Comparing Fruit Consumer Acceptance Models Using Different Machine Learning Algorithms

Simon Kollaart[1], Jonas Inderbitzin[1], Moritz Köhle[2]
[1] *Agroscope, Food Microbial Systems,*
[2] *Agroscope, Competence Division Plants and Plant Products*

Introduction

Predicting consumer acceptance of fruit, based on instrumentally measurable traits, can guide variety testing in the scientific field and help optimize product quality at retail. The aim of this study was to model the acceptance of plums with multiple machine learning algorithms.

Method

Individual-level data:
80 consumer liking scores x 30 samples x 10 analytical predictors = 2400 observations.
Aggregate-level data:
1 mean consumer liking score x 30 samples x 10 analytical predictors = 30 observations.
Goal:
Predicting consumer liking scores with various regression algorithms and different resampling techniques.
Analytical predictors included °Brix, Acid, Firmness, Color, etc.

Results

Individual-level		Aggregate-level				
Cross-Validation		Bootstraps		Cross-Validation		
RMSE	R2	RMSE	R2	RMSE	R2	Model
		8.5	0.32	8.3	0.29	Boosttree (lightgbm)
		8.2	0.23	8.1	0.30	Boosttree (xboost)
19.2	0.21	7.3	0.32	3.2	0.92	Random Forest
20.0	0.14	5.3	0.66	4.9	0.65	LinearRegression (glm)
20.0	0.14	5.3	0.66	4.9	0.65	Linear Regression (lm)
19.3	0.20	8.9	0.20	3.0	0.87	Decision Tree
19.7	0.19	5.2	0.67	3.3	0.86	Support Vector Machine
19.2	0.21	5.6	0.62	3.0	0.87	Neuralnet

Table 1: Performance metrics of consumer liking models. Cross-Validation (n=10), Bootstraps n=100

Model performance reveals good accuracy at the aggregate level across cross-validation and bootstrapping methods for a number of different algorithms. Conversely, individual-level models show poor predictive power.

Boosttree models did not perform well and may require more careful hyperparameter tuning to achieve optimal performance.

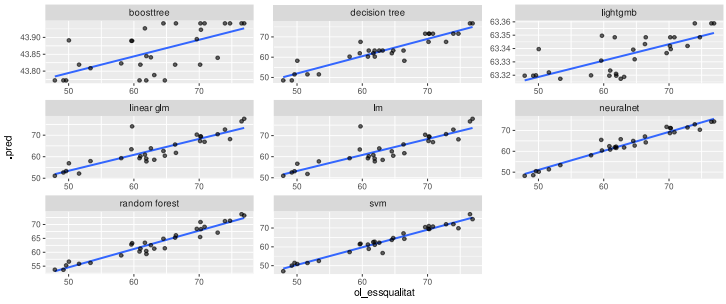

Figure 1: Scatterplots of consumer liking models using the aggregate-level data

To determine the smallest subset size that is statistically similar to the full dataset with 80 observations at a 95 % confidence level we used a Monte Carlo Simulation with 10'000 iterations. The results in Figure 2 suggest that using a subset size of 50 would provide a mean that is statistically similar to that of the full dataset, with a 95% confidence level.

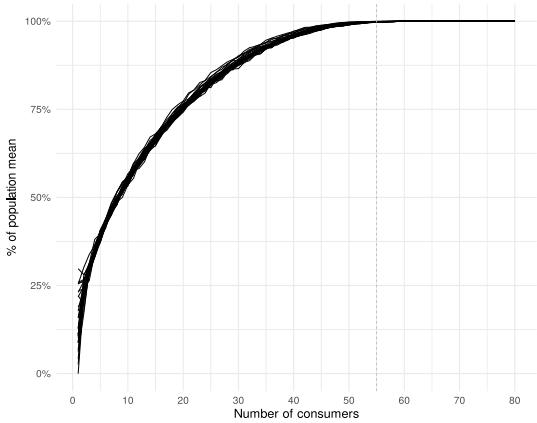

Figure 2: Plot showing the results of a simulation estimating the % of means that fall into the 95 % confidence interval of the population mean (10'000 iterations).

Conclusion

All models struggled to accurately predict consumer liking scores at the individual level. Aggregating the data by creating means reduces noise and the influence of outliers, leading to a cleaner signal that may be easier to model. Reducing the number of consumers from 80 to 50 would not only maintain a statistically similar mean but also enhance the efficiency of modeling by potentially increasing the number of samples and reducing costs through the use of fewer participants.