How to make toast 🍞
With delicious CURST?

CC BY NC

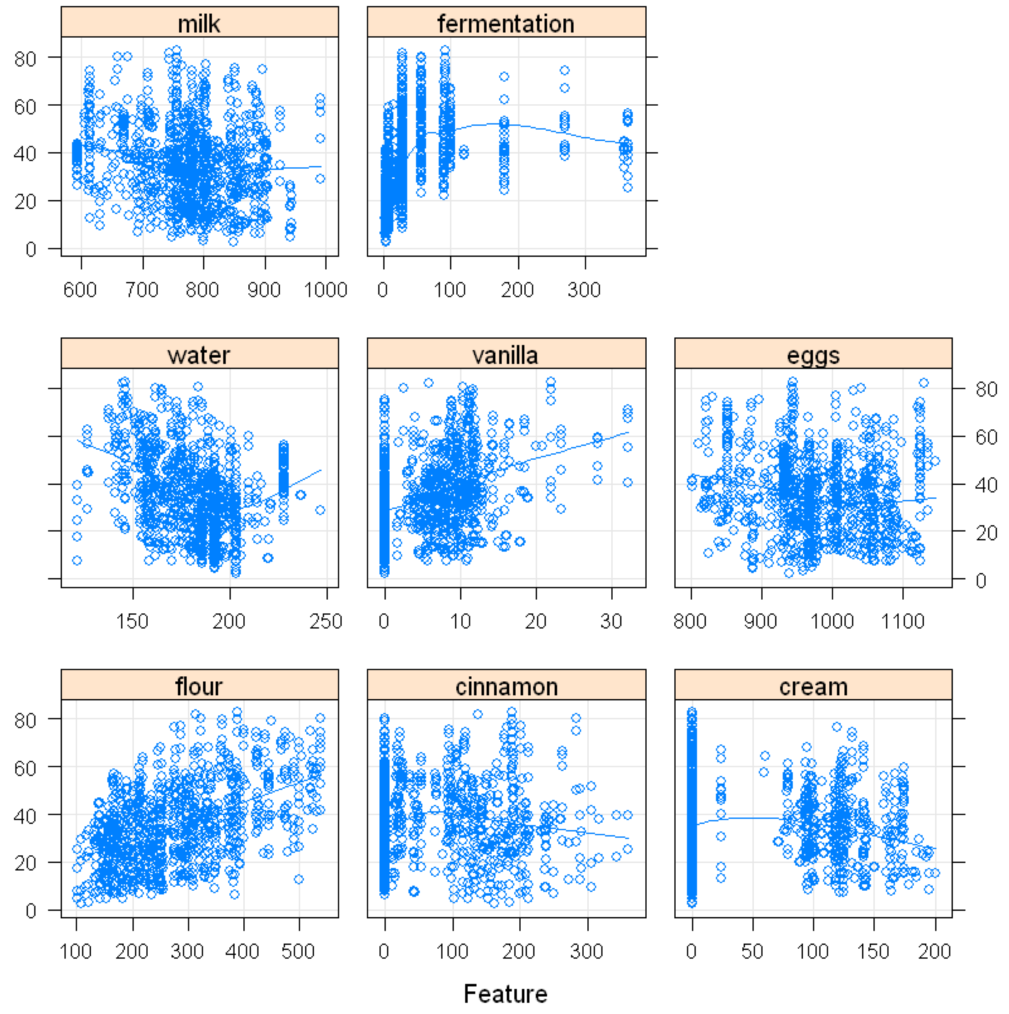
Let Machine Learning do t. magic

A Quick Look at the Data Structure

1030 obs. of 9 variables

	flour	cinnamon	cream	water	vanilla	eggs	milk	fermentation	rating
	540	0	0	162	2.5	1040	676	28	79.99
	540	0	0	162	2.5	1055	676	28	61.89
	332.5	142.5	0	228	0	932	594	270	40.27
	...	...	...	...	...	...	...	...	...
	198.6	132.4	0	192	0	978.4	825.5	360	44.3
	266	114	0	228	0	932	670	90	47.03
	380	95	0	228	0	932	594	365	43.7
	380	95	0	228	0	932	594	28	36.45
Summary									
1st Qu.	192.4	0	0	164.9	0	932	731	7	23.71
3rd Qu.	350	142.9	118.3	192	10.2	1029.4	824	56	46.13
Max.	540	359.4	200.1	247	32.2	1145	992.6	365	82.6
Mean	281.2	73.9	54.19	181.6	6.205	972.9	773.6	45.66	35.82
Median	272.9	22	0	185	6.4	968	779.5	28	34.45
Min.	102	0	0	121.8	0	801	594	1	2.33

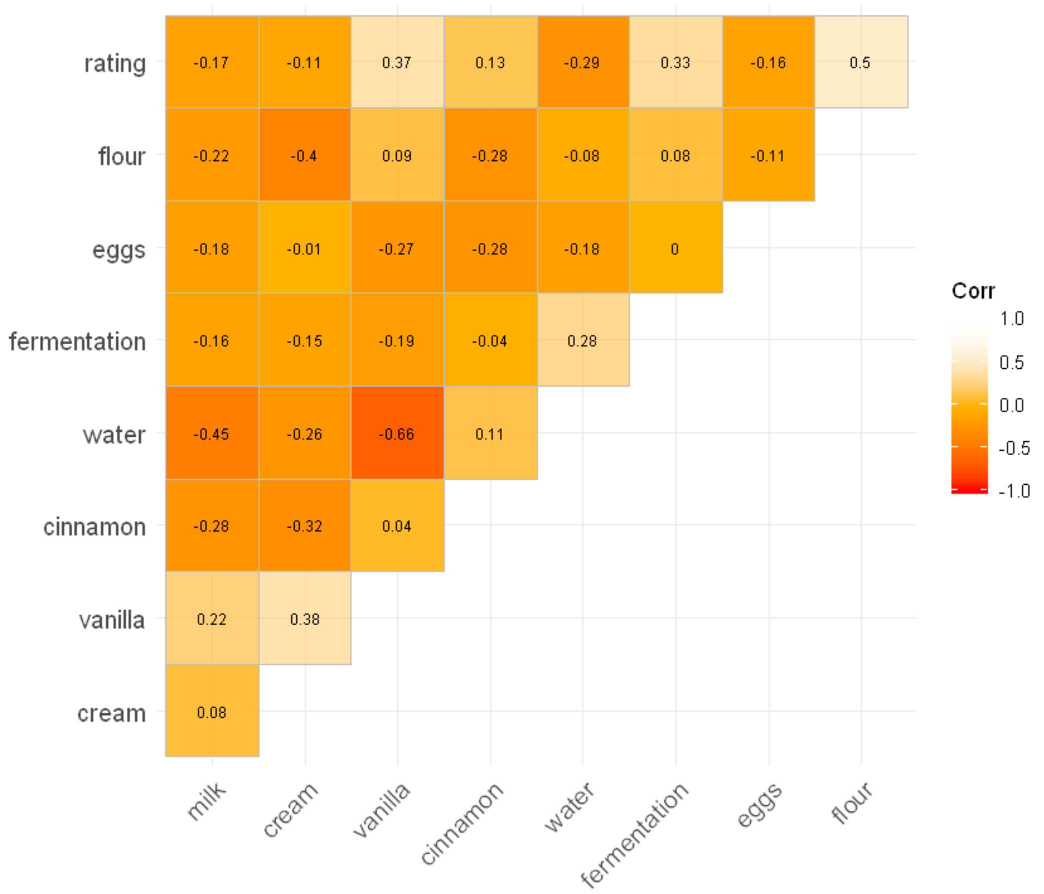
Discover & Visualize the Data



Predictor versus Response

- fermentation shows a strong nonlinear relationship with the predictor, and the flour amount has a linear relationship.
- several of the ingredients have a large frequency of a single amount, such as zero for the vanilla and cream.
- In these cases, the consumer rating of how delicious curst taste varies widely for those values of the predictors.

Looking for Correlations



Pairwise Correlations

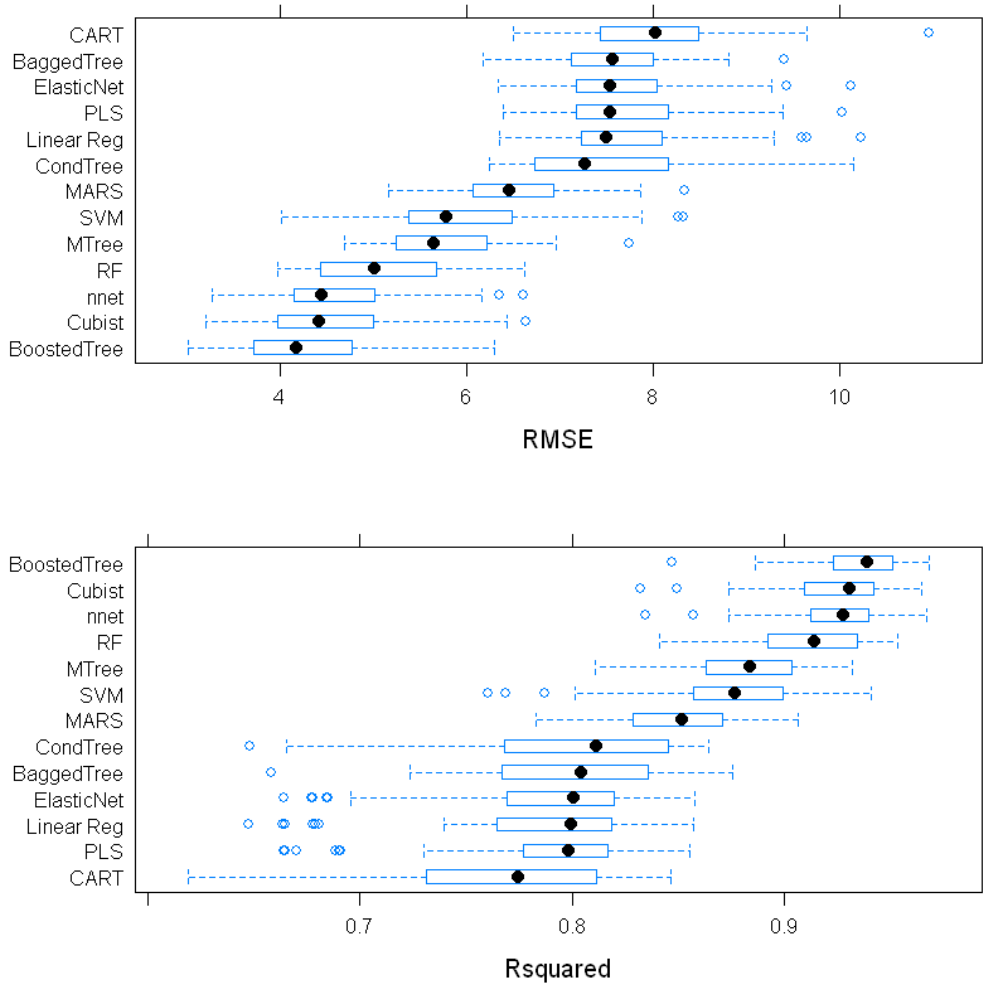
- In this case, the predictors will enter the models as the proportion of the total amount. Because of this, there is a built-in dependency in the predictor values.
- Despite this, the pairwise correlations are not large, and, therefore, we would not expect methods that are designed to deal with collinearity (e.g., PLS, ridge regression) to have performance that is superior to other models.

Model Building Strategy

A suite of models were tested

Linear Regression	Nonlinear Regression	Regression Trees
➢ Linear regression	➢ SVMs	➢ Regression trees
➢ Partial least squares	➢ Neural network	➢ Cubist
➢ Elastic net	➢ MARS models	➢ Bagged & boosted reg. trees

Model Performance



Resampling results across all models

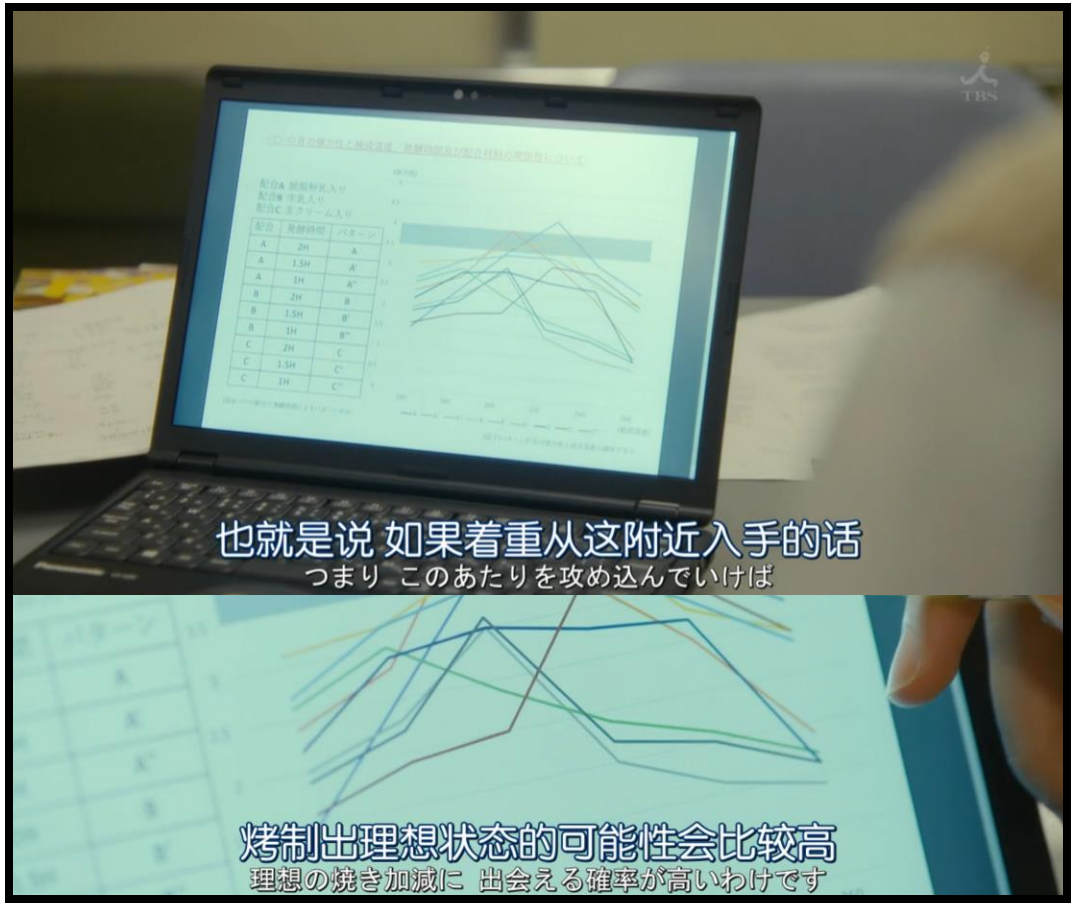
- From this, the top performing models were tree ensembles (random forest and boosting), rule ensembles (Cubist), and neural networks.
- Linear models and simple trees did not perform well.
- Bagged trees, SVMs, and MARS showed modest results but are clearly worse than the top cluster of models.

Optimizing Consumer Rating

Numerical search

- The neural network and Cubist models were used to determine possible mixtures with improved consumer rating.
- To do this, a numerical search routine can be used to find formulations with high consumer rating (as predicted by the model).
- Once a candidate set of mixtures is found, additional experiments would then be executed for the mixtures to verify that the taste of curst has indeed improved.

Optimizing Consumer Rating



The top 3 optimal mixtures

	Cubist			Neural network		
New Mix	mix 1	mix 2	mix 3	mix 4	mix 5	mix 6
flour	12.7	21.7	14.6	34.4	22.2	40.8
cinnamon	14.9	3.4	13.7	7.9	11.6	4.9
cream	6.8	5.7	0.4	0.2	0.1	6.7
vanilla	0.5	0.3	2	0.3	1.1	0.7
eggs	34	33.7	35.8	21.1	27.8	20.5
milk	25.7	29.9	27.5	21.1	27.8	20.5
water	5.4	5.3	6	5.1	5.8	6.1
fermentation	28	28	28	28	28	28
Prediction	89.1	88.4	88.2	88.7	85.7	83.9

Bon Appetit!

Special Acknowledgement & References

- Applied Predictive Modeling by Kuhn and Johnson (2013)
- 義母と娘のブルース /Gibo to Musume no Blues by TBS
- Yeh I (1998). "Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks." Cement and Concrete research, 28(12),1797–1808

randallzoeng@qq.com