CENTENARY OF THE OIV
45th WORLD CONGRESS OF VINE AND WINE
14-18 OCTOBER 2024 FRANCE-DIJON
OIV
Available on IVES Conference Series vine & wine

DOI : https://doi.org/10.58233/wciWpJNC

Anticipating consumer preference for low-alcohol wine: A machine learning analysis based on consumption habits and socio-demographics

Mingze Rui[1], Federica Rosa[2], Andrea Viberti[2], Filippo Brun[1], Simone Blanc[1] and Stefano Massaglia[1,3]

[1] Department of Agricultural, Forest, and Food Sciences, University of Turin, 10095 Grugliasco, Italy
[2] RV Studio s.r.l, 12051 Alba, Italy
[3] Centro Interdipartimentale Viticoltura e Vino (CONViVi), University of Turin, 12051 Alba, Italy

Abstract. The global wine consumption landscape is undergoing a transformation, marked by a growing trend towards reduced consumption and a preference for healthier lifestyles. However, comprehensive studies investigating consumer preferences for low-alcohol wine remain scarce. This study aims to anticipate consumer preference for low-alcohol wine based on consumers wine consumption habits and their socio-demographics. The data were collected from 2022 to 2024 through online surveys to consumers in France, Germany, Italy and the United States, the four largest wine-consuming countries in 2021. The survey comprised questions regarding consumers' socio-demographics, wine consumption habits and preference for alcoholic content of wine. We employed various machine learning algorithms, treating each survey year as a repetition to identify the best modeling approach. After evaluating the performance of these models, Random Forest emerged as the top performer. The analysis revealed that interactions between socio-demographic factors and buying habits factors are significant predictors of low-alcohol wine preferences. The study provides actionable insights for wine producers and marketers, offering a data-driven approach to more effectively target low-alcohol wine markets using accessible customer data such as demographics and sales records.

1. Introduction

Wine consumers have expressed increasing concerns regarding the alcohol content of wines. Despite evidence indicating that light to moderate wine consumption can confer health benefits, such as reduced mortality and morbidity due to its ethanol and phenolic contents [1], there is a growing preference for lower-alcohol wines [2]. This trend highlights consumer anxiety about the potential negative health impacts of alcohol consumption, even at moderate levels. The rising interest in low-alcohol wines underscores a significant shift in consumer behavior towards minimizing alcohol intake [3]. This shift is driven by broader health-related concerns, including the potential risks of alcohol consumption on overall well-being, and reflects an increasing awareness of the need for healthier lifestyle choices [4]. Furthermore, consumers are becoming more aware of the ethical implications of wine production and marketing practices, particularly how they may affect public health and societal norms surrounding alcohol consumption [5]. This awareness has led to varying levels of concern about the effects of wine production practices on health and the environment, offering an opportunity for wine marketers to positively influence consumer perceptions by adopting more responsible and transparent practices [6].

The global consumption of alcohol-free or low-alcohol beverages has seen a notable increase, with a 5% rise in key markets in 2023 [7]. Among the world's top ten no/low alcohol markets, low-alcohol wine is expected to lead the growth, with a projected compound annual growth rate (CAGR) of 12% from 2023 to 2027 [8]. The growing trend of the low-alcohol wine market is driven by multiple factors, including perceived health benefits [1,9,10], environmental advantages [11,12], and shifting consumer preferences [13,14]. However, the market faces many challenges such as product availability, quality perceptions, taste issues, and a lack of awareness about low-alcohol wine options [13]. Efforts are ongoing to balance acidity, sweetness, and body taste in dealcoholized wines to improve their overall appeal [3,15]. Although there is a high level of awareness of health issues pertaining to alcohol consumption, initial reluctance to try non-alcoholic options persists.

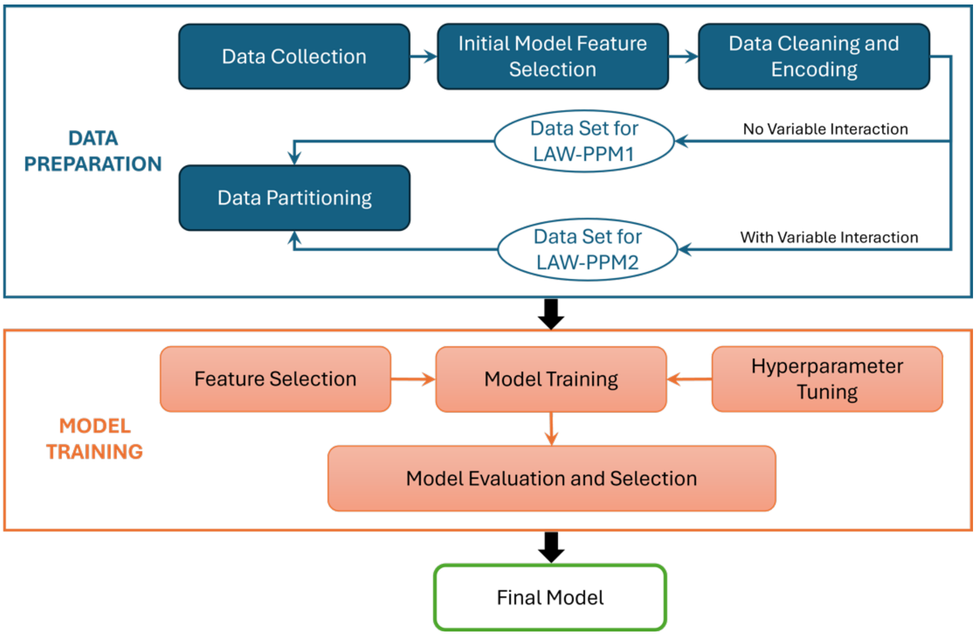

Figure 1. Workflow for generation of models using machine learning algorithms.

Consumers' preferences for low-alcohol or non-alcohol wines are closely linked to various socio-demographic factors and drinking habits. Specifically, females, younger individuals, and less frequent wine drinkers are more likely to favor low-alcohol wines, especially in the UK [13] and Italy [16,17]. Specifically, in the UK, low-alcohol wine consumers are typically Millennials and Baby Boomers with mid-to-low incomes, who tend to prefer white and rosé wines [13]. In the US, the growing consumer interest in moderating alcohol consumption is largely driven by health-conscious Millennials [18]. In France, less frequent wine drinkers show greater acceptance of partially dealcoholized wines [4]. Similarly, in Australia, younger consumers are more receptive to non-alcoholic red wines [19]. These findings and facts highlight the significant role that socio-demographic factors and drinking habits play in shaping consumer preferences for low-alcohol and non-alcoholic wines.

Machine learning has transformed how we predict consumer preferences, with advanced algorithms like random forests, decision tree, and gradient boosting offering deeper insights than traditional methods like the Affinity Diagram and Analytic Hierarchy Process (AHP) methods [20–22]. Studies have proved the effectiveness and reliability these machine learning methods in predicting consumer preferences and market trends [22,23]. Building on the observation that socio-demographic factors and drinking habits significantly influence consumer preferences for low-alcohol wines across various countries, this study aims to develop and evaluate predictive models using advanced machine learning algorithms. Our goal is to develop a model that can accurately predict consumer preferences for low-alcohol wines, thereby enabling wine marketers to better target their audience.

Although studies on consumer preferences for low-alcohol wines have not yet explored the moderating role of socio-demographic factors, previous research on wine consumer behavior has identified the influence of demographic variables such as age, income and gender [24]. To address this gap, we will develop two models: one that considers demographic factors as predictors and another that examines them as moderators. This will allow us to assess both direct and interactive effects of demographic variables on consumer preferences, giving a more comprehensive understanding of how these factors shape wine consumption behavior.

2. Methods

An overview of the machine learning analysis workflow is presented in Figure 1.

2.1. Data collection and preparation

Data were collected from 2022 to 2024 through online surveys distributed to agro-food consumers in France, Germany, Italy, and the United States, the four largest wine-consuming countries in 2021 according to the International Organization of Vine and Wine (OIV). The questionnaire was conducted anonymously and did not collect any sensitive information. Prior to participation, respondents provided informed consent after reviewing a project disclosure sheet detailing the survey's objectives. The research was carried out in accordance with the principles of the Declaration of Helsinki. The survey included questions about consumers' preferences and purchasing habits related to agro-food products, including wine, as well as their socio-demographic information. Based on the literature review presented in the previous section, we selected relevant features that influence consumer preferences for low-alcohol wine, including socio-demographic factors and wine-buying habits. The variables used in the machine learning analysis are detailed in Table 1.

Table 1. Variables used in the prediction models.

Variable Category	Variable in Original Survey	Variable Transformation
Socio-demographics	Country of residence	No transformation applied
	Gender	No transformation applied
	Age (Year of born)	Generation (Gen Z, Millennials, Gen X, Baby Boomers)
	Income (Annual income range)	Income level in their respective country (low, medium and high)
	Occupation type (Specific category)	General occupation type (entrepreneur/self-employed, not currently employed, employed, other)
Wine buying habits	Frequency of buying different types of wine (still white, still red, sparkling, rosé, dessert)	No transformation applied
	Channels where habitually consume or buy wine	Habitual buying or consuming wine on-premise
		Habitual buying or consuming wine off-premise
	Country of origin of wine habitually consume or buy	Habitual buying or consuming imported wine
Preference for low-alcohol wine	Preference for alcohol content of wine	Preference for low-alcohol wine (lower than 11% ABV)

To facilitate the machine learning analysis, some variables were transformed for simplification. These transformations included categorizing participants by generational cohorts based on their birth year, standardizing income into low, medium, and high levels relative to their country, and grouping occupations into broader categories such as employed or self-employed. Wine consumption and purchase channels were grouped into on-premise and off-premise categories, while wine origin was classified as either domestic or imported. Preferences for alcohol content were also transformed: those favoring low-alcohol wines were defined as respondents selecting wines with an alcohol content lower than 11% ABV, whereas those with no specific preference were those selecting wines with 11% ABV or higher or indicating that alcohol content was not important.

Furthermore, to facilitate the machine learning analysis, these categorical data were one-hot encoded. To determine the most effective method for predicting consumer preferences for low-alcohol wines, we propose two models. The Low-Alcohol Wine Preference Prediction Model 1 (LAW-PPM1) considers all factors with a direct effect on preferences, while the Low-Alcohol Wine Preference Prediction Model 2 (LAW-PPM2) explores socio-demographic factors as moderators by examining two-way interactions between socio-demographic variables and drinking habits. Following the construction of the datasets for LAW-PPM1 and LAW-PPM2, each dataset was split into training (70%) and testing (30%) subsets.

2.2. Model training using machine learning algorithms

We employed seven machine learning methods: Gradient Boosting Machines (GBM), Decision Trees, Random Forest (RF), Rotation Forest (RTF), K-Nearest Neighbors (k-NN), Naïve Bayes (NB) classifier, and Support Vector Machines (SVM). Each method was applied to both LAW-PPM1 and LAW-PPM2 datasets across all survey years. Feature selection was refined using backward elimination to enhance model performance. Additionally, hyperparameter tuning was conducted for each classification model to determine the optimal settings. Model performance was evaluated using several metrics, including accuracy, precision, recall, and F1-score, with a primary focus on accuracy and F1-score. All analyses were executed using R Studio [25].

3. Results

3.1. Survey response summary and preference trends

A total of 7,703 valid responses were collected across the survey years 2022 to 2024 (Table 2). The distribution of responses from France, Germany, Italy, and the USA was relatively balanced each year, with Italy having the highest number of respondents in 2022 and 2023, and the USA in 2024. Additionally, a growing participation over time is observed with the number of respondents increased in each subsequent year.

Table 2. Valid survey responses by country and survey year.

Survey Year	France	Germany	Italy	USA	Total
2022	594	441	640	528	2203
2023	544	657	844	609	2654
2024	642	594	803	807	2846

Figure 2 shows survey respondents' preferences for low-alcohol wine over three survey years. In 2022, 806 respondents (36%) expressed a preference for low-alcohol wine. This percentage increased to 41% in 2023 and slightly decreased to 40% in 2024. However, across all three years, the majority of respondents showed no preference for low-alcohol wine.

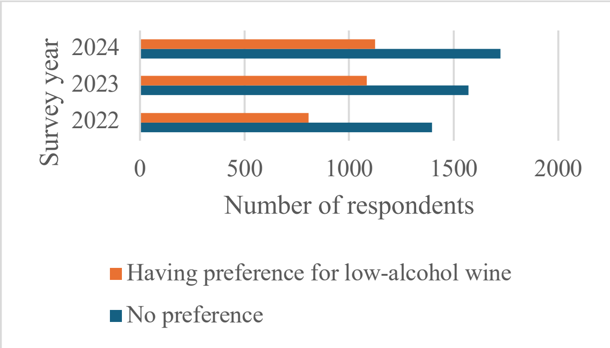

Figure 4. Survey respondents' preferences for low-alcohol wine (2022-2024).

3.2. Experimental results

3.2.1. Assessments of model performance and comparison

The results of the model testing for predicting consumer preference for low-alcohol wine show a varied performance across different machine learning algorithms and model configurations. Heatmap of metrics for different models are shown in Figure 3.

For the LAW-PPM1 model, which considers both demographic and habit variables as predictors, the accuracy ranges from 0.5330 (KNN) to 0.6314 (SVM).

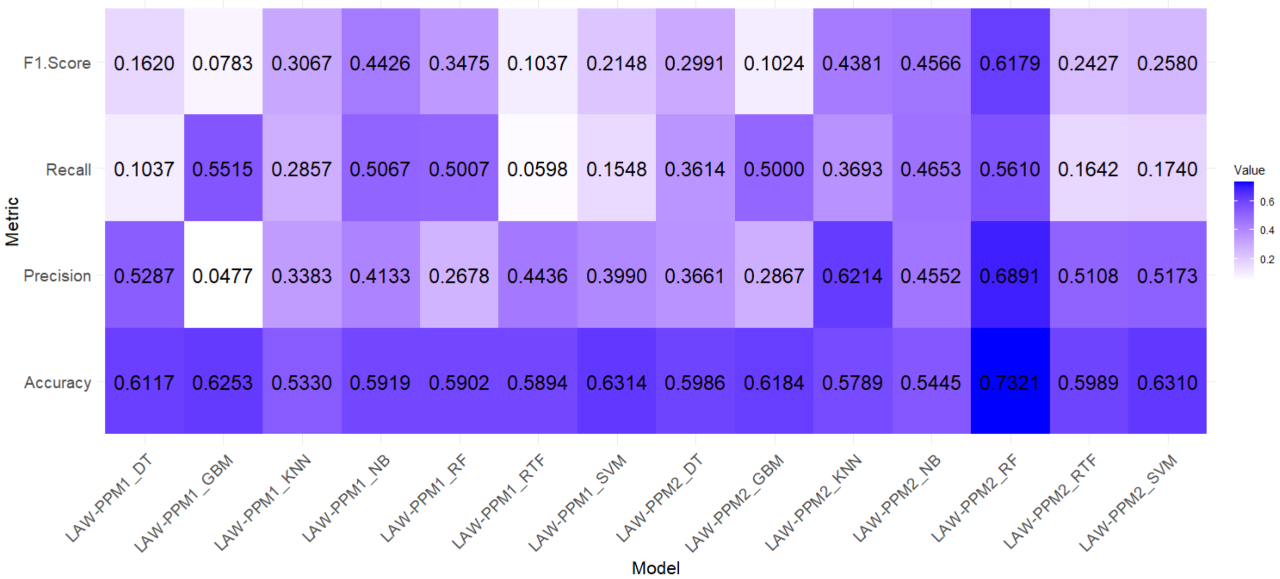

Figure 3. Heatmap of metrics by model.

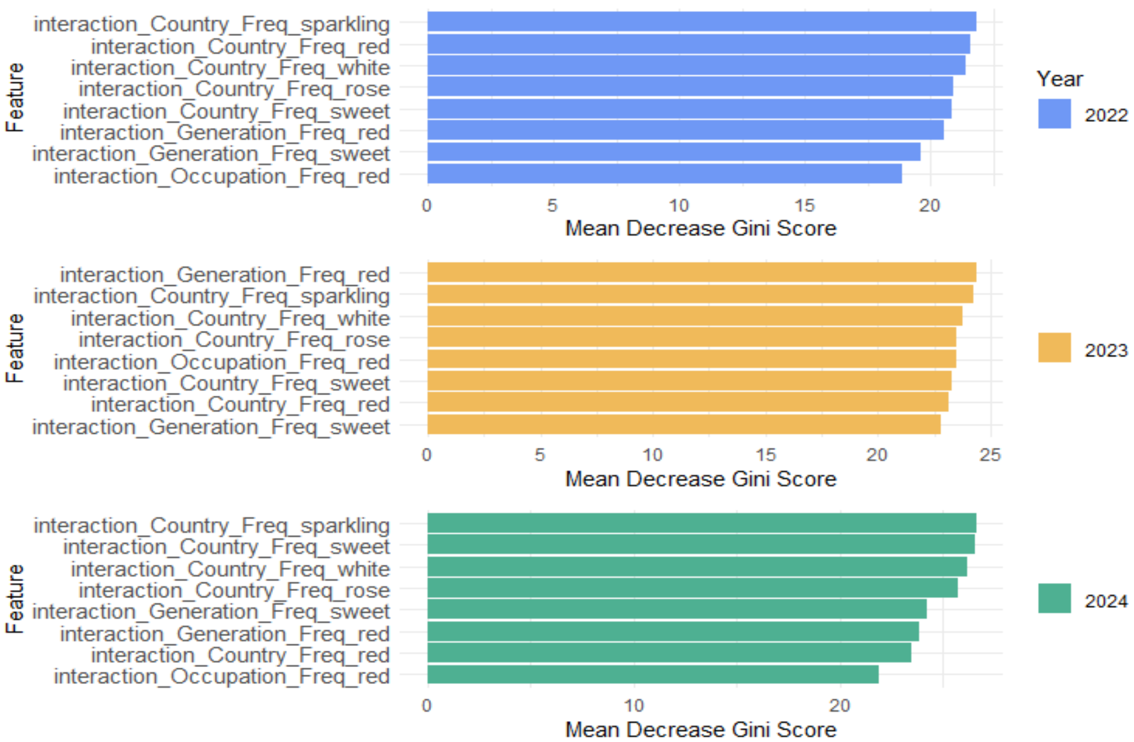

Figure 2. Top 8 important features using LAW-PPM2 with Random Forest algorithm.

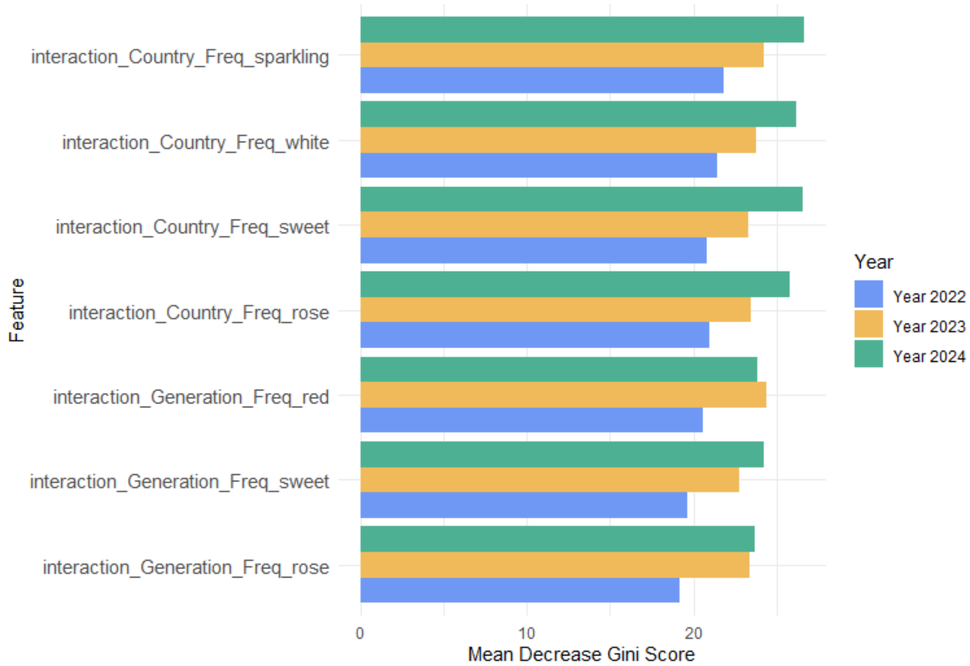

Figure 5. Important features ranked in top 10 across all years, using LAW-PPM2 with Random Forest algorithm.

The GBM model achieved the highest accuracy (0.6253) but had a very low precision (0.0477) and F1-Score (0.0783), indicating that despite correctly classifying many instances, it struggled with correctly identifying true positives. The RF model demonstrated a balanced trade-off with a reasonable F1-Score (0.3475) but had lower accuracy (0.5902). The NB model also showed a balanced performance with moderate scores across the metrics, achieving a slightly higher F1-Score (0.4426) compared to others.

On the other hand, the LAW-PPM2 model, which treats demographic variables as mediators and only considers habit variables as predictors, generally showed better performance, especially with the RF algorithm, which achieved the highest accuracy (0.7321) and also had the best overall performance across all metrics, including a high precision (0.6891), recall (0.5610), and F1-Score (0.6179). This suggests that when demographic factors are considered as mediators rather than direct predictors, the model's predictive power improves significantly. Other algorithms like KNN and NB also showed reasonable performance under the LAW-PPM2 model, with KNN achieving a high precision (0.6214) and NB maintaining a balanced performance across metrics.

In conclusion, the results suggest that the LAW-PPM2 model with the Random Forest (RF) algorithm is the most effective in predicting consumer preference for low-alcohol wine. This configuration not only achieved the highest accuracy but also balanced precision, recall, and F1-Score, indicating a relatively strong overall performance.

3.2.2. Feature importance for selected model

To further understand the roles of factors in predicting low-alcohol wine preference, feature importance was calculated using LAW-PPM2 model with the RF algorithm. Figure 4 shows the top eight important features for each survey year.

The importance of features is assessed based on the Mean Decrease Gini Score, which reflects how much each feature contributes to reducing uncertainty (or impurity) in the model's predictions. The most important features for predicting consumer preference for low-alcohol wine vary slightly across the years 2022, 2023, and 2024, but consistently highlight the significance of geographic and buying frequency patterns. Interactions between a consumer's country of residence and their frequency of consuming different types of wine (sparkling, red, white, rosé, and sweet) are consistently top predictors. The importance of generational differences in wine consumption, particularly for red wine, becomes more pronounced in 2023, reflecting changing preferences among red wine drinkers in different age groups. Additionally, the interaction between occupation and the frequency of red wine consumption remains a significant predictor, indicating that occupational factors, likely linked to lifestyle, influence consumer preferences. These findings emphasize the dynamic nature of wine preferences and the critical role of both demographic and behavioral variables in predicting consumer choices over time.

Seven features consistently ranked within the top 10 most important predictors of consumer preference for low-alcohol wine across the years 2022, 2023, and 2024, as shown in Figure 5. These features are all interactions between a consumer's country of residence or generation and their frequency of consuming different types of wine.

The consistent presence of these seven features among the top predictors across all years suggests that wine type preferences linked to a consumer's location, age group and consumption frequency are crucial for understanding their inclination toward low-alcohol options. The most important predictor across all years is the interaction

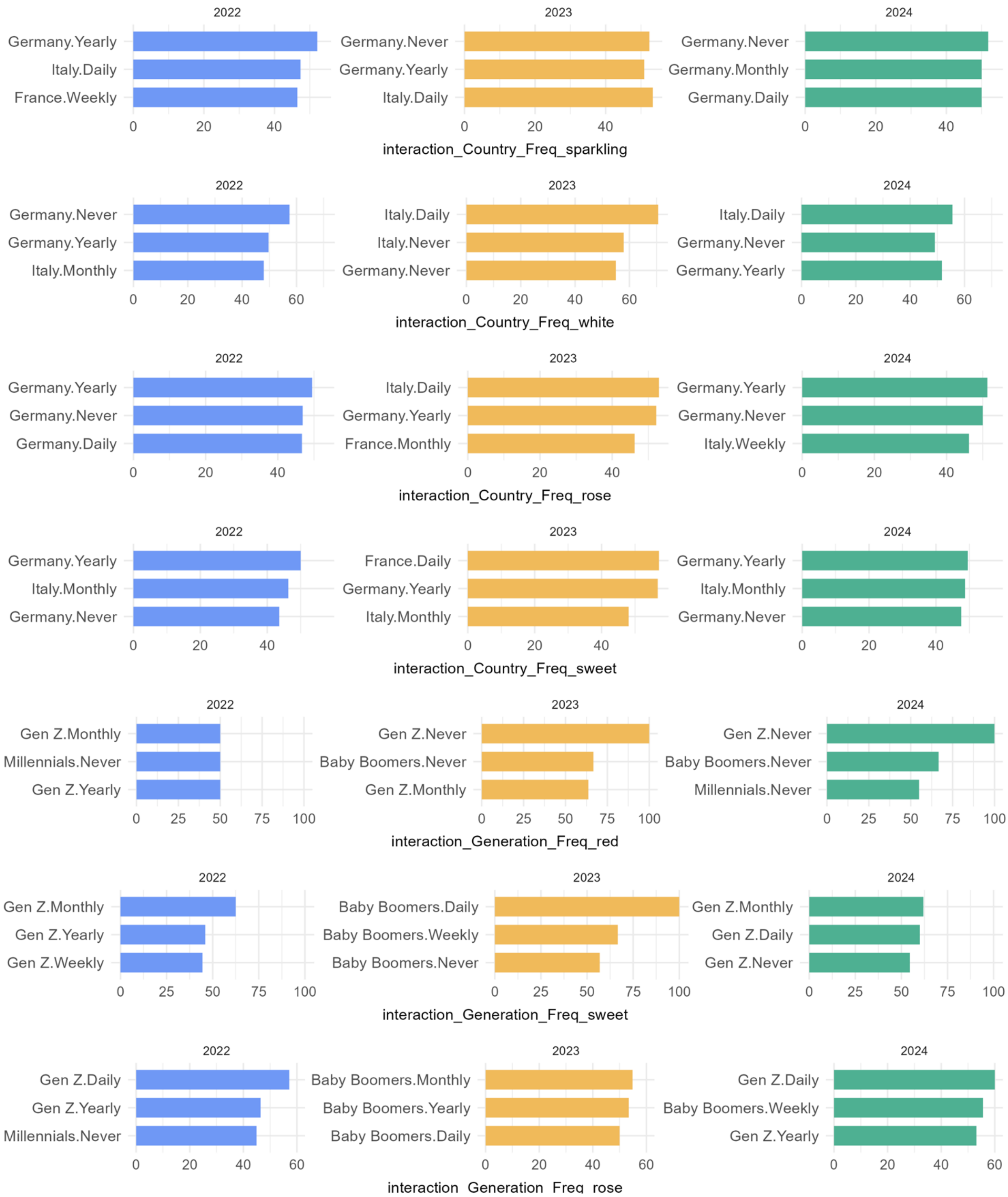

Figure 6. Descriptive Statistics Summary for Important Features, the percentage of respondents with a preference for low-alcohol wine is displayed.

between country and sparkling wine consumption, indicating the strong influence that a consumer's geographic location and their sparkling wine consumption habits have on predicting their preference for low-alcohol wine.

The other features, particularly the interaction between country and frequency of white, sweet, and rosé wine consumption, also consistently rank highly, though with slightly varying importance each year. This indicates that consumers' preferences for low-alcohol wine are significantly shaped by their habitual consumption of these wine types, influenced by their country of residence. Lastly, the generation-specific consumption patterns, particularly for red and sweet wines, also play an important role in predicting low-alcohol wine preferences.

Descriptive statistics were summarized for these seven features to further understand the patterns and variations in consumer preferences for low-alcohol wine.

It can be seen from Figure 6 that German consumers with low consumption frequency regardless the wine type consistently prefer low alcohol options, while Italian consumers show fluctuating preferences depending on the consumption pattern of different wine types. Generationally, Gen Z exhibits a strong and consistent preference for low alcohol wine, particularly for the monthly sweet wine drinkers. Moreover, Millennials who never drank rosé or red wine, as well as Baby Boomers who drank rosé more frequently, showed a greater preference for low-alcohol options. It is important to note that some of the data may be biased due to the small number of respondents in certain groups. Specifically, the number of Gen Z respondents who drink rosé, red, and sweet wine either not at all or daily was fewer than 10 each year across all three survey years. Similarly, the group of Baby Boomers who drink sweet wine daily was also very small. These limited sample sizes may affect the reliability of the findings for these particular groups.

4. Discussion

By testing and evaluating the LAW-PPM1 and LAW-PPM2 models with various machine learning algorithms, we have identified the LAW-PPM2 model using the Random Forest (RF) algorithm as the most effective for predicting consumer preferences for low-alcohol wine. Empirically, RF has proven to be effective in predicting wine preferences, as evidenced by its use in wine recommendation systems [26]. Moreover, another study on predicting consumer preferences for labeled food products using machine learning highlighted the impact of demographics and purchasing habits [27], which aligns with our findings.

It was found that incorporating both genetic and socio-demographic data significantly improved the accuracy of predicting consumers' preferences for alcoholic beverages compared to using socio-demographic data alone. Genetic information, including polygenic scores for behavioral traits such as depression and height, as well as genetic variants related to bitter taste perception, made substantial contributions to these predictions. [28]. While our study does not include genetic data, the significant role of demographic interactions, such as generation and country with wine consumption habits, suggests that integrating additional data sources could further refine predictive models.

The study has several limitations. One key limitation is the reliance on self-reported data for demographic information and wine consumption habits, which may introduce biases such as social desirability or inaccurate recall. Furthermore, the models were tested with a relatively narrow range of machine learning algorithms; other methods, such as ensemble approaches, might offer better predictive performance [29]. Additionally, the generalizability of the models may be limited, as they were developed and tested on a specific dataset, and their performance could vary across different populations or contexts.

5. Conclusion

This study aimed to predict consumer preferences for low-alcohol wine by utilizing machine learning algorithms that incorporate both socio-demographic factors and wine buying habits. Two models, LAW-PPM1 and LAW-PPM2, were developed and tested across various algorithms, with LAW-PPM2, particularly when using Random Forest, showing superior predictive performance. The findings emphasize the significance of interactions between socio-demographic factors and buying habits in predicting consumer preferences for low-alcohol wine. The proposed model provides valuable insights for wine producers and marketers, enabling more effective targeting of low-alcohol wine markets. By leveraging readily accessible customer data, such as demographic profiles and sales records, the model helps refine marketing strategies and optimize market reach.

6. References

1. R. S. Jackson, in *Wine Sci.* (Elsevier, 2014), pp. 889–920
2. C. S. Stockley, C. Varela, A. Coulter, P. R. Dry, I. L. Francis, R. Muhlack, and I. S. Pretorius, in *Wine Types Prod. Health* (2012), pp. 281–299
3. T. Bucher, E. Frey, M. Wilczynska, K. Deroover, and S. Dohle, Public Health Nutr. **23**, 1939 (2020)
4. S. Meillon, V. Dugas, C. Urbano, and P. Schlich, Am. J. Enol. Vitic. **61**, 42 (2010)
5. M. Ramšak, in *Soc. Impact Wine Mark.* (Springer International Publishing, Cham, 2022), pp. 107–109
6. S. L. Forbes, R. Cullen, D. A. Cohen, S. D. Wratten, and J. Fountain, J. Wine Res. **22**, 79 (2011)
7. IFN, Ital. Food News (2023)
8. S. Aswani, IWSR (2024)
9. S. Abreu, M. Liz Martins, and C. Morais, in *Adv. Bot. Res.* (Elsevier, 2024), p. S0065229624001307
10. R. Golan, Y. Gepner, and I. Shai, Eur. J. Clin. Nutr. **72**, 55 (2019)
11. A. Contreras, C. Hidalgo, S. Schmidt, P. A. Henschke, C. Curtin, and C. Varela, Int. J. Food Microbiol. **205**, 7 (2015)
12. C. Varela, P. R. Dry, D. R. Kutyna, I. L. Francis, P. A. Henschke, C. D. Curtin, and P. J. Chambers, Aust. J. Grape Wine Res. **21**, 670 (2015)
13. J. Bruwer, V. Jiranek, L. Halstead, and A. Saliba, Br. Food J. **116**, 1143 (2014)
14. A. Querol, R. Pérez-Torrado, J. Alonso-del-Real, R. Minebois, J. Stribny, B. M. Oliveira, and E. Barrio, in *Adv. Food Nutr. Res.* (Elsevier, 2018), pp. 177–210

15. L. Liguori, D. Albanese, A. Crescitelli, M. Di Matteo, and P. Russo, J. Food Sci. Technol. **56**, 3707 (2019)
16. S. Naspetti, F. Alberti, M. Mozzon, S. Zingaretti, and R. Zanoli, Br. Food J. **122**, 2621 (2019)
17. A. Stasi, F. Bimbo, R. Viscecchia, and A. Seccia, Wine Econ. Policy **3**, 54 (2014)
18. S. Aswani, Wine Intell. (2022)
19. I. Day, K. Deroover, M. Kavanagh, E. Beckett, T. Akanbi, M. Pirinen, and T. Bucher, Int. J. Gastron. Food Sci. **35**, 100886 (2024)
20. P. S. Dahake, R. V. Mohare, and N. Somani, in *2023 Second Int. Conf. Smart Technol. Smart Nation SmartTechCon* (IEEE, Singapore, Singapore, 2023), pp. 1347–1352
21. V. R. S. Mahala, N. Garg, D. Saxena, and R. Kumar, in *Data Sci. Appl.*, edited by S. J. Nanda, R. P. Yadav, A. H. Gandomi, and M. Saraswat (Springer Nature Singapore, Singapore, 2024), pp. 207–218
22. J. Martínez-Garmendia, J. Mark. Anal. (2023)
23. Prakash, S. M. Babu, P. P. Kumar, S. Devi, K. P. Reddy, and M. Satish, in *2023 IEEE 5th Int. Conf. Cybern. Cogn. Mach. Learn. Appl. ICCCMLA* (IEEE, Hamburg, Germany, 2023), pp. 687–692
24. V. Neelavar Kelkar, J. Mallya, V. Payini, and V. Kamath, F1000Research **13**, 228 (2024)
25. Posit team, (2024)
26. W. Mu, Y. Feng, H. Shu, B. Wang, and D. Tian, Ital. J. Food Sci. **34**, 98 (2022)
27. Y. Shen, J. A. Hamm, F. Gao, E. T. Ryser, and W. Zhang, J. Food Prot. **84**, 1560 (2021)
28. C. Zhu, T. Beatty, Q. Zhao, W. Si, and Q. Chen, China Agric. Econ. Rev. **15**, 685 (2023)
29. T. Larkin and D. McManus, Stat. Anal. Data Min. ASA Data Sci. J. **13**, 451 (2020)