[2]Karlsruhe Institute of Technology (KIT), ITIV, Karlsruhe, 76131, Germany

ABSTRACT

Taste has always been a decisive factor in food and beverage preparation. Yet, in times of increasing ecological awareness, optimizing recipes requires balancing subjective user satisfaction with measurable sustainability goals. Coffee, one of the most widely consumed beverages, provides a particularly relevant case: small changes in the Coffee-to-Water Ratio (C2WR) not only influence taste perception but also have a measurable impact on the environmental footprint. Building on previous work that established a universal architecture for context-aware food and beverage preparation systems (CONFES) and developed a large-scale data acquisition framework for a context-aware coffee machine, this paper extends the research toward machine learning modelling approaches capable for prediction of recipe parameters like C2WR. Tree-based ensemble models, such as Random Forest, Gradient Boosting and AdaBoost explained a higher proportion of variance (R^2 = 61.5%) compared to Neural Networks, k-Nearest Neighbour, and Support Vector Machines.

Keywords: Context-aware systems, Coffee-to-water ratio (C2WR), Recipe optimization, Machine learning, Taste perception, Sustainability, Random forest, Gradient boosting, AdaBoost

INTRODUCTION

The environmental impact of food consumption and preparation is dependent on both the type of food and the quantity of ingredients used (Notarnicola et al., 2017) (Nemecek et al., 2016). The substitution of food items with a high environmental impact with food items with a lower environmental impact, or the reduction of the quantity of ingredients with a high environmental impact, can contribute to the sustainability of food consumption (Clark et al., 2019). A single cup of coffee, prepared using 8 grams of coffee powder and 120 millilitres of water, requires an average of 140 litres (L) of water (Chapagain & Hoekstra, 2007) and generates around 0.045 kilograms of CO2 equivalent (kg CO^2e) (Reinhardt & Wagner, 2020) (across the entire production chain). Approximately two billion cups of coffee are consumed on a daily basis (Surma & Oparil, 2021). Even minor reductions in the quantity of coffee powder can result in a notable reduction in the environmental impact of coffee consumption. However, any reduction must be within the tolerance range of the consumer, ensuring that the taste is not adversely affected. In our previous work, we proposed an architectural framework for a **CON**text-aware **F**ood and b**E**verage preparation **S**ystem (CONFES) with

Received September 4, 2025; Revised November 28, 2025; Accepted December 9, 2025; Available online April 1, 2026

the objective of optimising the environmental footprint of prepared food and beverages, while simultaneously pursuing further optimization goals in the domains of nutrition and taste (Mueller et al., 2024). Research in automotive domain pursues similar objectives in enhancing user experience and reduce costs through provision of large-scale datasets (Pistorius et al., 2020). Also in terms of changing comfort functions in vehicles the context of a given situation can have a large impact on the preferences with track information and for example, weather data a routine can be detected in the preferred vehicle comfort functions (Guinea et al., 2021). Testing self-learning systems is essential in safety-critical settings like vehicles, making scenario-based validation of systems that interact with individual user behaviour necessary (Stang, 2025). However modern vehicle architectures are designed for safety critical environments, where incorrect predictions may result in severe consequences. Such stringent safety requirements are not applicable to coffee machines, allowing the use of standard architectures (Zink et al., 2025). To date, one database exists that stores recipes with adapted amounts of ingredients, user feedback, and collected context (Mueller et al., 2025). The objective is to evaluate different machine learning models against each other, to find models that are capable to predict the recipe parameters, such as the C2WR. Identifying appropriate models lies the foundation for context-dependent optimization of the C2WR, thereby achieving the desired taste and promoting the sustainable use of ingredients. In accordance with the No Free Lunch theorem, no single learning algorithm demonstrates universal superiority across all problem domains, necessitating empirical evaluation of multiple candidate models for each specific task (Guillén & Rojas, 2016).

Problem

Although datasets are now being collected that link recipes, contextual information, and user feedback, at present only a relatively small dataset exists that systematically integrates these dimensions (Mueller et al., 2024). Current systems often rely on static presets or heuristic adjustments, which overlook the demanding and dynamic interplay between human factors (e.g., fatigue, stress, mood), environmental conditions (e.g., temperature, humidity), and recipe variables. Moreover, while numerous statistical and machine learning (ML) models are available, their relative strengths and limitations in detecting and interpreting such influential drivers for coffee preparation remain unexplored. Without a structured comparison, the development of adaptive, context-aware systems remains fragmented and lacks methodological grounding.

Aim

The aim of this study is to investigate how ML models can be employed to predict via regression the C2WR for the target of recipe optimization. Rather than emphasizing model quality alone, the study highlights the comparative performance of different model classes. The analysis considers how well

each approach can handle subjectivity in human taste feedback, variability in context, and heterogeneity across participants. By doing so, our work seeks to establish methodological guidance for researchers and practitioners: which models are best suited for interpretability, which are more effective at capturing nonlinear effects, and how they may complement each other in designing adaptive systems for food and beverage preparation.

Contribution

This paper contributes a structured overview of ML approaches applied to the context of recipe optimization through prediction of the C2WR through regression. Machine learning methods, such as tree ensemble techniques, excel at capturing interactions and nonlinearities. By comparing their performance on a shared dataset, the paper delivers a balanced perspective on their relative merits. Importantly, the contribution does not lie in identifying significant contextual drivers, but in the assessment and evaluation of different ML models for the suitability of the task of C2WR prediction. The comparative analysis offers an informed view on which approaches are more promising for real-world deployment in context-aware food preparation systems. This work lays the foundation for future systems that align taste preferences with sustainability goals.

RELATED WORK

Context-Aware Food and Beverage Preparation Systems

The analysed data set, was gathered with a data acquisition (DAQ) system which is an implementation of the universal architecture for a CONFES described in (Mueller et al., 2024). The architecture is based on the state of the art of Cyber-Physical-System development and considers machine learning algorithms used in other food preparation systems. The CONFES comprises four subcomponents: a sensor system, a computing unit, a database, and a food and beverage preparation machine (See Figure 1). The sensor system detects pre-selected context information (e.g., ambient temperature, humidity, and time of day). The computing unit processes the data, while also receiving data from the user and the database in order to create a personalised recipe, which is then sent to the food and beverage preparation machine. The preparation machine serves the food or beverage to the user. Finally, the computing unit interacts further with the user to collect feedback and update the database accordingly.

ML Regression Models in Taste Prediction

Recent studies demonstrate that tree-based ensemble models and deep learning approaches achieve the best model quality for taste and flavour estimation. Among these, Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost)

consistently outperform conventional regression methods, particularly when modelling, non-linear relationships inherent in sensory and chemical data (Yavas et al., 2024) (Schreurs et al., 2024) (Song et al., 2023) (Wang et al., 2021) (Fernie & Alseekh, 2022) (Zeng et al., 2023) (Goel et al., 2022). In parallel, deep learning and graph-based models have emerged as promising tools for molecular-level taste prediction. Graph Neural Networks (GNNs), particularly when combined with molecular fingerprints in hybrid or consensus architectures, exhibit strong predictive power for both classification and regression of taste attributes (Song et al., 2023). Neural networks are suitable of modelling nonlinear behaviour as their activation functions are built to model this relationship (Malavolta et al., 2022) (Miao et al., 2023). The ability of tree based ensemble models to model interactions among numerous input variables makes them highly effective for predicting multi-factorial taste responses. For example, RF and GB models have achieved Pearson correlation coefficients of up to r = 0.94 in sweetness prediction tasks, significantly surpassing traditional models (Song et al., 2023). Similarly, in applications such as wine and beer flavour prediction, ensemble methods outperformed both linear and kernel-based algorithms, confirming their suitability in sensory modelling (Fernie & Alseekh, 2022) (Wang et al., 2021) (Schreurs et al., 2024).

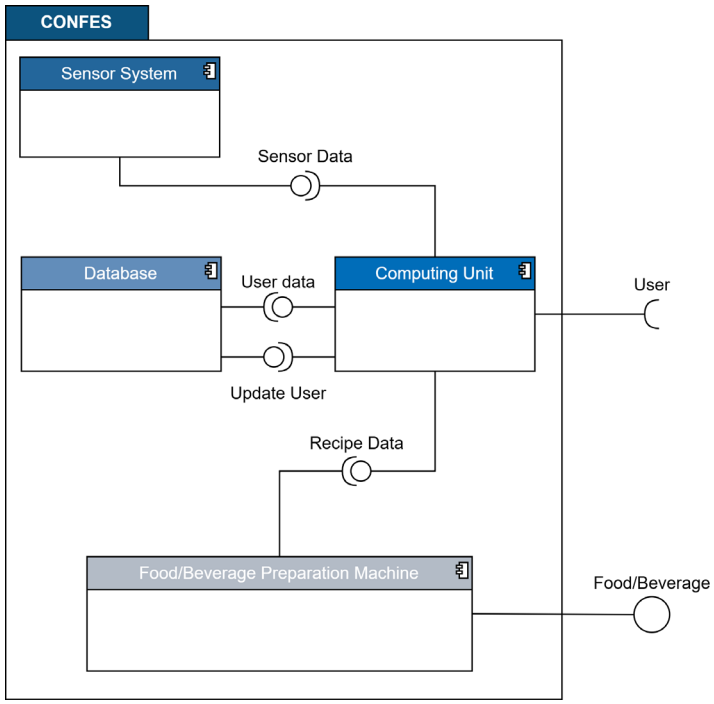

Figure 1: Component diagram of general CONFES setup (Mueller et al., 2024).

PREPARATION OF DATASET

The experiment that generated the dataset presented in the related work was continued to collect additional data. At the time of publication, 89 participants had consumed a total of 2,142 cups of coffee. For the classification of user feedback, two variables were defined: **isGood** (**iG**) and **isSustGood** (**iSG**). Their construction was based on two feedback parameters: (1) the general rating of the beverage, ranging from "1" (bad), "2" (ok), "3" (good), to "4" (very good); and (2) the intensity rating, coded as "0" (too strong), "1" (good), and "2" (too weak). Beverages with a general rating of at least "2" were assigned to **iG**, as they were considered acceptable and not rated negatively by participants. For **iSG**, beverages rated as "too strong" were excluded, since a higher coffee powder dosage than necessary implies a lower level of sustainability. The resulting datasets comprise 1,758 samples for **iG** and 1,488 samples for **iSG**. In total, the dataset includes 83 features encompassing user attributes, personal context, environmental context, feedback variables, and recipe parameters, comprising both directly measured and feature-engineered data.

MODEL TRAINING AND CONFIGURATION

To evaluate the predictive performance of different modelling approaches, several algorithms were trained and compared using distinct parameter configurations. For model training, user feedback on dispensed drinks, participant identifiers, and both personal and environmental context parameters were used as input features, whereas the C2WR served as the output variable. The selection of models covers both statistical and machine learning methods, allowing a broad assessment of linear and non-linear relationships within the dataset. Prior to model training, a cross-validation procedure was conducted (Krstajic et al., 2014) (Bergmeir & Benítez, 2012). The dataset was partitioned into 10 subsets, where nine subsets were used for training and one subset for testing in each iteration. This 10-fold cross-validation process was repeated with deterministic sampling to guarantee identical data splits across all model runs (Wong & Yeh, 2020). The approach ensures comparability between models and minimizes the influence of random sampling on performance metrics (Lei, 2017).

Decision Tree. A binary decision tree classifier was trained with a minimum of two instances per leaf node and a minimum subset size of five for splitting. The maximum tree depth was limited to 100 levels to avoid overfitting. The classification process was terminated once the majority class purity in a node reached 95%.

Neural Network. A multi-layer perceptron algorithm with backpropagation with one hidden layer was implemented. For the hidden layer, the rectified linear unit (ReLU) activation function was applied, and weight optimization was performed using the limited-memory quasi-Newton method for bound-constrained problems (L-BFGS-B). Ridge regression (L2) penalty (regularization term) was applied with a coefficient of $\alpha = 0.0001$ to prevent overfitting, and the maximal number of epochs was set to 200. All trainings were conducted under replicable conditions.

Linear Regression Models. Three regularized linear regression variants were trained for comparison:

- Lasso Regression (L1): Regularization strength $\alpha = 0.0001$.
- Ridge Regression (L2): Regularization strength $\alpha = 0.0001$.
- Elastic Net Regression: Regularization strength $\alpha = 0.0001$ with an equal mixing ratio of L1:L2 = 0.5:0.5.

All models included an intercept term and aimed to balance model interpretability with predictive performance.

k-Nearest Neighbours (kNN). The kNN classifier used k = 10 neighbours and uniform weighting. Distance was measured using three metrics—Euclidean, Manhattan, and Chebyshev—to evaluate metric sensitivity.

Support Vector Machine (SVM). Several kernel functions were evaluated to explore both linear and non-linear decision boundaries. The model used a cost parameter of $C = 1.0$ and, for regression variants, an epsilon-insensitive loss with $\varepsilon = 0.1$.

- Linear kernel: $x \cdot y$
- Radial Basis Function (RBF) kernel: $\exp(-\gamma |x-y|^2)$, with γ set to "auto"
- Sigmoid kernel: $\tanh(\gamma\, x \cdot y + c)$, with γ = "auto" and $c = 1.0$
- Polynomial kernel: $(\gamma\, x \cdot y + c)^d$, with γ = "auto," $c = 1.0$, and degree = 3.

Optimization was performed with a numerical tolerance of 0.001 and an iteration limit of 100.

Extreme Gradient Boosting. The XGBoost model was configured with 100 trees, a learning rate of 0.3, and a regularization parameter $\lambda = 1$. The maximum depth of individual trees was limited to six. Full subsampling was used, with all training instances and features considered at each tree, level, and split.

Gradient Boosting. A GB model was trained with 100 estimators and a learning rate of 0.1. Each tree was restricted to a maximum depth of three, and subsets smaller than two instances were not split further. Subsampling was set to 1.0 for all training instances.

Random Forest. The RF classifier consisted of 10 trees, each constrained to avoid splitting subsets smaller than five instances, ensuring model robustness while limiting overfitting.

Adaptive Boosting. AdaBoost was configured with a decision tree as base estimator, 50 estimators, a learning rate of 1.0., and regression variants were evaluated with linear (lin), square (squ), and exponential (exp) loss functions.

Model Validation Procedure. The combination of models was chosen to represent a comprehensive spectrum of algorithmics. Linear models (Lasso, Ridge, Elastic Net) provide a transparent baseline to quantify linear dependencies between contextual factors and recipe parameters (Allen & Tkatchenko, 2022) (Rousseeuw, 2024). Tree-based ensemble methods (RF, GB, XGBoost, AdaBoost) enable the capture of higher-order, non-linear relationships and feature interactions (Huynh-Thu et al., 2010) (Lundberg

et al., 2020). The SVM introduces kernel-based non-linear modelling capabilities (Chatterjee & Yu, 2016), while the Neural Network represents a data-driven approach where conventional feature engineering is obsolete due to its capability of learning feature hierarchies (Alsallakh et al., 2017). Together, these methods ensure a balanced comparison of interpretable, statistically grounded, and advanced machine learning approaches for predicting and optimizing context-dependent beverage preparation.

Global and Personal Modelling Approach. To investigate whether inter-individual differences should be considered in future implementations of machine learning models for recipe prediction, all models were trained using both a global (G) and a personal (P) approach for C2WR prediction. Accordingly, the training process was conducted in two configurations: (1) **a global model**, trained without any participant-specific information, where the participant identifier (ID) was removed from the dataset, and (2) **a personalized model**, trained with participant knowledge included in the form of the participant ID. This distinction enables an assessment of how individual user characteristics influence model performance and the potential benefits of personalization for context-aware recipe optimization. For future implementations, the global model can be applied immediately, even without prior knowledge of the user, making it suitable for new coffee-machine users. The personalized model, in turn, becomes advantageous once individual user data is available.

EVALUATION OF ML REGRESSION MODELS FOR PREDICTION OF C2WR

To compare the predictive performance of all trained models, three evaluation metrics were applied: the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the coefficient of determination (R^2). While MAE and RMSE are useful primarily for relative comparisons between models, the R^2 value directly reflects model quality by indicating how much of the variance of the target variable is explained. To identify the models with the highest model quality and comparatively low error, we applied the following filtering criteria, reducing the set from 68 to 10 models: MAE > 0.018, RMSE > 0.024, or R^2 < 0.529. After applying these filter criteria, only tree-based ensemble models remained among the top ten performing configurations. Across all evaluated models, the observed performance metrics ranged between MAE = 0.016 – 0.050, RMSE = 0.021 – 0.056, and R^2 = –1.656 – 0.615. Within this range, the **RF PiS model, GB PiS model**, and the **GB GiS model** achieved the highest overall performance (see Table 1). These results indicate that ensemble-based learning methods, particularly those leveraging boosting strategies, are well-suited for capturing the non-linear dependencies within the dataset. The consistent superiority of the PiS models further suggests that incorporating participant-specific knowledge enhances model quality for context-aware recipe optimization. This highlights the performance gap between tree-based ensemble models and other algorithmic categories, reinforcing their suitability for the CONFES framework.

Table 1: Comparison of model performance across all configurations using MAE, RMSE, and R^2 metrics for the top ten models, sorted by R^2.

Modell	RMSE	MAE	R^2
RF PiS	**0.021**	**0.016**	**0.615**
GB PiS	**0.022**	**0.017**	**0.593**
GB GiS	**0.022**	**0.017**	**0.588**
AdaBoost Lin PiS	0.023	0.016	0.573
AdaBoost Squ PiS	0.023	0.016	0.573
AdaBoost Exp PiS	0.023	0.016	0.559
XGBoost PiS	0.024	0.017	0.535
XGBoost PiG	0.024	0.018	0.531
AdaBoost Lin GiS	0.024	0.017	0.530
XGBoost GiS	0.024	0.017	0.529

Detached from the previous evaluation of model quality, the prediction behaviour of the models is analysed by parity plots, which display the relationship between predicted and the actual C2WR values (Figure 2, Figure 3, and Figure 4). The three models with the highest R^2, also show a similar, over the whole range of C2WR evenly prediction behaviour. All three models exhibit a strong linear relationship with a regression line coefficients between **r** = 0.76 and 0.77, indicating a high degree of agreement between predicted and observed values, and the suitability of the three models across the full range of measured C2WR values. The confidence ellipses are narrow and long all plots, further confirming that all three models capture the underlying data variability with comparable precision.

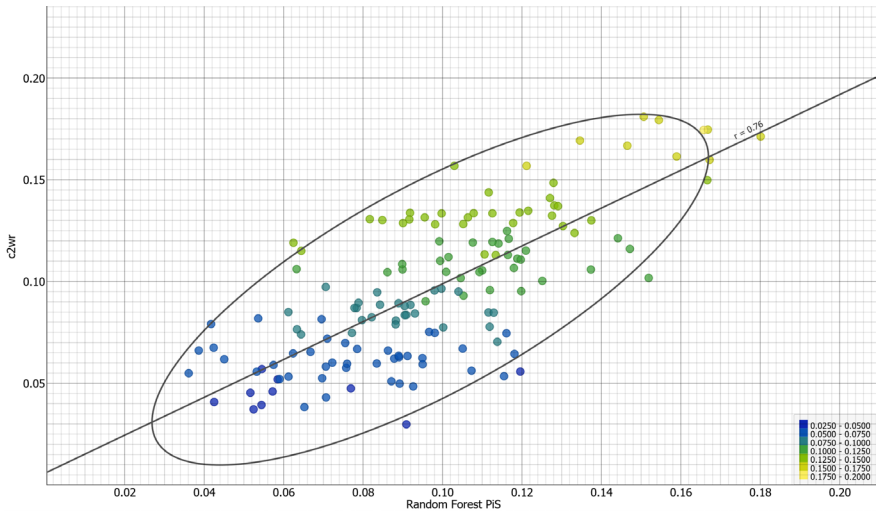

Figure 2: Parity plot of the RF PiS model showing predicted vs. actual C2WR values.

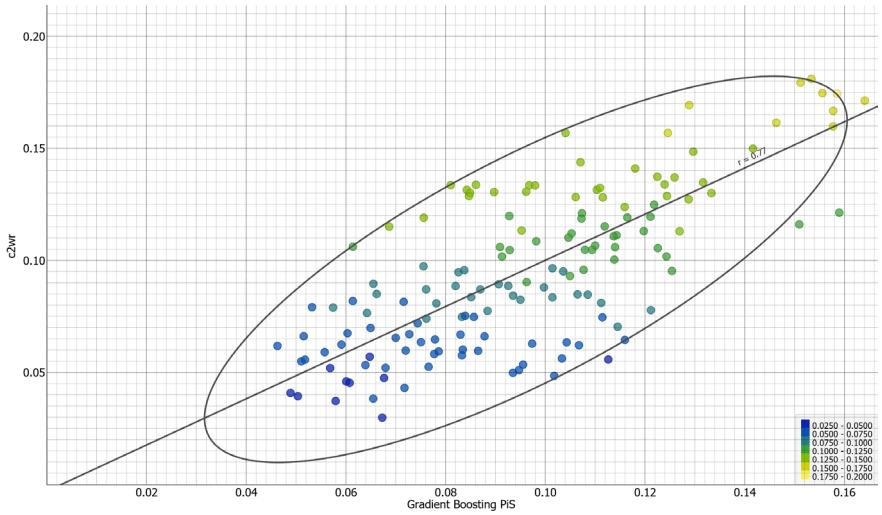

Figure 3: Parity plot of the GB PiS model showing predicted vs. actual C2WR values.

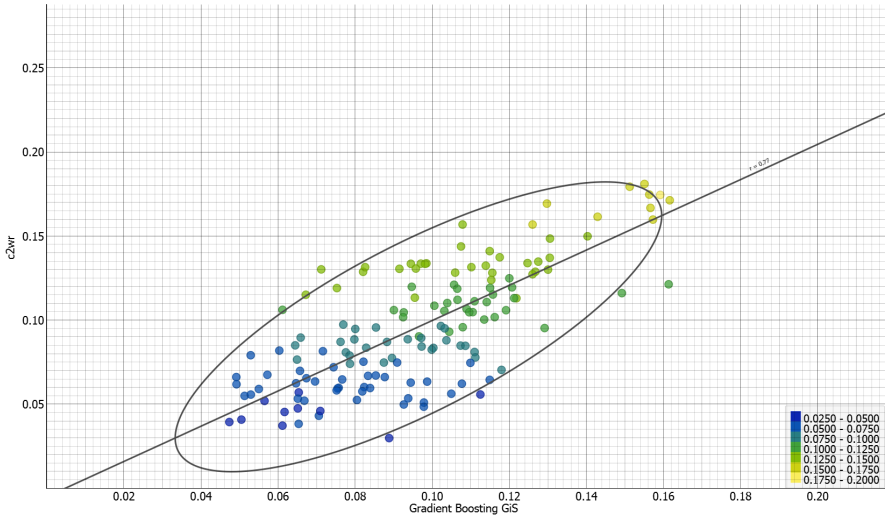

Figure 4: Parity plot of the GB GiS model showing predicted vs. actual C2WR values.

DISCUSSION

The results demonstrate that tree-based ensemble models consistently outperformed all other approaches. This superior performance can be attributed to two main factors: the size and structure of the dataset, and the inherent variability introduced by real-world data collection. Since the dataset was obtained under real world conditions, it contains experimental noise, participant-specific behavioural differences, and contextual fluctuations. Ensemble methods, particularly boosting algorithms, are known for their suitability to such noise and their ability to capture non-linear dependencies, which likely explains their dominance in this study. The models proved suitable for predicting the C2WR based on contextual and feedback

information. Considering the predictive capability of the models and the individual tolerance range of good-rated C2WR values, the system can select a C2WR at the lower bound of an individual's acceptable range, yielding at least a 10% reduction in coffee powder usage. Using the initial example (140 L of water and 0.045 kg CO_2e), this corresponds to a minimum saving of 14 L of water and 0.0045 kg CO_2e per cup of coffee. Scaled to the global consumption of approximately two billion cups of coffee per day, this reduction translates to roughly 9×10^6 kg CO_2e and 2.8×10^{10} L of water saved per day. The achieved performance metrics are consistent with values reported in related literature, indicating that the chosen modelling approach is appropriate. Expanding the dataset would provide opportunities for a more detailed assessment of the models and allow the integration of additional explanatory variables. As the data were collected through an embedded experimental setup in the coffee machine, further observations could contribute to a broader representation of user behaviour and taste perception. A larger sample would also enhance the stability of parameter estimates and may enable the identification of additional context-related patterns. Although the most relevant features were selected through feature engineering, it cannot be ruled out that other contextual or physiological factors influenced the results. Overall, these findings provide a solid empirical basis upon which further research can build. The study highlights substantial variability in individual coffee preferences, reinforcing the importance of personalisation in beverage preparation systems. Such variability, observable in the participant-specific intercepts, suggests that even relatively simple adjustments to brewing parameters could lead to noticeable improvements in user satisfaction. Even without deploying the models or a full CONFES framework, a coffee machine can achieve a more personalized C2WR through a simple preference-adaptation algorithm, similar to the one used in the experiment (Mueller et al., 2025). This approach enables users to iteratively calibrate their taste profile by providing continuous feedback. The results also indicate that integrating contextual and personal information enables a more precise adaptation of recipes. This finding is supported by the superior performance of the PiS models, which demonstrated that personalisation enhances both taste and resource efficiency. Future research should therefore focus on understanding how implicit and explicit information about user preferences can be captured and used to dynamically adjust brewing parameters. In the long term, these insights could guide the design of coffee machines that optimize both flavour and sustainability, as previously proposed in our earlier work (Mueller et al., 2024). Such systems could reduce coffee powder consumption while maintaining user satisfaction and improving sustainability, particularly interesting in business-to-business applications where resource optimization is critical.

CONCLUSION AND OUTLOOK

The objective of this study was to demonstrate how machine learning models can be employed on an existing dataset to predict the C2WR, and thereby establish a foundation for a CONFES. A possible implementation, of such

a system, within the machine operates as follows: upon user authentication and beverage selection, the model gets the following input parameters, user identity, personal and environmental context, and historical feedback, and outputs a predicted C2WR that is chosen to be as low as possible, yet as high as necessary, to ensure both sustainability and acceptable taste.

It is important to acknowledge that the current study and its findings are subject to several limitations that must be addressed before implementation. To date, data have been collected from 89 consumers, comprising a total of 2,142 cups of coffee. The participant group does not include individuals younger than 20 or older than 65 years, as the experimental setup was deployed in professional environments. Moreover, the majority of participants were male (57 %) and non-smokers (90 %), and more than 90 % of all participants were located in Europe. Consequently, the present dataset is not yet sufficient to systematically investigate the influence of personal attributes and regional or cultural factors on taste perception. Future studies should therefore extend the experiment to other continents in order to include a broader diversity of cultural backgrounds, consumption habits, and environmental contexts.

The study confirmed the suitability of **tree-based ensemble** models, particularly **RF**, **GB** and **AdaBoost**, for predicting C2WR from contextual information. However, these findings are specific to the coffee blend used during the experiment and are therefore not yet generalizable. To improve external validity, future research should extend the experimental setup to include additional coffee types and blends, such as mixtures of Arabica and Canephora, or 100 % Arabica and 100 % Canephora beans. Given the natural variability of coffee as an agricultural product, even repeated trials under identical conditions may yield different results. Long-term experimentation over several years will therefore be necessary to capture potential effects of varying crop years. While we focused on the C2WR, other recipe parameters, such as grind size, brewing temperature, and brewing time, should be explored in future work. The next research phase should involve the collection of additional data from a larger and more balanced participant base, using multiple coffee blends. Subsequent analyses could then employ additional machine-learning and statistical approaches to validate and extend the current findings. This would not only improve model quality but also facilitate the identification of features with a statistically significant influence on recipe optimization. Overall, this study provides an empirical and methodological foundation for the development of context-aware recipe optimization systems. By combining real-world data with interpretable and robust machine-learning models, it demonstrates the feasibility of personalized, sustainable, and data-driven food and beverage preparation. The findings represent a significant step toward realizing the CONFES vision, an adaptive system capable of optimizing recipes based on user preferences, contextual conditions, and sustainability goals.

REFERENCES

Allen, A. & Tkatchenko, A., 2022. Machine learning of material properties: Predictive and interpretable multilinear models. *Science Advances*, Issue 8.

Alsallakh, B. et al., 2017. Do Convolutional Neural Networks Learn Class Hierarchy?. *IEEE Transactions on Visualization and Computer Graphics*, Issue 24, pp. 152–162.

Anon., 2023. Boosting Predictive Power: Random Forest and Gradient Boosted Trees in Ensemble Learning. *Proceeding Book of 2nd International Conference on Contemporary Academic Research* ICCAR 2023.

Bergmeir, C. & Benítez, J., 2012. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.*, Issue 191, pp. 192–213.

Chapagain, A. & Hoekstra, A., 2007. The water footprint of coffee and tea consumption in the Netherlands. *Ecological Economics*, 1(64), P. 109–118.

Chatterjee, R. & Yu, T., 2016. Generalized coherent states, reproducing kernels, and quantum support vector machines. *Quantum Inf. Comput.*, Issue 17, pp. 1292–1306.

Clark, M., Springmann, M., Hill, J. & and Tilman, D., 2019. Multiple health and environmental impacts of foods. *Proceedings of the National Academy of Sciences of the United States of America*, Issue 116, p. 23357–23362.

Fernie, A. & Alseekh, S., 2022. Metabolomic selection–based machine learning improves fruit taste prediction. *Proceedings of the National Academy of Sciences of the United States of America*, Issue 119.

Goel, M. et al., 2022. Machine learning models to predict sweetness of molecules. *Computers in biology and medicine*, Issue 152, P. 106441.

Guillén, D. & Rojas, A., 2016. An Empirical Overview of the No Free Lunch Theorem and Its Effect on Real-World Machine Learning Classification. *Neural Computation*, Issue 28, pp. 216–228.

Guinea, M. et al., 2021. A proactive context-aware recommender system for in-vehicle use. *Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (ICVISP 2020)*.

Huynh-Thu, V., Irrthum, A., Wehenkel, L. & Geurts, P., 2010. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, Issue 5.

Krstajic, D., Buturovic, L., Leahy, D. & Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, Issue 6.

Lei, J., 2017. Cross-Validation With Confidence. *Journal of the American Statistical Association*, Issue 115, pp. 1978–1997.

Lundberg, S. et al., 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, Issue 2, pp. 56–67.

Malavolta, M. et al., 2022. A survey on computational taste predictors. *European Food Research and Technology*, Issue 248, pp. 2215–2235.

Miao, Z. et al., 2023. Gustation-Inspired Dual-Responsive Hydrogels for Taste Sensing Enabled by Machine Learning. *Small*.

Mueller, M., Kraus, D., Lukezic, N., Guissouma, H., Sax, E., 2024. An architecture for context-aware food and beverage preparation systems. In: S. N. S. AG, ed. Intelligent Systems and Applications – *Proceedings of the 2024 Intelligent Systems Conference (IntelliSys)*. Amsterdam: s.n., pp. 1–15.

Mueller, M., Kraus, D., Lukezic, N. & Sax, E., 2025. A Data Acquisition System for a Context-Aware Fully Automated Coffee Machine. *Intelligent Systems and Applications*, pp. 486–500.

Mueller, M., Lukezic, N., Luenztel, V., Kraus, D. & Sax, E., 2024. *Dataset: Coffee consumption and user experience dataset with machine and contextual data*. [Online] Available at: https://dx.doi.org/10.35097/h8qfpu2hdzvsxem7

Nemecek, T., Jungbluth, N., Canals, L. & and Schenck, R., 2016. Environmental impacts of food consumption and nutrition: where are we and what is next?. *The International Journal of Life Cycle Assessment*, Issue 21, pp. 607–620.

Notarnicola, B. et al., 2017. Environmental impacts of food consumption in Europe. *Journal of Cleaner Production*, Volume 140, P. 753–765.

Park, S. & Kim, C., 2022. Comparison of tree-based ensemble models for regression. *Communications for Statistical Applications and Methods*.

Pistorius, F., Baumann, D., Seidel, L. & Sax, E., 2020. Intuitive time-series-analysis-toolbox for inexperienced data scientists. *International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 401–406.

Reinhardt, G. & Wagner, 2020. *Environmental footprint of food and dishes in Germany*, Heidelberg: Institut für Energie- und Umweltforschung Heidelberg.

Rousseeuw, P., 2024. Explainable Linear and Generalized Linear Models by the Predictions Plot. *The American Statistician*.

Schreurs, M. et al., 2024. Predicting and improving complex beer flavor through machine learning. *Nature Communications*, Volume 15.

Song, Y. et al., 2023. A Comprehensive Comparative Analysis of Deep Learning Based Feature Representations for Molecular Taste Prediction. *Foods*, Issue 12.

Stang, M., 2025. *Szenariobasierte Validierung selbstlernender Systeme mit individueller Benutzeraktion durch metamorphe Beziehungen*. Dissertation ed. Karlsruhe: Karlsruher Institut für Technologie (KIT).

Surma, S. & Oparil, S., 2021. Coffee and arterial hypertension. *Current Hypertension Reports*, Issue 23.

Wang, Y. et al., 2021. Prediction of flavor and retention index for compounds in beer depending on molecular structure using a machine learning method. *RSC Advances*, Issue 11, pp. 36942–36950.

Wong, T. & Yeh, P., 2020. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, Issue 32, pp. 1586–1594.

Yavas, C., Kim, J. & Chen, L., 2024. Exploring Flavors Through AI: The Future of Culinary Taste Prediction. *IEEE/ACIS 22nd International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 139–147.

Zeng, X. et al., 2023. Food flavor analysis 4.0: A cross-domain application of machine learning. *Trends in Food Science & Technology*.

Zink, M., Grimm, D. & Sax, E., 2025. AURORA Networks: Auto-associative Universal Real-Time Outlier Risk Assessment Networks. *Computer Safety, Reliability, and Security. SAFECOMP 2025 Workshops. SAFECOMP 2025. Lecture Notes in Computer Science*, Issue 15955.