Technical report: beverage pH prediction

Mauricio Claudio

2023-04-10

Executive Summary

This report presents the findings of an analysis to determine the factors predictive of pH in ABC Beverage products. The analysis applies five predictive Machine Learning models to beverage manufacture process data for each of the four ABC Beverage products - Brand A, B, C and D - and selects one model based on predictive performance on resampled data. Twenty models were tested in total, and their error performance measured with the Root Mean Square Error (RMSE) metric. The results of the analysis are summarized below:

- The data has a substantial number of missing entries, pointing to possible gaps in data collection and management, and/or sensor malfunction and need for maintenance.
- Random Forest best predicts pH for all four ABC Beverage products. Model predictive error is as follows:

> **Brand A**, 0.11
> **Brand B**, 0.09
> **Brand C**, 0.14
> **Brand D**, 0.08

- Manufacturing process factors most predictive of pH are, in descending order of importance, as follows:

> Brand A, **Mnf.Flow**, **Filler.Level**, **Usage.cont**, **Bowl.Setpoint**
> Brand B, **Mnf.Flow**
> Brand C, **Oxygen.Filler**, **Carb.Rel**
> Brand D, **Usage.cont**, **Mnf.Flow**, **Pressure.Vacuum**, **Carb.Pressure1**, **Temperature**

A limitation of the analysis is that its estimates are based on resampled, cross-validated data, rather than on testing on unseen data. A next step, pending approval from Management, is to test the models with unseen test data.

Data Overview & Transformation

The training dataset consists of 2,571 observations, 32 predictor variables and one predictor variable, *pH*, related to the composition of the beverages manufactured by ABC Beverage. Of the 32 predictor

variables, 31 predictors are continuous numerical and one predictor is a categorical beverage brand code denoting individual beverage products. The dataset is summarized in the table below.

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
1	Brand.Code [character]	1. (Empty string) 2. A 3. B 4. C 5. D	120 (4.7%) 293 (11.4%) 1239 (48.2%) 304 (11.8%) 615 (23.9%)	2571 (100.0%)	0 (0.0%)
2	Carb.Volume [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5 ≤ 5.3 ≤ 5.7 IQR (CV) : 0.2 (0)	101 distinct values	2561 (99.6%)	10 (0.4%)
3	Fill.Ounces [numeric]	Mean (sd) : 24 (0.1) min ≤ med ≤ max: 23.6 ≤ 24 ≤ 24.3 IQR (CV) : 0.1 (0)	92 distinct values	2533 (98.5%)	38 (1.5%)
4	PC.Volume [numeric]	Mean (sd) : 0.3 (0.1) min ≤ med ≤ max: 0.1 ≤ 0.3 ≤ 0.5 IQR (CV) : 0.1 (0.2)	454 distinct values	2532 (98.5%)	39 (1.5%)
5	Carb.Pressure [numeric]	Mean (sd) : 68.2 (3.5) min ≤ med ≤ max: 57 ≤ 68.2 ≤ 79.4 IQR (CV) : 5 (0.1)	106 distinct values	2544 (98.9%)	27 (1.1%)
6	Carb.Temp [numeric]	Mean (sd) : 141.1 (4) min ≤ med ≤ max: 128.6 ≤ 140.8 ≤ 154 IQR (CV) : 5.4 (0)	123 distinct values	2545 (99.0%)	26 (1.0%)
7	PSC [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0.1 ≤ 0.3 IQR (CV) : 0.1 (0.6)	129 distinct values	2538 (98.7%)	33 (1.3%)
8	PSC.Fill [numeric]	Mean (sd) : 0.2 (0.1) min ≤ med ≤ max: 0 ≤ 0.2 ≤ 0.6 IQR (CV) : 0.2 (0.6)	32 distinct values	2548 (99.1%)	23 (0.9%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
9	PSC.CO2 [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.2 IQR (CV) : 0.1 (0.8)	13 distinct values	2532 (98.5%)	39 (1.5%)
10	Mnf.Flow [numeric]	Mean (sd) : 24.6 (119.5) min ≤ med ≤ max: -100.2 ≤ 65.2 ≤ 229.4 IQR (CV) : 240.8 (4.9)	487 distinct values	2569 (99.9%)	2 (0.1%)
11	Carb.Pressure1 [numeric]	Mean (sd) : 122.6 (4.7) min ≤ med ≤ max: 105.6 ≤ 123.2 ≤ 140.2 IQR (CV) : 6.4 (0)	140 distinct values	2539 (98.8%)	32 (1.2%)
12	Fill.Pressure [numeric]	Mean (sd) : 47.9 (3.2) min ≤ med ≤ max: 34.6 ≤ 46.4 ≤ 60.4 IQR (CV) : 4 (0.1)	108 distinct values	2549 (99.1%)	22 (0.9%)
13	Hyd.Pressure1 [numeric]	Mean (sd) : 12.4 (12.4) min ≤ med ≤ max: -0.8 ≤ 11.4 ≤ 58 IQR (CV) : 20.2 (1)	245 distinct values	2560 (99.6%)	11 (0.4%)
14	Hyd.Pressure2 [numeric]	Mean (sd) : 21 (16.4) min ≤ med ≤ max: 0 ≤ 28.6 ≤ 59.4 IQR (CV) : 34.6 (0.8)	207 distinct values	2556 (99.4%)	15 (0.6%)
15	Hyd.Pressure3 [numeric]	Mean (sd) : 20.5 (16) min ≤ med ≤ max: -1.2 ≤ 27.6 ≤ 50 IQR (CV) : 33.4 (0.8)	192 distinct values	2556 (99.4%)	15 (0.6%)
16	Hyd.Pressure4 [integer]	Mean (sd) : 96.3 (13.1) min ≤ med ≤ max: 52 ≤ 96 ≤ 142 IQR (CV) : 16 (0.1)	40 distinct values	2541 (98.8%)	30 (1.2%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
17	Filler.Level [numeric]	Mean (sd) : 109.3 (15.7) min ≤ med ≤ max: 55.8 ≤ 118.4 ≤ 161.2 IQR (CV) : 21.7 (0.1)	288 distinct values	2551 (99.2%)	20 (0.8%)
18	Filler.Speed [integer]	Mean (sd) : 3687.2 (770.8) min ≤ med ≤ max: 998 ≤ 3982 ≤ 4030 IQR (CV) : 110 (0.2)	244 distinct values	2514 (97.8%)	57 (2.2%)
19	Temperature [numeric]	Mean (sd) : 66 (1.4) min ≤ med ≤ max: 63.6 ≤ 65.6 ≤ 76.2 IQR (CV) : 1.2 (0)	56 distinct values	2557 (99.5%)	14 (0.5%)
20	Usage.cont [numeric]	Mean (sd) : 21 (3) min ≤ med ≤ max: 12.1 ≤ 21.8 ≤ 25.9 IQR (CV) : 5.4 (0.1)	481 distinct values	2566 (99.8%)	5 (0.2%)
21	Carb.Flow [integer]	Mean (sd) : 2468.4 (1073.7) min ≤ med ≤ max: 26 ≤ 3028 ≤ 5104 IQR (CV) : 2042 (0.4)	533 distinct values	2569 (99.9%)	2 (0.1%)
22	Density [numeric]	Mean (sd) : 1.2 (0.4) min ≤ med ≤ max: 0.2 ≤ 1 ≤ 1.9 IQR (CV) : 0.7 (0.3)	78 distinct values	2570 (100.0%)	1 (0.0%)
23	MFR [numeric]	Mean (sd) : 704 (73.9) min ≤ med ≤ max: 31.4 ≤ 724 ≤ 868.6 IQR (CV) : 24.7 (0.1)	587 distinct values	2359 (91.8%)	212 (8.2%)
24	Balling [numeric]	Mean (sd) : 2.2 (0.9) min ≤ med ≤ max: -0.2 ≤ 1.6 ≤ 4 IQR (CV) : 1.8 (0.4)	217 distinct values	2570 (100.0%)	1 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
25	Pressure.Vacuum [numeric]	Mean (sd) : -5.2 (0.6) min ≤ med ≤ max: -6.6 ≤ -5.4 ≤ -3.6 IQR (CV) : 0.6 (-0.1)	16 distinct values	2571 (100.0%)	0 (0.0%)
26	PH [numeric]	Mean (sd) : 8.5 (0.2) min ≤ med ≤ max: 7.9 ≤ 8.5 ≤ 9.4 IQR (CV) : 0.2 (0)	52 distinct values	2567 (99.8%)	4 (0.2%)
27	Oxygen.Filler [numeric]	Mean (sd) : 0 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.4 IQR (CV) : 0 (1)	338 distinct values	2559 (99.5%)	12 (0.5%)
28	Bowl.Setpoint [integer]	Mean (sd) : 109.3 (15.3) min ≤ med ≤ max: 70 ≤ 120 ≤ 140 IQR (CV) : 20 (0.1)	11 distinct values	2569 (99.9%)	2 (0.1%)
29	Pressure.Setpoint [numeric]	Mean (sd) : 47.6 (2) min ≤ med ≤ max: 44 ≤ 46 ≤ 52 IQR (CV) : 4 (0)	8 distinct values	2559 (99.5%)	12 (0.5%)
30	Air.Pressurer [numeric]	Mean (sd) : 142.8 (1.2) min ≤ med ≤ max: 140.8 ≤ 142.6 ≤ 148.2 IQR (CV) : 0.8 (0)	32 distinct values	2571 (100.0%)	0 (0.0%)
31	Alch.Rel [numeric]	Mean (sd) : 6.9 (0.5) min ≤ med ≤ max: 5.3 ≤ 6.6 ≤ 8.6 IQR (CV) : 0.7 (0.1)	53 distinct values	2562 (99.6%)	9 (0.4%)
32	Carb.Rel [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5 ≤ 5.4 ≤ 6.1 IQR (CV) : 0.2 (0)	42 distinct values	2561 (99.6%)	10 (0.4%)
33	Balling.Lvl [numeric]	Mean (sd) : 2.1 (0.9) min ≤ med ≤ max:	82 distinct values	2570 (100.0%)	1 (0.0%)

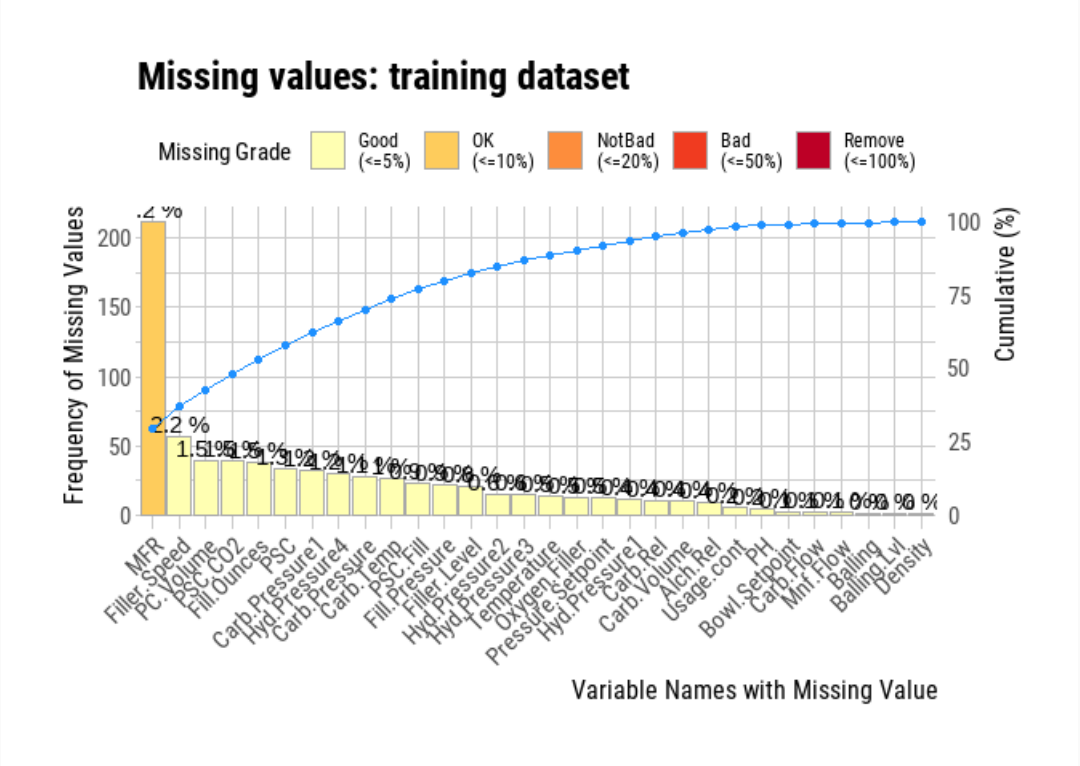
No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
		0 ≤ 1.5 ≤ 3.7 IQR (CV) : 1.8 (0.4)			

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

/// Missing Values

Missing values are present in 30 variables, including the target variable *pH*, ranging in prevalence from 0.001% to 8.2% of observations. In addition, there are 120 observations for which the grouping variable *Brand.Code* is empty, denoting no beverage product. Absent explicit guidance from management on how to treat these observations with a missing brand code, we do not consider these 120 observations further.

Therefore, to correct for missing values we perform the following transformations on the training dataset:
1. Drop 120 observations for which the beverage **Brand.Code** variable is empty.
2. Drop four (4) observations for which the target variable **PH** is empty.
3. Impute missing values with the Bagging Trees method.



variables	types	missing_count	missing_percent
MFR	numeric	212	8.2
Filler.Speed	integer	57	2.2

variables	types	missing_count	missing_percent
PC.Volume	numeric	39	1.5
PSC.CO2	numeric	39	1.5
Fill.Ounces	numeric	38	1.5
PSC	numeric	33	1.3
Carb.Pressure1	numeric	32	1.2
Hyd.Pressure4	integer	30	1.2
Carb.Pressure	numeric	27	1.1
Carb.Temp	numeric	26	1.0
PSC.Fill	numeric	23	0.9
Fill.Pressure	numeric	22	0.9
Filler.Level	numeric	20	0.8
Hyd.Pressure2	numeric	15	0.6
Hyd.Pressure3	numeric	15	0.6
Temperature	numeric	14	0.5
Oxygen.Filler	numeric	12	0.5
Pressure.Setpoint	numeric	12	0.5
Hyd.Pressure1	numeric	11	0.4
Carb.Volume	numeric	10	0.4
Carb.Rel	numeric	10	0.4
Alch.Rel	numeric	9	0.4
Usage.cont	numeric	5	0.2
PH	numeric	4	0.2
Mnf.Flow	numeric	2	0.1
Carb.Flow	integer	2	0.1
Bowl.Setpoint	integer	2	0.1

variables	types	missing_count	missing_percent
Density	numeric	1	0.0
Balling	numeric	1	0.0
Balling.Lvl	numeric	1	0.0

/// Variance

A check of variable variance reveals no predictor variable with zero or near zero variance.

	freqRatio	percentUnique	zeroVar	nzv
Brand.Code	2.0	0.2	FALSE	FALSE
Carb.Volume	1.1	4.3	FALSE	FALSE
Fill.Ounces	1.2	5.0	FALSE	FALSE
PC.Volume	1.1	19.5	FALSE	FALSE
Carb.Pressure	1.0	5.0	FALSE	FALSE
Carb.Temp	1.0	5.4	FALSE	FALSE
PSC	1.2	6.4	FALSE	FALSE
PSC.Fill	1.1	2.1	FALSE	FALSE
PSC.CO2	1.1	2.0	FALSE	FALSE
Mnf.Flow	1.1	19.6	FALSE	FALSE
Carb.Pressure1	1.0	6.3	FALSE	FALSE
Fill.Pressure	1.8	4.9	FALSE	FALSE
Hyd.Pressure1	30.7	10.1	FALSE	FALSE
Hyd.Pressure2	7.6	8.4	FALSE	FALSE
Hyd.Pressure3	12.1	7.8	FALSE	FALSE
Hyd.Pressure4	1.0	1.8	FALSE	FALSE
Filler.Level	1.1	11.7	FALSE	FALSE
Filler.Speed	1.1	11.1	FALSE	FALSE

	freqRatio	percentUnique	zeroVar	nzv
Temperature	1.1	2.7	FALSE	FALSE
Usage.cont	1.1	19.6	FALSE	FALSE
Carb.Flow	1.4	21.4	FALSE	FALSE
Density	1.1	3.1	FALSE	FALSE
MFR	1.1	24.2	FALSE	FALSE
Balling	1.2	8.4	FALSE	FALSE
Pressure.Vacuum	1.4	0.7	FALSE	FALSE
PH	1.1	2.1	FALSE	FALSE
Oxygen.Filler	1.3	13.8	FALSE	FALSE
Bowl.Setpoint	3.0	0.5	FALSE	FALSE
Pressure.Setpoint	1.3	0.6	FALSE	FALSE
Air.Pressurer	1.1	1.3	FALSE	FALSE
Alch.Rel	1.2	2.1	FALSE	FALSE
Carb.Rel	1.0	1.9	FALSE	FALSE
Balling.Lvl	1.3	3.3	FALSE	FALSE

/// Duplicates

A check for duplicate observations reveals that the dataset is free of duplicates.

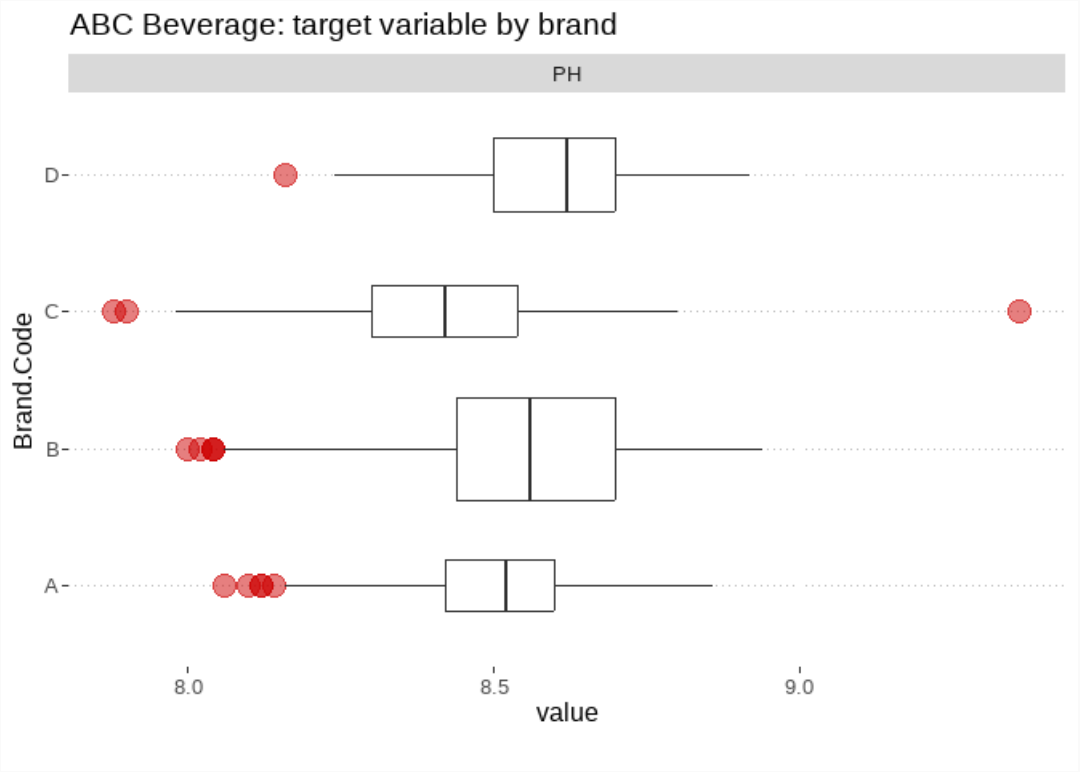
```
### Check for duplicate observations ###
sum(duplicated(StudentData))
```

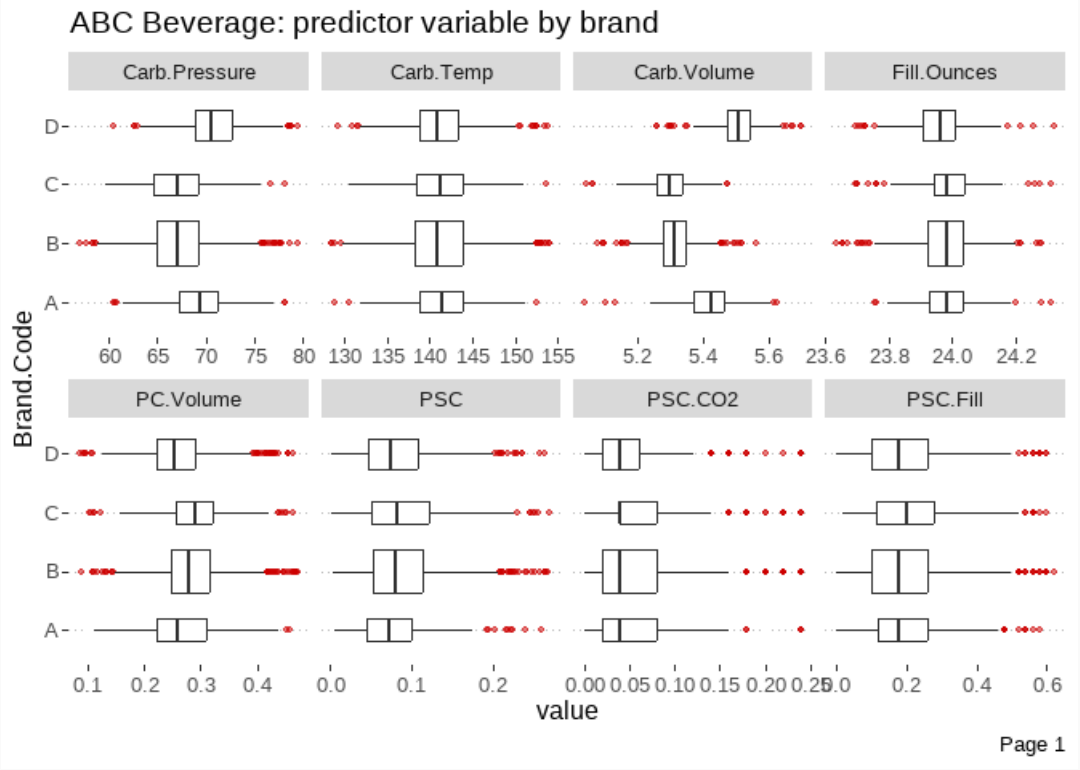
```
## [1] 0
```

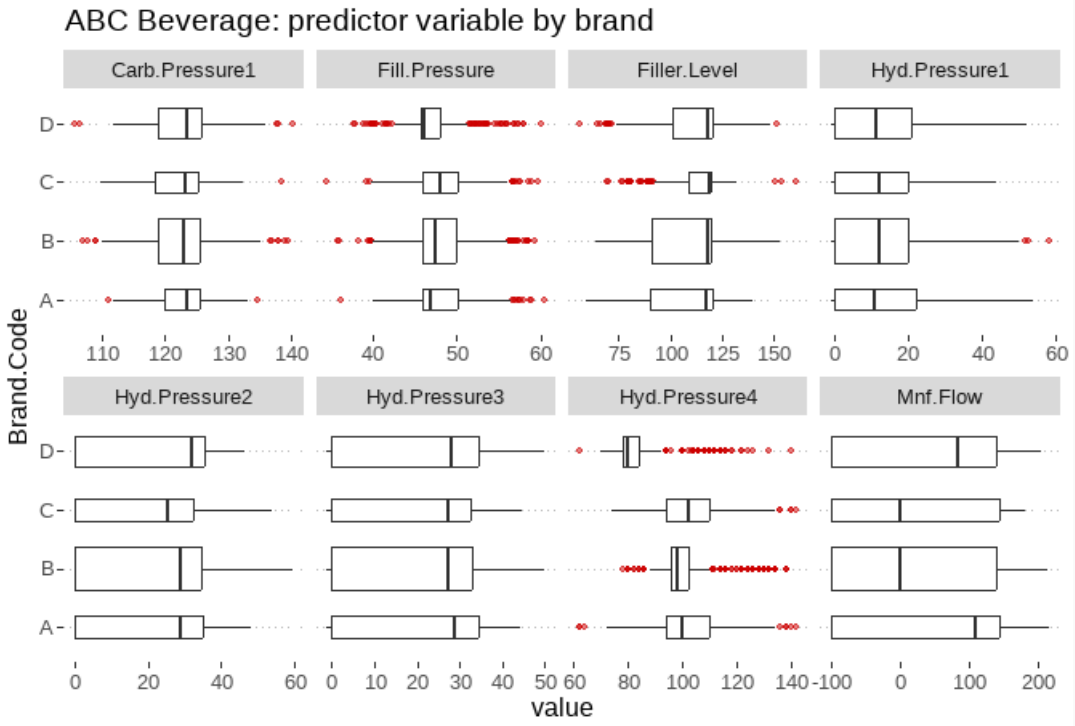
/// Outliers

Box plots of target and predictor variables grouped by beverage brand reveal outliers (shown in red) across predictors.

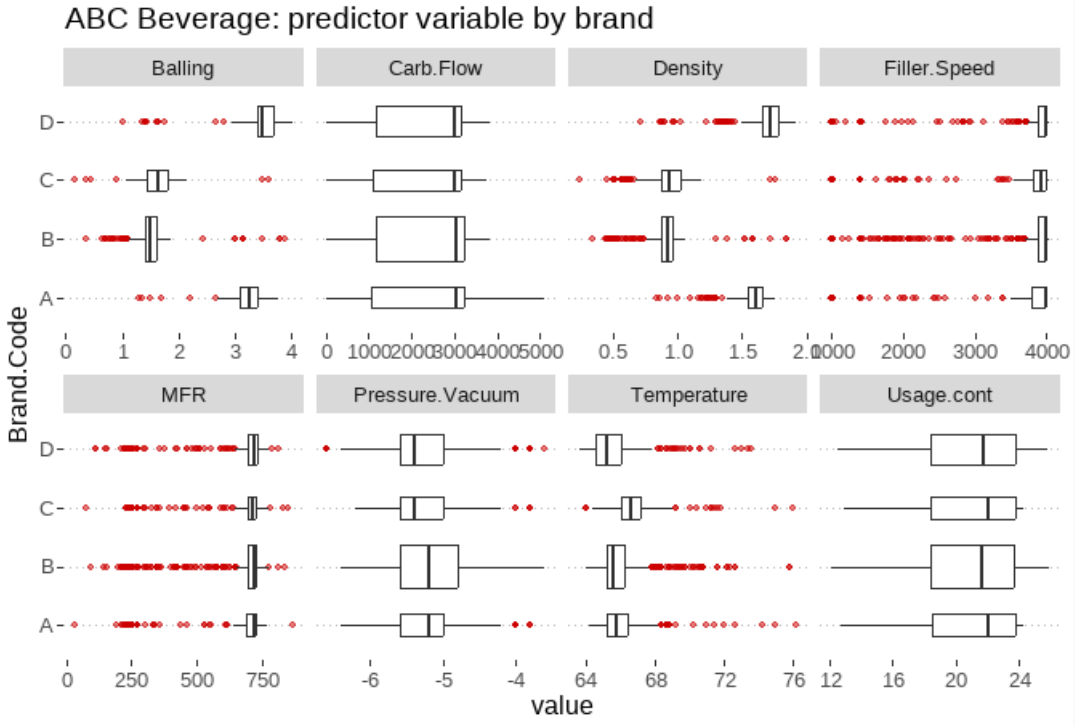
Nevertheless, because the task is at hand is predictive rather than inferential modeling, we leave outliers intact lest we risk model over-fitting and diminished predictive performance later.



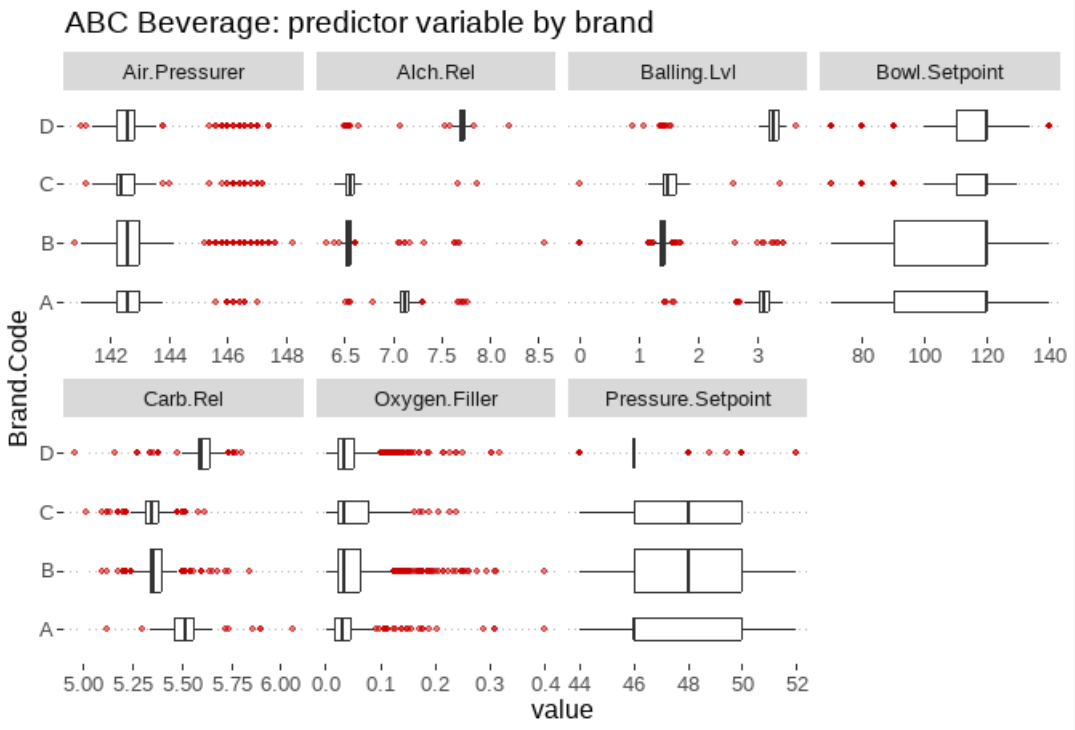



ABC Beverage: predictor variable by brand
Carb.Pressure1
Fill.Pressure
Filler.Level
Hyd.Pressure1
Hyd.Pressure2
Hyd.Pressure3
Hyd.Pressure4
Mnf.Flow
Brand.Code
D
C
B
A
value

Page 2


ABC Beverage: predictor variable by brand
Balling
Carb.Flow
Density
Filler.Speed
MFR
Pressure.Vacuum
Temperature
Usage.cont
Brand.Code
D
C
B
A
value

Page 3



Variable	n	% n	Outlier mean	Variable mean w/ outliers	Variable mean w/o outliers
Filler.Speed	362	14.8	2102.4	3675.0	3948.1
MFR	299	12.2	354.9	673.6	718.0
Air.Pressurer	220	9.0	146.4	142.8	142.5
Oxygen.Filler	178	7.3	0.2	0.0	0.0
Pressure.Vacuum	116	4.7	-4.0	-5.2	-5.3
Hyd.Pressure4	111	4.5	127.8	96.4	95.0
Temperature	104	4.3	70.3	65.9	65.7
PC.Volume	79	3.2	0.3	0.3	0.3
Fill.Pressure	72	2.9	51.7	47.9	47.8
PSC.CO2	69	2.8	0.2	0.1	0.1
PSC	54	2.2	0.2	0.1	0.1
PSC.Fill	52	2.1	0.6	0.2	0.2
Fill.Ounces	45	1.8	23.9	24.0	24.0

Variable	n	% n	Outlier mean	Variable mean w/ outliers	Variable mean w/o outliers
Carb.Temp	34	1.4	147.2	141.1	141.1
Carb.Pressure	21	0.9	74.4	68.3	68.2
PH	18	0.7	8.1	8.5	8.6
Carb.Pressure1	17	0.7	127.2	122.5	122.5
Hyd.Pressure1	6	0.2	53.3	12.6	12.5
Filler.Level	4	0.2	84.2	109.1	109.1
Carb.Rel	4	0.2	5.7	5.4	5.4
Carb.Volume	1	0.0	5.0	5.4	5.4
Mnf.Flow	0	0.0	NaN	24.5	24.5
Hyd.Pressure2	0	0.0	NaN	21.2	21.2
Hyd.Pressure3	0	0.0	NaN	20.5	20.5
Usage.cont	0	0.0	NaN	21.0	21.0
Carb.Flow	0	0.0	NaN	2468.5	2468.5
Density	0	0.0	NaN	1.2	1.2
Balling	0	0.0	NaN	2.2	2.2
Bowl.Setpoint	0	0.0	NaN	109.2	109.2
Pressure.Setpoint	0	0.0	NaN	47.6	47.6
Alch.Rel	0	0.0	NaN	6.9	6.9
Balling.Lvl	0	0.0	NaN	2.1	2.1

/// Collinearity

The dataset has eight predictor variable pairs with correlation ranging from moderate to very high, pointing some collinearity in the data.

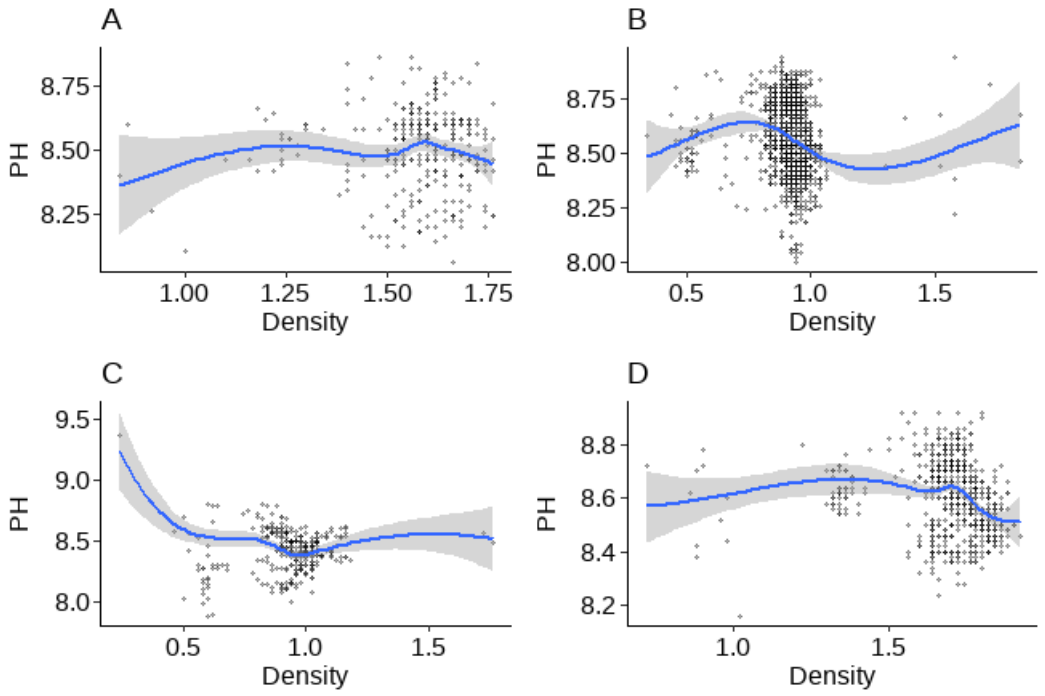
We do not effect any transformations to de-correlate predictor variables due to the uncertain, premature nature of this task at this point and to the relatively high number of predictors involved.

Predictor 1	Predictor 2	Correlation	p-value
Carb.Pressure	Carb.Temp	0.91	0
Filler.Level	Bowl.Setpoint	0.87	0
Filler.Speed	MFR	0.84	0
Hyd.Pressure2	Hyd.Pressure3	0.78	0
Hyd.Pressure1	Hyd.Pressure2	0.68	0
Density	Balling	0.68	0
Mnf.Flow	Hyd.Pressure3	0.66	0
Alch.Rel	Balling.Lvl	-0.78	0

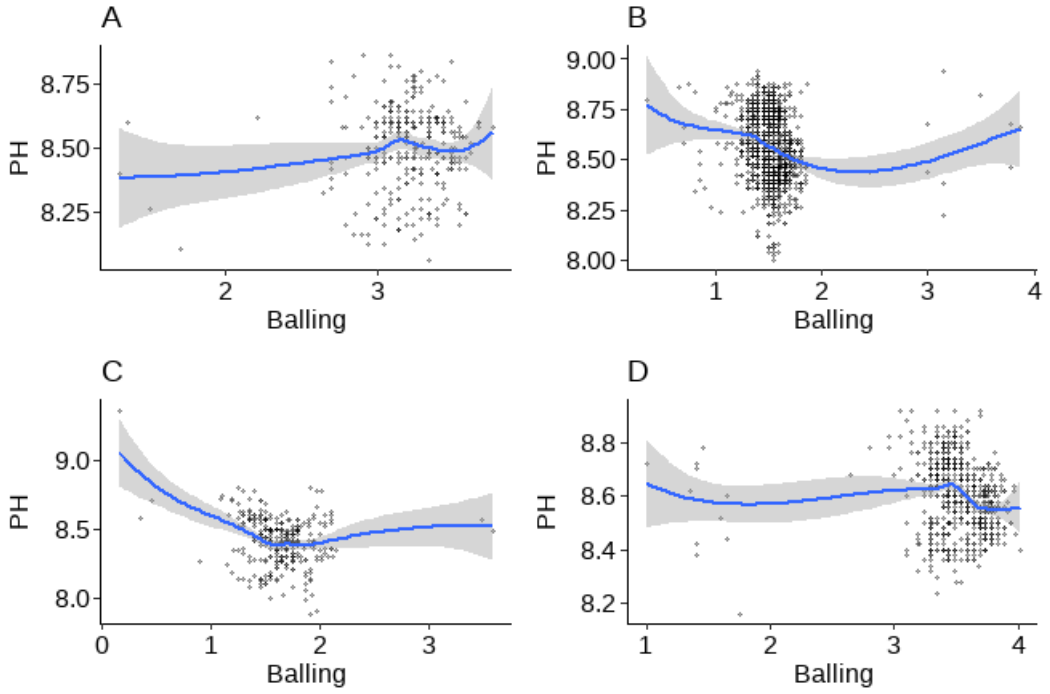
/// Non-linearity

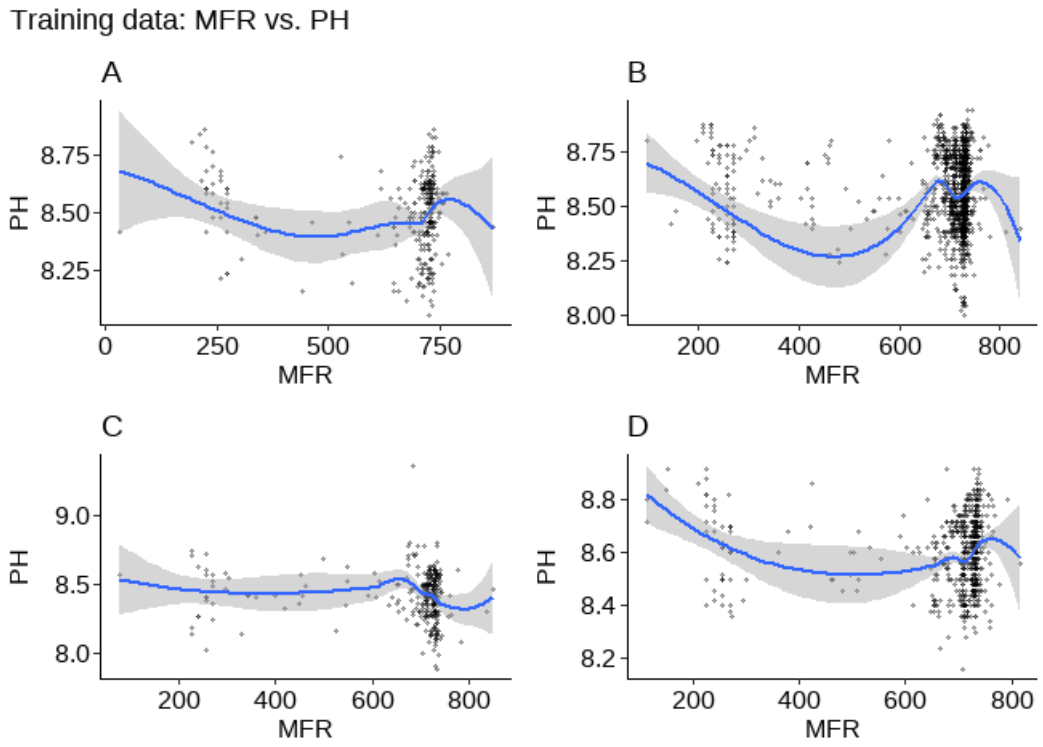
The dataset shows nonlinear relationships in varying degrees between the predictor and target variable for most predictors. The graphs below for five selected predictor vs. target variable *PH* pairs illustrate this phenomenon.

Training data: Density vs. PH



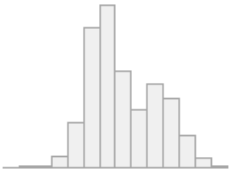
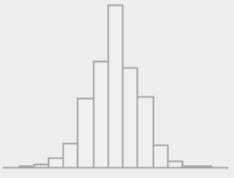
Training data: Balling vs. PH

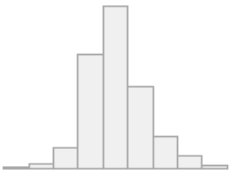
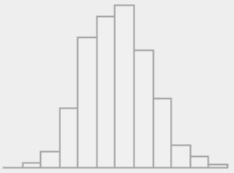
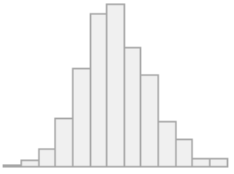
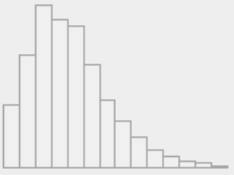
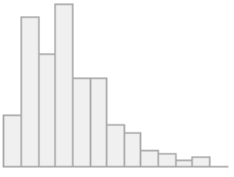


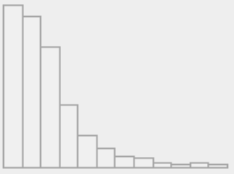
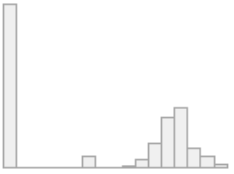
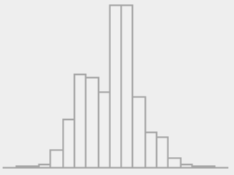
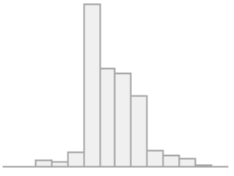
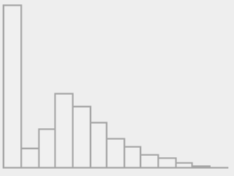


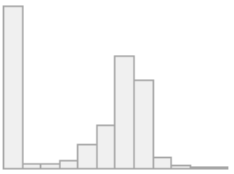
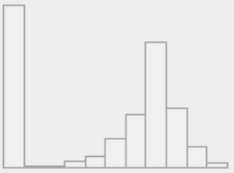
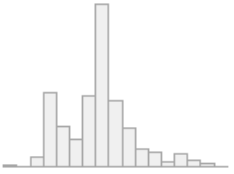
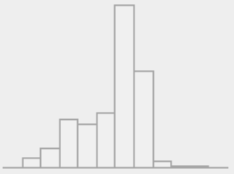
In summary, the training dataset is characterized by 1.) a fair number of outliers and by 2.) collinearity across several predictor variables, and 3.) non-linearity between the predictors and the target variable.

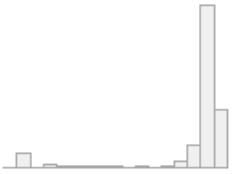
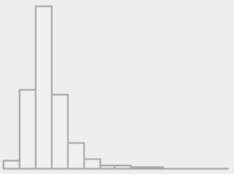
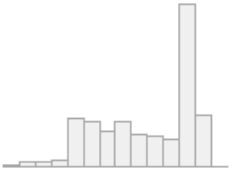
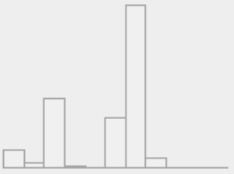
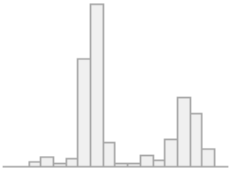
The final training data after the foregoing transformations, now ready for next step of model fitting, is summarized below.

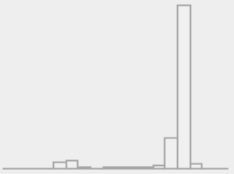
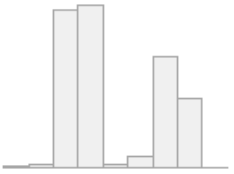
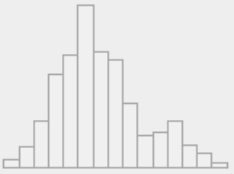
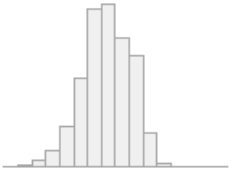
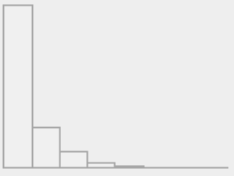
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Brand.Code [character]	1. A 2. B 3. C 4. D	293 (12.0%) 1235 (50.5%) 304 (12.4%) 615 (25.1%)		2447 (100.0%)	0 (0.0%)
2	Carb.Volume [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5 ≤ 5.3 ≤ 5.7 IQR (CV) : 0.2 (0)	104 distinct values		2447 (100.0%)	0 (0.0%)
3	Fill.Ounces [numeric]	Mean (sd) : 24 (0.1) min ≤ med ≤ max:	122 distinct values		2447 (100.0%)	0 (0.0%)

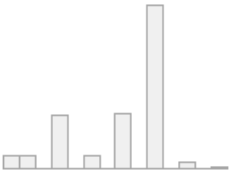
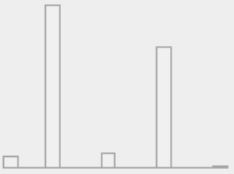
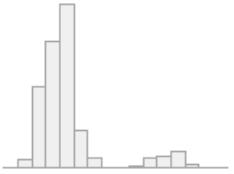
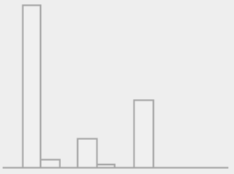
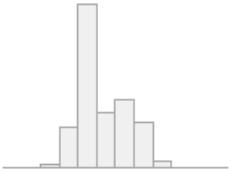
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		23.6 ≤ 24 ≤ 24.3 IQR (CV) : 0.1 (0)				
4	PC.Volume [numeric]	Mean (sd) : 0.3 (0.1) min ≤ med ≤ max: 0.1 ≤ 0.3 ≤ 0.5 IQR (CV) : 0.1 (0.2)	476 distinct values		2447 (100.0%)	0 (0.0%)
5	Carb.Pressure [numeric]	Mean (sd) : 68.3 (3.5) min ≤ med ≤ max: 57 ≤ 68.2 ≤ 79.4 IQR (CV) : 4.8 (0.1)	123 distinct values		2447 (100.0%)	0 (0.0%)
6	Carb.Temp [numeric]	Mean (sd) : 141.1 (4.1) min ≤ med ≤ max: 128.6 ≤ 140.8 ≤ 154 IQR (CV) : 5.4 (0)	133 distinct values		2447 (100.0%)	0 (0.0%)
7	PSC [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0.1 ≤ 0.3 IQR (CV) : 0.1 (0.6)	157 distinct values		2447 (100.0%)	0 (0.0%)
8	PSC.Fill [numeric]	Mean (sd) : 0.2 (0.1) min ≤ med ≤ max:	51 distinct values		2447 (100.0%)	0 (0.0%)

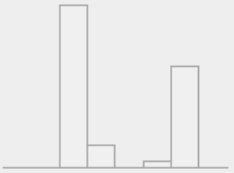
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		0 ≤ 0.2 ≤ 0.6 IQR (CV) : 0.2 (0.6)				
9	PSC.CO2 [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.2 IQR (CV) : 0.1 (0.7)	49 distinct values		2447 (100.0%)	0 (0.0%)
10	Mnf.Flow [numeric]	Mean (sd) : 24.5 (119.7) min ≤ med ≤ max: -100.2 ≤ 70.2 ≤ 216.2 IQR (CV) : 241.2 (4.9)	479 distinct values		2447 (100.0%)	0 (0.0%)
11	Carb.Pressure1 [numeric]	Mean (sd) : 122.5 (4.7) min ≤ med ≤ max: 105.6 ≤ 123.2 ≤ 140.2 IQR (CV) : 6.6 (0)	154 distinct values		2447 (100.0%)	0 (0.0%)
12	Fill.Pressure [numeric]	Mean (sd) : 47.9 (3.1) min ≤ med ≤ max: 34.6 ≤ 46.4 ≤ 60.4 IQR (CV) : 4 (0.1)	120 distinct values		2447 (100.0%)	0 (0.0%)
13	Hyd.Pressure1 [numeric]	Mean (sd) : 12.6 (12.4) min ≤ med ≤ max:	246 distinct values		2447 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		-0.8 ≤ 11.6 ≤ 58 IQR (CV) : 20.4 (1)				
14	Hyd.Pressure2 [numeric]	Mean (sd) : 21.2 (16.4) min ≤ med ≤ max: 0 ≤ 28.8 ≤ 59.4 IQR (CV) : 34.8 (0.8)	206 distinct values		2447 (100.0%)	0 (0.0%)
15	Hyd.Pressure3 [numeric]	Mean (sd) : 20.5 (15.9) min ≤ med ≤ max: -1.2 ≤ 27.8 ≤ 50 IQR (CV) : 33.2 (0.8)	190 distinct values		2447 (100.0%)	0 (0.0%)
16	Hyd.Pressure4 [numeric]	Mean (sd) : 96.4 (12.9) min ≤ med ≤ max: 62 ≤ 96 ≤ 142 IQR (CV) : 14 (0.1)	45 distinct values		2447 (100.0%)	0 (0.0%)
17	Filler.Level [numeric]	Mean (sd) : 109.1 (15.7) min ≤ med ≤ max: 55.8 ≤ 118.4 ≤ 161.2 IQR (CV) : 22.7 (0.1)	286 distinct values		2447 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
18	Filler.Speed [numeric]	Mean (sd) : 3675 (777.8) min ≤ med ≤ max: 998 ≤ 3982 ≤ 4030 IQR (CV) : 128 (0.2)	272 distinct values		2447 (100.0%)	0 (0.0%)
19	Temperature [numeric]	Mean (sd) : 65.9 (1.3) min ≤ med ≤ max: 63.6 ≤ 65.6 ≤ 76.2 IQR (CV) : 1.2 (0)	65 distinct values		2447 (100.0%)	0 (0.0%)
20	Usage.cont [numeric]	Mean (sd) : 21 (3) min ≤ med ≤ max: 12.1 ≤ 21.8 ≤ 25.9 IQR (CV) : 5.4 (0.1)	479 distinct values		2447 (100.0%)	0 (0.0%)
21	Carb.Flow [numeric]	Mean (sd) : 2468.5 (1070) min ≤ med ≤ max: 26 ≤ 3030 ≤ 5104 IQR (CV) : 2024 (0.4)	524 distinct values		2447 (100.0%)	0 (0.0%)
22	Density [numeric]	Mean (sd) : 1.2 (0.4) min ≤ med ≤ max: 0.2 ≤ 1 ≤ 1.9	77 distinct values		2447 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		IQR (CV) : 0.7 (0.3)				
23	MFR [numeric]	Mean (sd) : 673.6 (133) min ≤ med ≤ max: 31.4 ≤ 722.2 ≤ 868.6 IQR (CV) : 33.5 (0.2)	591 distinct values		2447 (100.0%)	0 (0.0%)
24	Balling [numeric]	Mean (sd) : 2.2 (0.9) min ≤ med ≤ max: 0.2 ≤ 1.6 ≤ 4 IQR (CV) : 1.8 (0.4)	205 distinct values		2447 (100.0%)	0 (0.0%)
25	Pressure.Vacuum [numeric]	Mean (sd) : -5.2 (0.6) min ≤ med ≤ max: -6.6 ≤ -5.4 ≤ -3.6 IQR (CV) : 0.6 (-0.1)	16 distinct values		2447 (100.0%)	0 (0.0%)
26	PH [numeric]	Mean (sd) : 8.5 (0.2) min ≤ med ≤ max: 7.9 ≤ 8.5 ≤ 9.4 IQR (CV) : 0.2 (0)	52 distinct values		2447 (100.0%)	0 (0.0%)
27	Oxygen.Filler [numeric]	Mean (sd) : 0 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.4	338 distinct values		2447 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		IQR (CV) : 0 (1)				
28	Bowl.Setpoint [numeric]	Mean (sd) : 109.2 (15.3) min ≤ med ≤ max: 70 ≤ 120 ≤ 140 IQR (CV) : 20 (0.1)	12 distinct values		2447 (100.0%)	0 (0.0%)
29	Pressure.Setpoint [numeric]	Mean (sd) : 47.6 (2) min ≤ med ≤ max: 44 ≤ 46 ≤ 52 IQR (CV) : 4 (0)	14 distinct values		2447 (100.0%)	0 (0.0%)
30	Air.Pressurer [numeric]	Mean (sd) : 142.8 (1.2) min ≤ med ≤ max: 140.8 ≤ 142.6 ≤ 148.2 IQR (CV) : 0.8 (0)	32 distinct values		2447 (100.0%)	0 (0.0%)
31	Alch.Rel [numeric]	Mean (sd) : 6.9 (0.5) min ≤ med ≤ max: 6.3 ≤ 6.6 ≤ 8.6 IQR (CV) : 1.1 (0.1)	51 distinct values		2447 (100.0%)	0 (0.0%)
32	Carb.Rel [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5 ≤ 5.4 ≤ 6.1	46 distinct values		2447 (100.0%)	0 (0.0%)

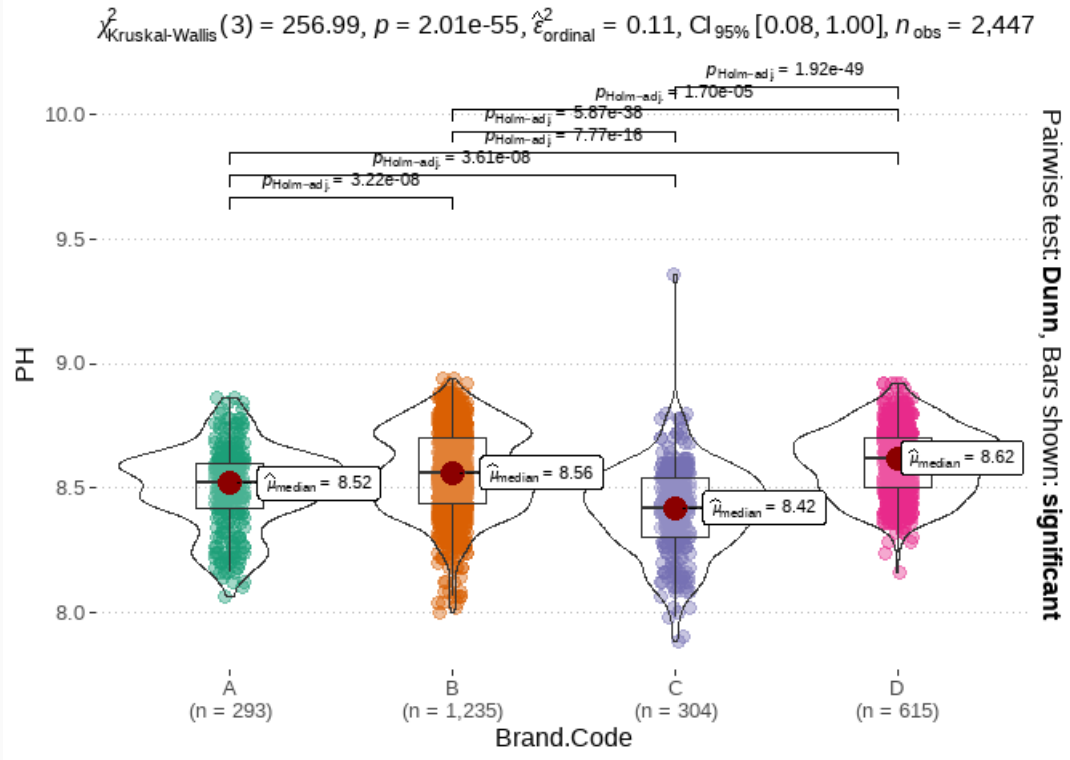
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		IQR (CV) : 0.2 (0)				
33	Balling.Lvl [numeric]	Mean (sd) : 2.1 (0.9) min ≤ med ≤ max: 0 ≤ 1.5 ≤ 3.7 IQR (CV) : 1.8 (0.4)	80 distinct values		2447 (100.0%)	0 (0.0%)

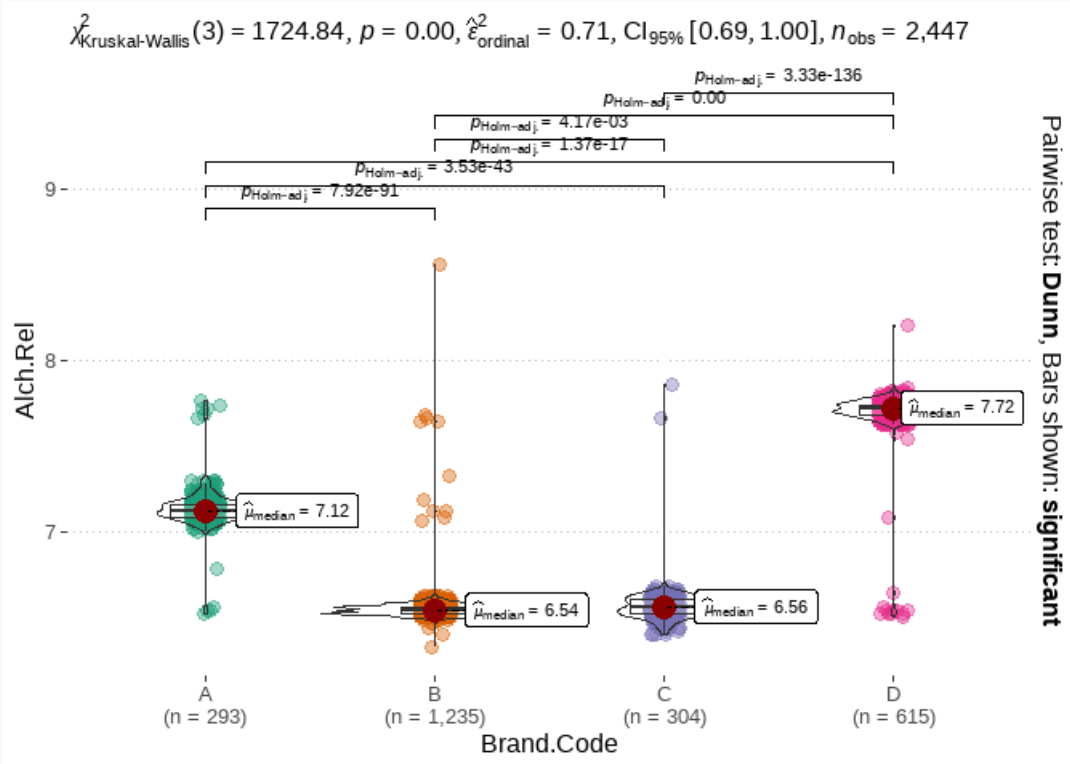
Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

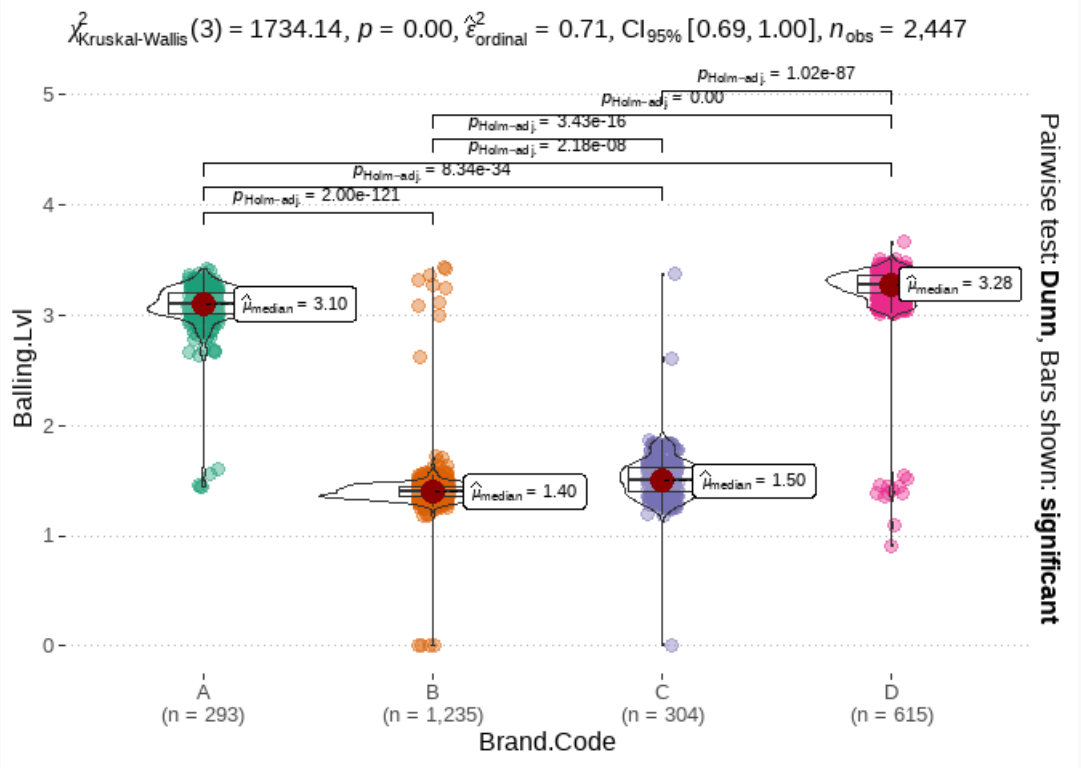
/ Model Fitting & Selection

Target variable **PH** displays statistically significant difference across the four A, B, C, and D beverage brands. This is demonstrated in the following graph showing pair-wise comparisons among the brands A, B, C and D for the **PH** target variable. First, a p-value of 2.01e-55 for the Kruskal-Wallis test (upper left in the plot) provides little evidence for the test null hypothesis that no one group median differs from the other, pointing to at least one group median differing from the others. Second the 0.11 value of the ϵ^2 statistic (upper center of the plot) points to the *moderate* impact of brand on *PH*. Finally, the p-values for the Dunn Test comparing each group against all other groups pair-wisely are below the 0.05 threshold, pointing to statistically significant group medians across all groups. Most predictor variables also show statistically significant difference across brand groups, often to a greater degree than the target variable. Three selected predictor variable plots illustrate this point. Note the very high ϵ^2 values.

Consequently, we chose the model the target variable **PH** individually by brand A, B, C and D, rather than to model it collectively for all four brands together. Likewise, we model each brand individually so we can achieve a more granular, detailed insight into drivers of pH for each beverage and better understand the manufacturing process.


$\chi^2_{\text{Kruskal-Wallis}}(3) = 256.99$, $p = 2.01e\text{-}55$, $\hat{\varepsilon}^2_{\text{ordinal}} = 0.11$, $\text{CI}_{95\%}[0.08, 1.00]$, $n_{\text{obs}} = 2{,}447$
$p_{\text{Holm-adj.}} = 1.92e\text{-}49$
$p_{\text{Holm-adj.}} = 1.70e\text{-}05$
$p_{\text{Holm-adj.}} = 5.87e\text{-}38$
$p_{\text{Holm-adj.}} = 7.77e\text{-}16$
$p_{\text{Holm-adj.}} = 3.61e\text{-}08$
$p_{\text{Holm-adj.}} = 3.22e\text{-}08$
PH
10.0
9.5
9.0
8.5
8.0
$\hat{\mu}_{\text{median}} = 8.52$
$\hat{\mu}_{\text{median}} = 8.56$
$\hat{\mu}_{\text{median}} = 8.42$
$\hat{\mu}_{\text{median}} = 8.62$
A
(n = 293)
B
(n = 1,235)
C
(n = 304)
D
(n = 615)
Brand.Code
Pairwise test: Dunn, Bars shown: significant


$\chi^2_{\text{Kruskal-Wallis}}(3) = 1724.84$, $p = 0.00$, $\hat{\varepsilon}^2_{\text{ordinal}} = 0.71$, $\text{CI}_{95\%}[0.69, 1.00]$, $n_{\text{obs}} = 2{,}447$
$p_{\text{Holm-adj.}} = 3.33e\text{-}136$
$p_{\text{Holm-adj.}} = 0.00$
$p_{\text{Holm-adj.}} = 4.17e\text{-}03$
$p_{\text{Holm-adj.}} = 1.37e\text{-}17$
$p_{\text{Holm-adj.}} = 3.53e\text{-}43$
$p_{\text{Holm-adj.}} = 7.92e\text{-}91$
Alch.Rel
9
8
7
$\hat{\mu}_{\text{median}} = 7.12$
$\hat{\mu}_{\text{median}} = 6.54$
$\hat{\mu}_{\text{median}} = 6.56$
$\hat{\mu}_{\text{median}} = 7.72$
A
(n = 293)
B
(n = 1,235)
C
(n = 304)
D
(n = 615)
Brand.Code
Pairwise test: Dunn, Bars shown: significant



Due to the non-linearity, collinearity and outlier contamination in the data, we avoid fitting simple linear models like Ordinary Least Squares regression. Instead, we fit five more complex predictive models. The first, our baseline linear model, is the linear ElasticNet regression model. We expect the L1 and L2 regularization of ElasticNet to account for collinearity and outliers, but we expect it to lag in performance due to its linear natures.

We fit also two non-linear regression models, Multivariate Adaptive Regression Splines (MARS) and Support Vector Machines (SVM). Finally, we fit two non-linear tree-based models, Gradient Boosted Machines (GBM) and Random Forest (RF). Together, we fit the five models to each one of the four beverage brands for a total of twenty different models.

Model resampled performance is determined by lowest estimate Root Mean Square Error (RMSE) metric via 10-fold cross-validation repeated three times for all models except for Gradient Boosted Machines. GBM performance is determined instead with bootstrap resampling with 25 repetitions. Each model is fitted with grid search hyper-parameter tuning in search of the model with lowest resampled RMSE value. For each brand of the four brands, the model with the lowest resampled RMSE is selected for prediction from among its four competitors.

Due to the short time alloted for the analysis, testing on unseen data was not attempted in this analysis. This is the next logical step in the analysis. Final model selection should await and depend on this testing.

The analysis was conducted in the **R** programming language with the RStudio IDE version **2022.12.0+353 for Windows Mozilla/5.0**. The following libraries were used in data processing and modeling:

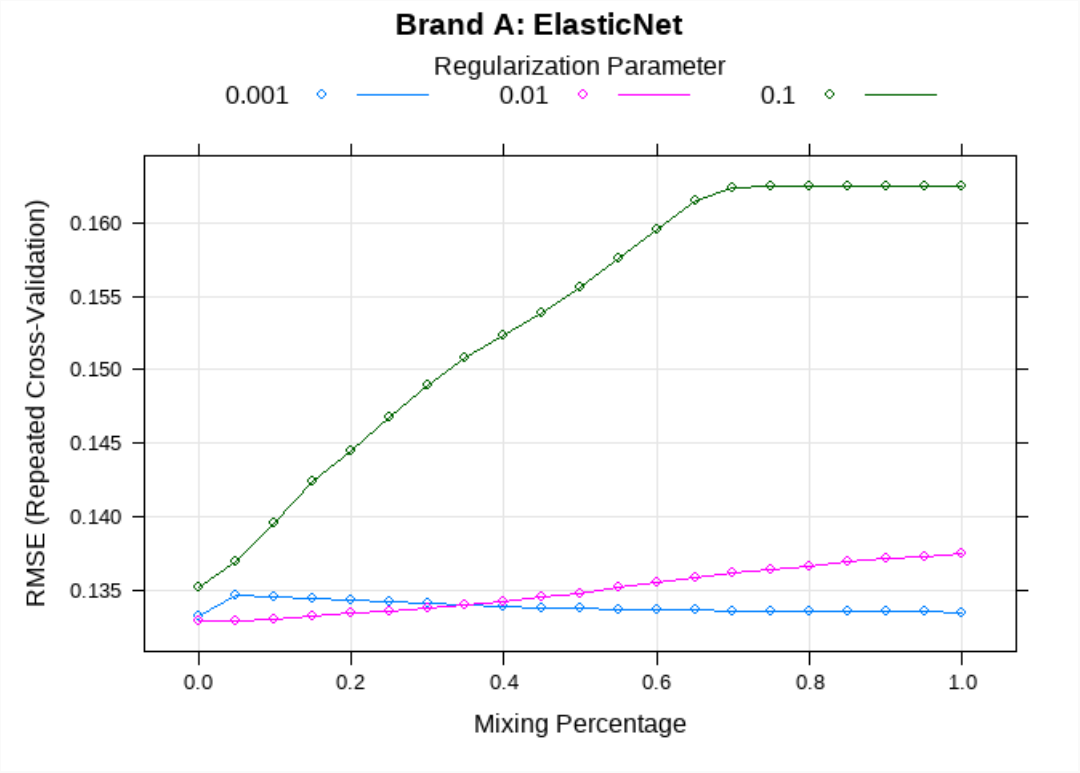
- Data processing and resampled model fit: `caret`, v6.0-94
- Elastic Net: `glmnet`, v4.1-3
- Multivariate Adaptive Regression Splines: `earth`, v5.3.2

- Support Vector Machines: `e1071`, v1.7-13
- Random Forest: `randomForest`, v4.7-1.1
- GBM: `gbm`, v2.1.8.1

All work is conducted with `set.seed(2023)`.

// Brand A: Model Fitting

/// ElasticNet

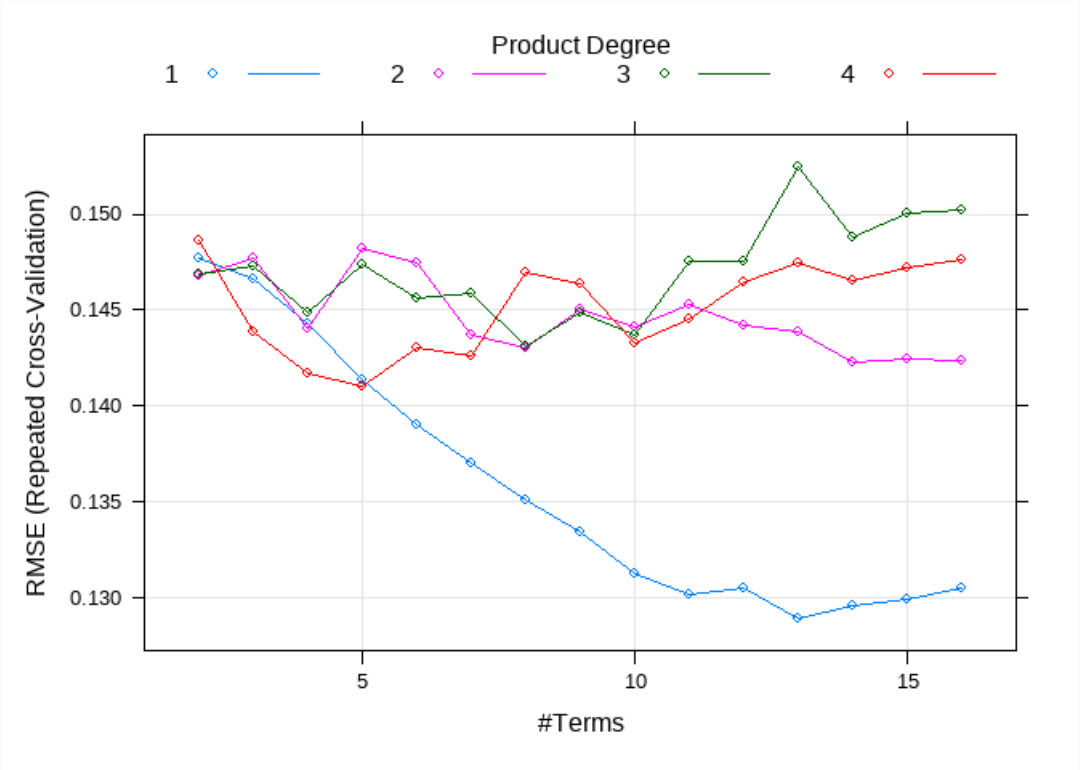



```
## glmnet
##
## 293 samples
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 264, 265, 263, 262, 262, 265, ...
## Resampling results across tuning parameters:
##
##   alpha  lambda  RMSE       Rsquared    MAE
##   0.00   0.001   0.1332388  0.35346492  0.1087408
##   0.00   0.010   0.1329327  0.35441146  0.1085980
##   0.00   0.100   0.1351911  0.33489367  0.1105705
##   0.05   0.001   0.1346085  0.34841967  0.1094744
##   0.05   0.010   0.1328785  0.35394311  0.1085175
##   0.05   0.100   0.1369321  0.32682131  0.1117463
##   0.10   0.001   0.1344931  0.34909492  0.1093828
```

```
##   0.10   0.010   0.1330139  0.35211412  0.1085260
##   0.10   0.100   0.1395226  0.31095225  0.1136004
##   0.15   0.001   0.1343786  0.34978470  0.1092907
##   0.15   0.010   0.1332024  0.35011650  0.1085425
##   0.15   0.100   0.1423583  0.28961710  0.1159275
##   0.20   0.001   0.1342655  0.35048881  0.1092079
##   0.20   0.010   0.1334126  0.34807614  0.1085730
##   0.20   0.100   0.1444951  0.27695873  0.1176361
##   0.25   0.001   0.1341475  0.35126427  0.1091328
##   0.25   0.010   0.1335843  0.34634699  0.1085659
##   0.25   0.100   0.1467114  0.26274468  0.1192910
##   0.30   0.001   0.1340409  0.35193757  0.1090683
##   0.30   0.010   0.1337804  0.34450983  0.1085840
##   0.30   0.100   0.1489935  0.24611201  0.1209590
##   0.35   0.001   0.1339531  0.35242796  0.1090137
##   0.35   0.010   0.1340090  0.34238017  0.1086801
##   0.35   0.100   0.1508001  0.23712873  0.1223073
##   0.40   0.001   0.1338675  0.35291814  0.1089543
##   0.40   0.010   0.1342323  0.34035023  0.1087967
##   0.40   0.100   0.1523609  0.23318035  0.1235431
##   0.45   0.001   0.1337962  0.35330528  0.1089009
##   0.45   0.010   0.1344752  0.33815123  0.1089523
##   0.45   0.100   0.1538975  0.23203462  0.1247510
##   0.50   0.001   0.1337328  0.35363382  0.1088529
##   0.50   0.010   0.1347877  0.33534931  0.1091564
##   0.50   0.100   0.1556150  0.22940971  0.1260504
##   0.55   0.001   0.1336764  0.35390041  0.1088097
##   0.55   0.010   0.1351581  0.33204917  0.1094145
##   0.55   0.100   0.1575306  0.22313968  0.1274061
##   0.60   0.001   0.1336309  0.35407040  0.1087729
##   0.60   0.010   0.1355206  0.32876642  0.1096763
##   0.60   0.100   0.1595672  0.21065905  0.1287831
##   0.65   0.001   0.1335962  0.35416587  0.1087479
##   0.65   0.010   0.1358442  0.32586564  0.1099134
##   0.65   0.100   0.1614697  0.18504668  0.1300252
##   0.70   0.001   0.1335656  0.35423374  0.1087288
##   0.70   0.010   0.1361436  0.32320169  0.1101381
##   0.70   0.100   0.1623911  0.05893092  0.1305879
##   0.75   0.001   0.1335405  0.35427060  0.1087133
##   0.75   0.010   0.1363954  0.32104986  0.1103108
##   0.75   0.100   0.1624776         NaN  0.1306230
##   0.80   0.001   0.1335211  0.35427261  0.1086990
##   0.80   0.010   0.1366421  0.31898851  0.1104783
##   0.80   0.100   0.1624776         NaN  0.1306230
##   0.85   0.001   0.1335084  0.35422645  0.1086920
##   0.85   0.010   0.1368800  0.31700694  0.1106448
##   0.85   0.100   0.1624776         NaN  0.1306230
##   0.90   0.001   0.1334950  0.35417484  0.1086821
##   0.90   0.010   0.1371007  0.31530740  0.1108316
##   0.90   0.100   0.1624776         NaN  0.1306230
##   0.95   0.001   0.1334855  0.35409007  0.1086746
```

```
##   0.95   0.010   0.1372956  0.31388330  0.1110105
##   0.95   0.100   0.1624776         NaN  0.1306230
##   1.00   0.001   0.1334781  0.35399110  0.1086678
##   1.00   0.010   0.1374636  0.31273315  0.1111582
##   1.00   0.100   0.1624776         NaN  0.1306230
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.05 and lambda = 0.01.
```

/// Multivariate Regression Adaptive Splines



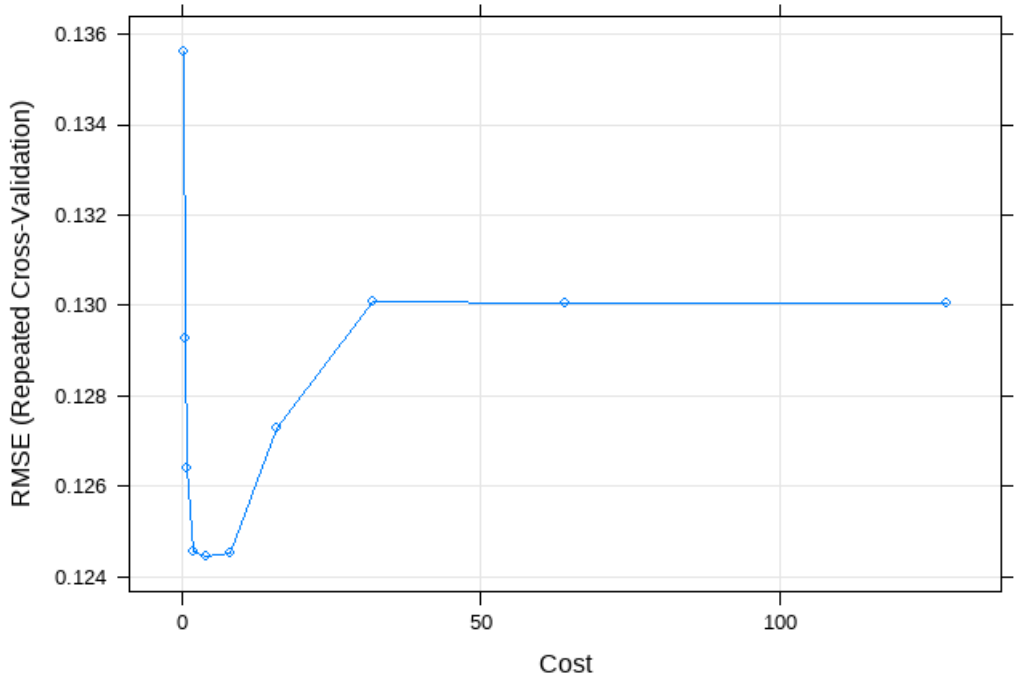

```
## Call: earth(x=data.frame[293,31], y=c(8.26,8.24,8.2...), keepxy=TRUE, degree=1,
##             nprune=13)
##
##                          coefficients
## (Intercept)                 8.1587727
## h(Mnf.Flow-10.2)            0.0010109
## h(103.4-Mnf.Flow)           0.0013696
## h(Usage.cont-22.42)        -0.1407731
## h(Usage.cont-23.8)          0.2990700
## h(Carb.Flow-1078)          -0.0000420
## h(Pressure.Vacuum- -4.8)   -0.2447176
## h(Oxygen.Filler-0.024)     -3.6878597
## h(Oxygen.Filler-0.0436)     3.1416247
## h(48-Pressure.Setpoint)     0.0595239
## h(Pressure.Setpoint-48)     0.0498486
## h(Balling.Lvl-2.8)          3.5821084
```

```
## h(Balling.Lvl-2.86)        -3.4998771
##
## Selected 13 of 51 terms, and 7 of 31 predictors (nprune=13)
## Termination condition: Reached nk 63
## Importance: Usage.cont, Balling.Lvl, Pressure.Setpoint, Carb.Flow, ...
## Number of terms at each degree of interaction: 1 12 (additive model)
## GCV 0.01349215    RSS 3.307373    GRSq 0.4943791    RSq 0.5740791
```

```
## Multivariate Adaptive Regression Spline
##
## 293 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 264, 265, 263, 262, 262, 265, ...
## Resampling results across tuning parameters:
##
##   degree  nprune  RMSE       Rsquared   MAE
##   1        2      0.1477307  0.2017492  0.1207944
##   1        3      0.1466039  0.2253840  0.1176397
##   1        4      0.1443140  0.2418771  0.1157783
##   1        5      0.1413796  0.2728738  0.1126792
##   1        6      0.1390735  0.3045130  0.1103752
##   1        7      0.1370549  0.3273328  0.1081837
##   1        8      0.1351049  0.3514680  0.1056816
##   1        9      0.1334433  0.3626924  0.1049138
##   1       10      0.1312693  0.3932143  0.1021944
##   1       11      0.1302183  0.3989301  0.1019381
##   1       12      0.1305454  0.4017721  0.1019833
##   1       13      0.1289224  0.4128399  0.1010213
##   1       14      0.1296121  0.4137858  0.1012863
##   1       15      0.1299676  0.4149265  0.1014190
##   1       16      0.1305020  0.4121160  0.1018437
##   2        2      0.1468020  0.2090953  0.1175196
##   2        3      0.1477433  0.2062772  0.1186215
##   2        4      0.1440059  0.2491768  0.1151062
##   2        5      0.1481662  0.2613755  0.1135765
##   2        6      0.1474748  0.2878849  0.1123709
##   2        7      0.1437272  0.3199114  0.1093055
##   2        8      0.1429992  0.3394200  0.1081049
##   2        9      0.1450655  0.3397401  0.1075969
##   2       10      0.1440856  0.3557123  0.1071488
##   2       11      0.1453033  0.3521093  0.1076478
##   2       12      0.1441850  0.3617336  0.1072866
##   2       13      0.1438450  0.3732275  0.1064592
##   2       14      0.1422968  0.3863047  0.1048706
##   2       15      0.1424164  0.3894277  0.1044528
```

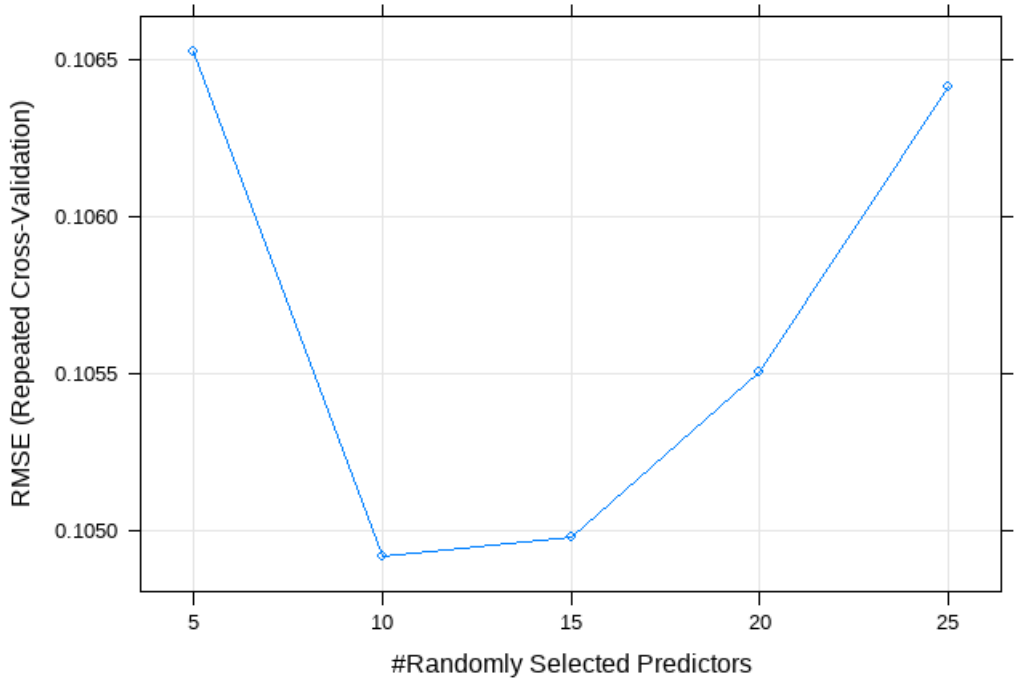
```
##   2       16      0.1423766  0.3918542  0.1047176
##   3        2      0.1468390  0.2065660  0.1176013
##   3        3      0.1472761  0.2074589  0.1179461
##   3        4      0.1448517  0.2367116  0.1139593
##   3        5      0.1473649  0.2640252  0.1132456
##   3        6      0.1456381  0.2879717  0.1106447
##   3        7      0.1458309  0.3007671  0.1094610
##   3        8      0.1431217  0.3185659  0.1079344
##   3        9      0.1448858  0.3265529  0.1090083
##   3       10      0.1437430  0.3411754  0.1074945
##   3       11      0.1475232  0.3450355  0.1081748
##   3       12      0.1475255  0.3587199  0.1073083
##   3       13      0.1524766  0.3527178  0.1093153
##   3       14      0.1488027  0.3804829  0.1064251
##   3       15      0.1500493  0.3712773  0.1083291
##   3       16      0.1501740  0.3730271  0.1077815
##   4        2      0.1486149  0.1980599  0.1183403
##   4        3      0.1438974  0.2369099  0.1157479
##   4        4      0.1416838  0.2752729  0.1125845
##   4        5      0.1410654  0.2986226  0.1107969
##   4        6      0.1430664  0.3163437  0.1100344
##   4        7      0.1426353  0.3289097  0.1084977
##   4        8      0.1469375  0.3207452  0.1108965
##   4        9      0.1463915  0.3327757  0.1100414
##   4       10      0.1432485  0.3504133  0.1081497
##   4       11      0.1445430  0.3531211  0.1080858
##   4       12      0.1464655  0.3631823  0.1081897
##   4       13      0.1474976  0.3612870  0.1084256
##   4       14      0.1465372  0.3703402  0.1083668
##   4       15      0.1472054  0.3677665  0.1083449
##   4       16      0.1476180  0.3688315  0.1091523
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were nprune = 13 and degree = 1.
```

/// Support Vector Machines



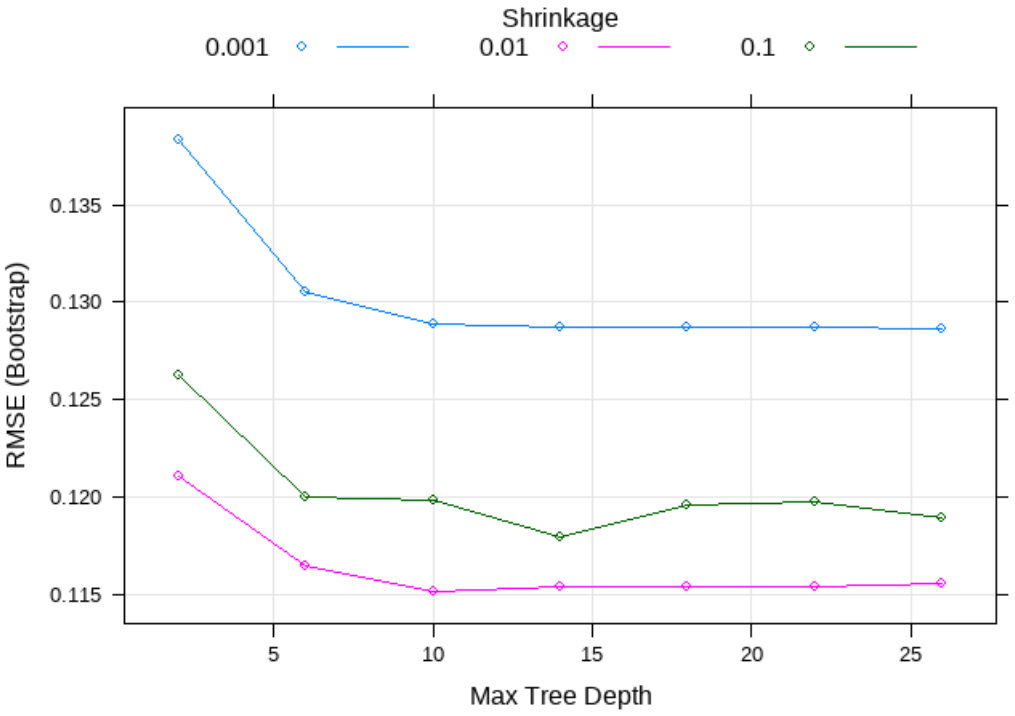

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 293 samples
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 264, 265, 263, 262, 262, 265, ...
## Resampling results across tuning parameters:
##
##   C       RMSE       Rsquared   MAE
##     0.25  0.1356039  0.3616765  0.10739658
##     0.50  0.1292759  0.4033335  0.10107426
##     1.00  0.1264060  0.4176263  0.09702177
##     2.00  0.1245550  0.4312279  0.09592860
##     4.00  0.1244594  0.4352141  0.09605752
##     8.00  0.1245393  0.4394588  0.09610741
##    16.00  0.1273084  0.4263824  0.09797349
##    32.00  0.1300875  0.4105093  0.10012951
##    64.00  0.1300589  0.4108182  0.10010790
##   128.00  0.1300589  0.4108182  0.10010790
##
## Tuning parameter 'sigma' was held constant at a value of 0.0220124
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.0220124 and C = 4.
```

/// Random Forest




```
## Random Forest
##
## 293 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 264, 265, 263, 262, 262, 265, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE       Rsquared   MAE
##    5    0.1065265  0.6120845  0.08319996
##   10    0.1049171  0.6089742  0.08124497
##   15    0.1049797  0.6015278  0.08062560
##   20    0.1055052  0.5928213  0.08075500
##   25    0.1064127  0.5829977  0.08102333
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 10.
```

/// Gradient Boosting Machines




```
## Stochastic Gradient Boosting
##
## 293 samples
##  31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 293, 293, 293, 293, 293, 293, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  RMSE       Rsquared   MAE
##   0.001       2                 0.1383642  0.3994768  0.11205111
##   0.001       6                 0.1305219  0.4656558  0.10468512
##   0.001      10                 0.1289055  0.4789570  0.10299587
##   0.001      14                 0.1287304  0.4812120  0.10284774
##   0.001      18                 0.1286952  0.4812779  0.10282721
##   0.001      22                 0.1287446  0.4809608  0.10283584
##   0.001      26                 0.1286422  0.4830290  0.10280721
##   0.010       2                 0.1210246  0.4598225  0.09633971
##   0.010       6                 0.1164711  0.4993123  0.09119509
##   0.010      10                 0.1151155  0.5115580  0.08959584
##   0.010      14                 0.1153853  0.5091378  0.08969473
##   0.010      18                 0.1154146  0.5083476  0.08980144
##   0.010      22                 0.1153741  0.5090933  0.08970877
##   0.010      26                 0.1155116  0.5078493  0.08993658
##   0.100       2                 0.1262696  0.4316460  0.09966744
##   0.100       6                 0.1199834  0.4725097  0.09417486
##   0.100      10                 0.1197990  0.4736872  0.09341431
##   0.100      14                 0.1179340  0.4876049  0.09208457
```

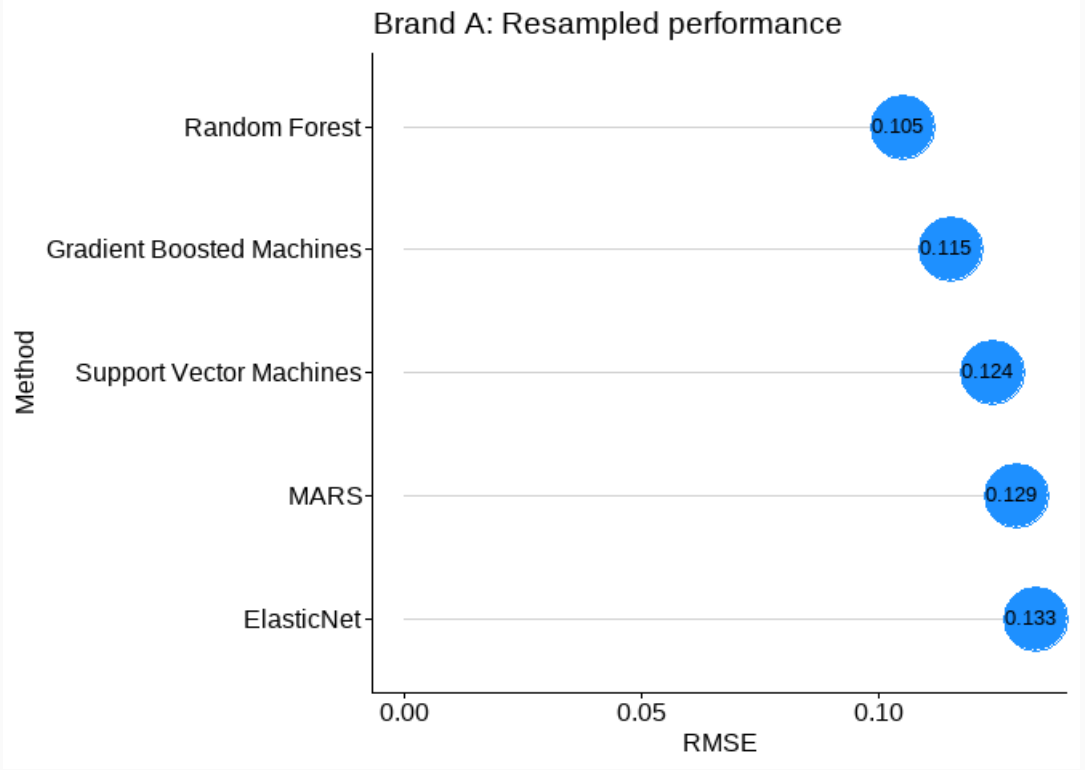
```
##   0.100      18                 0.1195896  0.4737482  0.09330322
##   0.100      22                 0.1197221  0.4740848  0.09309288
##   0.100      26                 0.1188972  0.4814946  0.09304033
##
## Tuning parameter 'n.trees' was held constant at a value of 1000
##
## Tuning parameter 'n.minobsinnode' was held constant at a value of 10
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 1000, interaction.depth
##  = 10, shrinkage = 0.01 and n.minobsinnode = 10.
```

// Brand A: Model Selection

The **Random Forest** model achieves the lowest resampled RMSE on brand **A** among the five models with score of **0.105**.

Brand A: Resampled performance

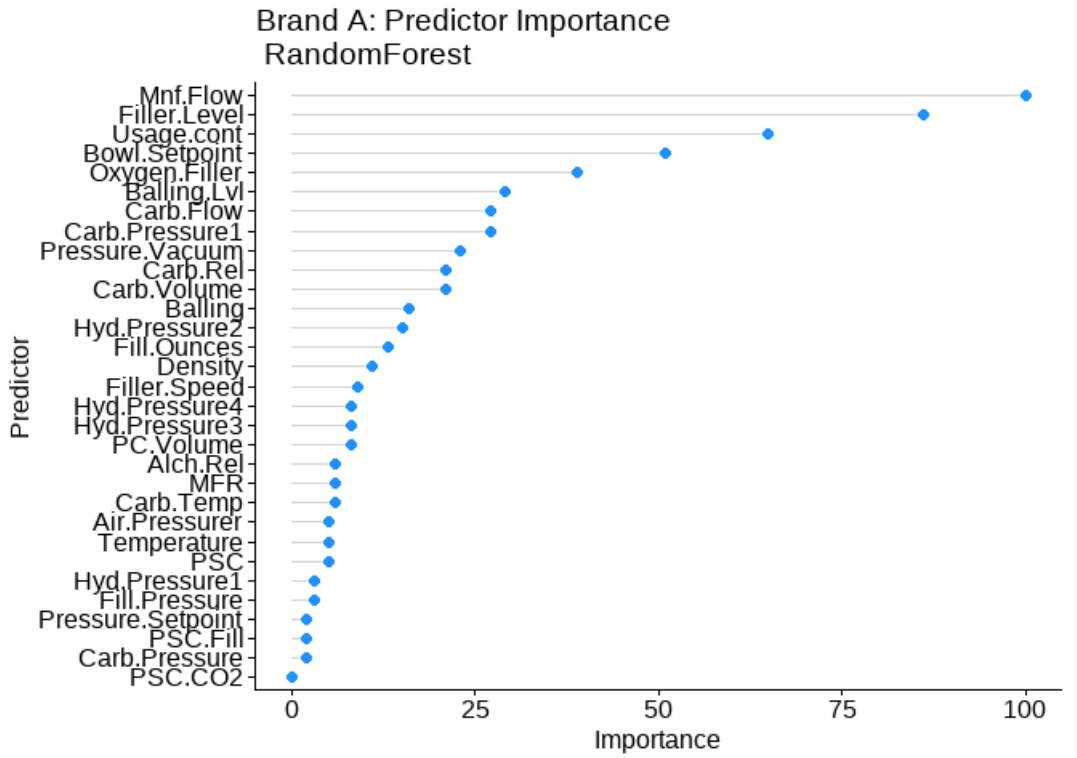
Method	RMSE
Random Forest	0.105
Gradient Boosted Machines	0.115
Support Vector Machines	0.124
MARS	0.129
ElasticNet	0.133



// Brand A: Model Variable Importance

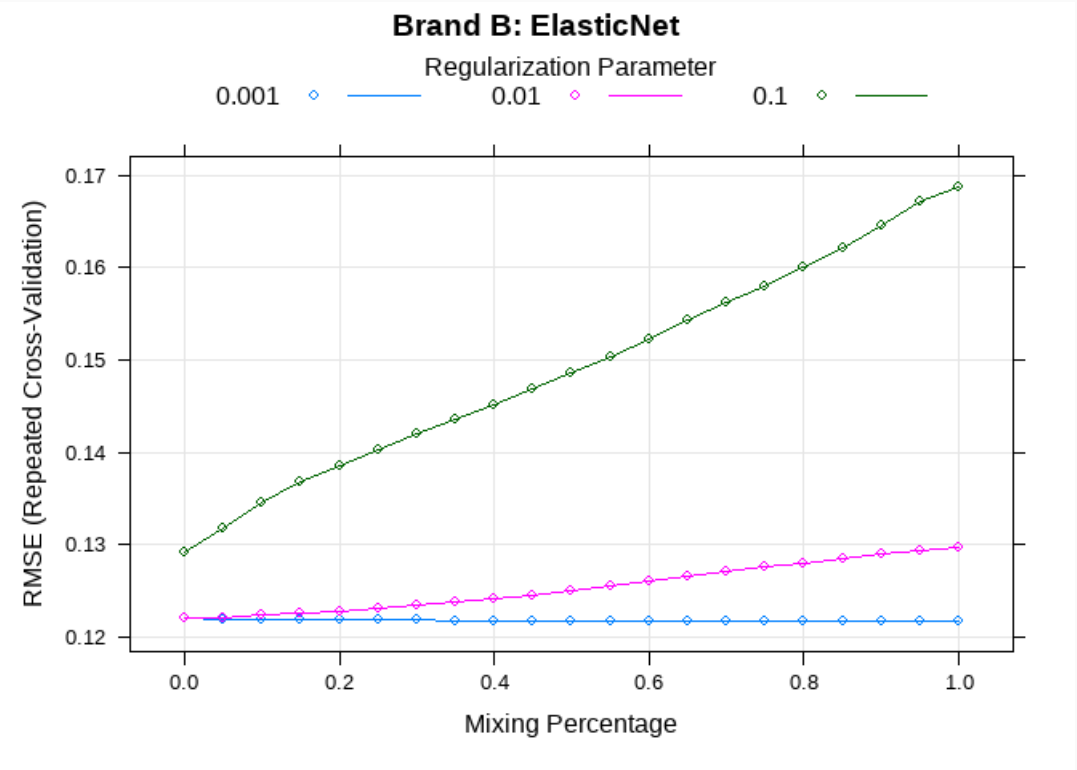
Brand A: Predictor Importance RandomForest

Predictor	Importance
Mnf.Flow	100
Filler.Level	86
Usage.cont	65
Bowl.Setpoint	51
Oxygen.Filler	39
Balling.Lvl	29
Carb.Pressure1	27
Carb.Flow	27
Pressure.Vacuum	23
Carb.Volume	21
Carb.Rel	21
Balling	16
Hyd.Pressure2	15
Fill.Ounces	13
Density	11
Filler.Speed	9
PC.Volume	8
Hyd.Pressure3	8
Hyd.Pressure4	8
Carb.Temp	6
MFR	6
Alch.Rel	6
PSC	5
Temperature	5
Air.Pressurer	5
Fill.Pressure	3
Hyd.Pressure1	3
Carb.Pressure	2
PSC.Fill	2
Pressure.Setpoint	2
PSC.CO2	0



// Brand B: Model Fitting

/// ElasticNet

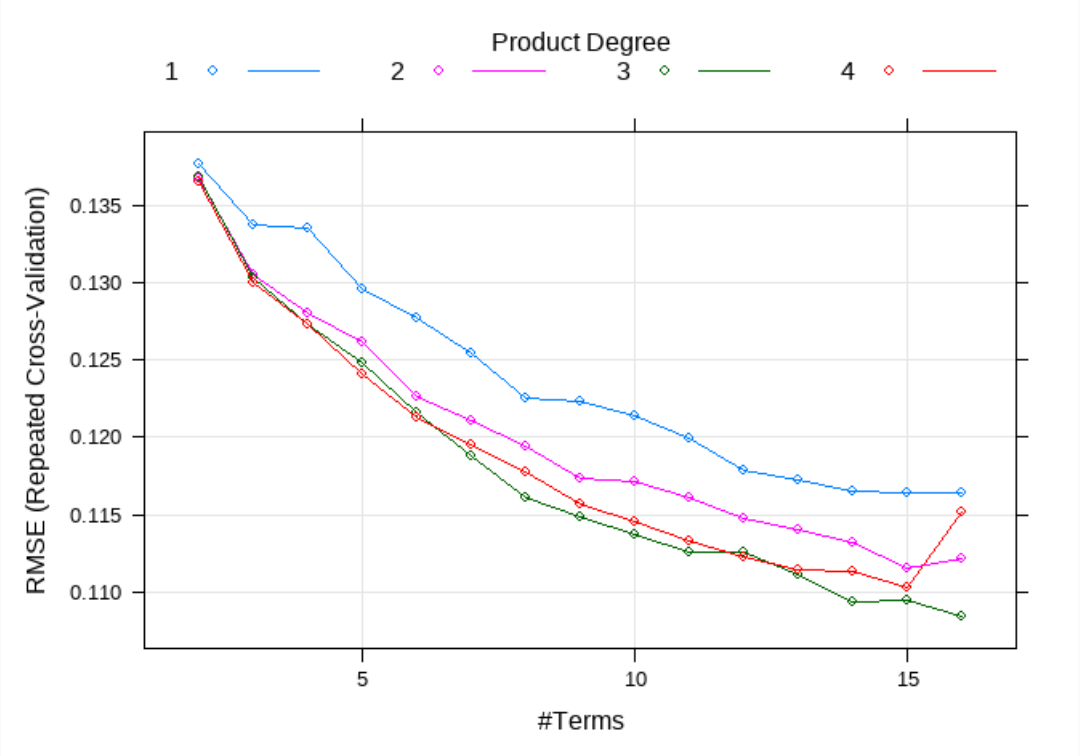



```
## glmnet
##
```

```
## 1235 samples
##   31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 1110, 1111, 1113, 1111, 1110, 1112, ...
## Resampling results across tuning parameters:
##
##   alpha  lambda  RMSE       Rsquared   MAE
##   0.00   0.001   0.1221352  0.4799921  0.09287054
##   0.00   0.010   0.1221496  0.4798982  0.09289133
##   0.00   0.100   0.1291649  0.4340038  0.10215511
##   0.05   0.001   0.1219647  0.4826834  0.09258527
##   0.05   0.010   0.1221806  0.4799844  0.09301128
##   0.05   0.100   0.1318411  0.4172356  0.10505941
##   0.10   0.001   0.1219373  0.4828500  0.09256673
##   0.10   0.010   0.1223660  0.4788175  0.09325143
##   0.10   0.100   0.1346500  0.3969166  0.10781761
##   0.15   0.001   0.1219182  0.4829419  0.09255723
##   0.15   0.010   0.1225902  0.4773433  0.09354689
##   0.15   0.100   0.1367647  0.3847534  0.10979899
##   0.20   0.001   0.1219032  0.4830139  0.09255091
##   0.20   0.010   0.1228385  0.4756795  0.09385095
##   0.20   0.100   0.1386139  0.3764203  0.11151436
##   0.25   0.001   0.1218827  0.4831312  0.09254118
##   0.25   0.010   0.1231186  0.4737624  0.09419823
##   0.25   0.100   0.1403068  0.3711069  0.11314012
##   0.30   0.001   0.1218592  0.4832776  0.09252887
##   0.30   0.010   0.1234252  0.4716238  0.09458447
##   0.30   0.100   0.1419997  0.3669859  0.11476703
##   0.35   0.001   0.1218366  0.4834220  0.09252177
##   0.35   0.010   0.1237530  0.4693149  0.09501067
##   0.35   0.100   0.1436230  0.3651437  0.11631029
##   0.40   0.001   0.1218153  0.4835595  0.09251438
##   0.40   0.010   0.1241226  0.4666279  0.09550680
##   0.40   0.100   0.1452298  0.3645507  0.11784864
##   0.45   0.001   0.1217963  0.4836835  0.09251113
##   0.45   0.010   0.1245437  0.4634647  0.09607134
##   0.45   0.100   0.1469247  0.3635090  0.11941280
##   0.50   0.001   0.1217772  0.4838139  0.09250663
##   0.50   0.010   0.1250194  0.4597568  0.09668039
##   0.50   0.100   0.1485941  0.3630187  0.12095807
##   0.55   0.001   0.1217597  0.4839345  0.09250096
##   0.55   0.010   0.1255332  0.4556553  0.09732062
##   0.55   0.100   0.1503504  0.3607207  0.12252324
##   0.60   0.001   0.1217497  0.4839869  0.09249761
##   0.60   0.010   0.1260669  0.4513087  0.09797545
##   0.60   0.100   0.1522634  0.3557047  0.12426989
##   0.65   0.001   0.1217454  0.4839958  0.09249632
##   0.65   0.010   0.1265829  0.4470687  0.09859053
##   0.65   0.100   0.1543099  0.3468180  0.12622547
```

```
##   0.70   0.001   0.1217532  0.4839040  0.09250420
##   0.70   0.010   0.1270804  0.4429531  0.09917720
##   0.70   0.100   0.1562174  0.3408097  0.12809353
##   0.75   0.001   0.1217619  0.4838108  0.09251191
##   0.75   0.010   0.1275728  0.4388444  0.09974577
##   0.75   0.100   0.1580529  0.3404698  0.12987987
##   0.80   0.001   0.1217721  0.4837070  0.09251966
##   0.80   0.010   0.1280768  0.4345794  0.10032254
##   0.80   0.100   0.1600396  0.3404698  0.13178525
##   0.85   0.001   0.1217822  0.4836067  0.09252530
##   0.85   0.010   0.1285281  0.4308178  0.10084867
##   0.85   0.100   0.1622080  0.3404698  0.13386866
##   0.90   0.001   0.1217924  0.4835086  0.09253083
##   0.90   0.010   0.1289575  0.4272349  0.10134308
##   0.90   0.100   0.1645884  0.3404698  0.13621463
##   0.95   0.001   0.1218030  0.4834077  0.09253605
##   0.95   0.010   0.1293727  0.4237689  0.10180787
##   0.95   0.100   0.1671895  0.3404698  0.13881835
##   1.00   0.001   0.1218141  0.4833026  0.09254049
##   1.00   0.010   0.1297422  0.4207636  0.10220735
##   1.00   0.100   0.1687128        NaN  0.14032212
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.65 and lambda = 0.001.
```

/// Multivariate Regression Adaptive Splines



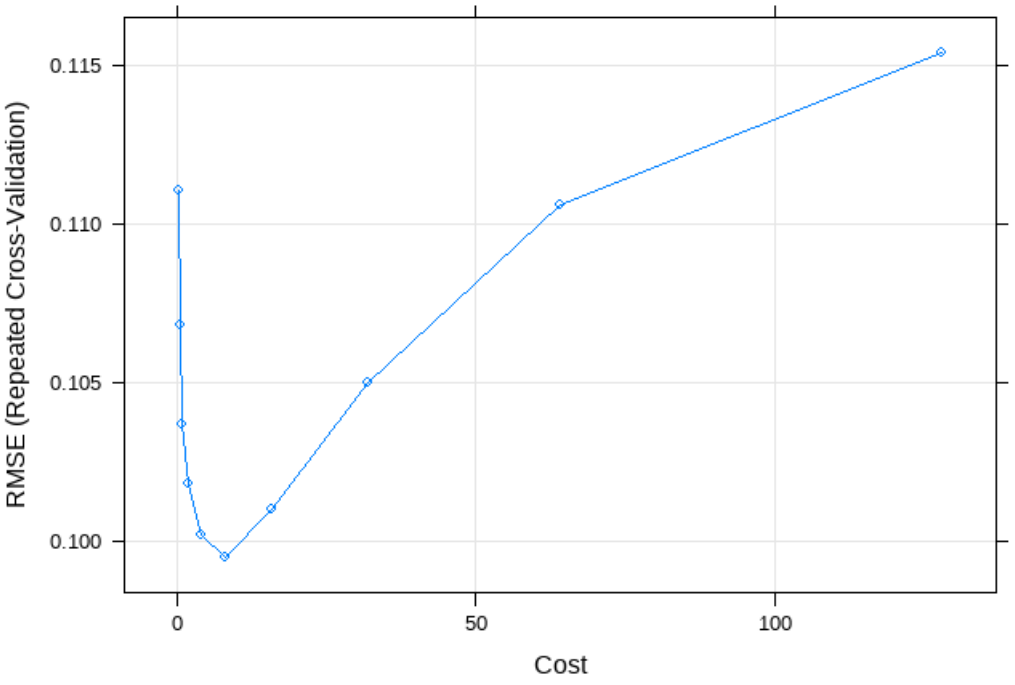
```
## Call: earth(x=data.frame[1235,31], y=c(8.36,8.94,8.3...), keepxy=TRUE,
##             degree=3, nprune=16)
##
##                                                          coefficients
## (Intercept)                                                 8.4756240
## h(0.2-Mnf.Flow)                                             0.0014662
## h(Carb.Flow-1762)                                           0.0000679
## h(Bowl.Setpoint-90)                                         0.0080027
## h(146.8-Air.Pressurer)                                     -0.0339108
## h(Mnf.Flow-0.2) * h(Balling-1.498)                          0.0010763
## h(Mnf.Flow-0.2) * h(1.498-Balling)                          0.0063983
## h(Mnf.Flow-0.2) * h(46-Pressure.Setpoint)                   0.0005416
## h(0.2-Mnf.Flow) * h(Air.Pressurer-143.8)                   -0.0010461
## h(Mnf.Flow-0.2) * h(Carb.Rel-5.34)                         -0.0073018
## h(Carb.Flow-1036) * h(Bowl.Setpoint-90)                    -0.0000030
## h(Mnf.Flow-0.2) * h(1.498-Balling) * h(Bowl.Setpoint-90)    -0.0001750
## h(0.2-Mnf.Flow) * h(Oxygen.Filler-0.106) * h(143.8-Air.Pressurer)   -0.0074590
## h(0.2-Mnf.Flow) * h(0.106-Oxygen.Filler) * h(143.8-Air.Pressurer)    0.0051947
## h(0.2-Mnf.Flow) * h(143.8-Air.Pressurer) * h(5.4-Carb.Rel)  -0.0070274
## h(67-Temperature) * h(Carb.Flow-1036) * h(Bowl.Setpoint-90)  0.0000009
##
## Selected 16 of 49 terms, and 9 of 31 predictors (nprune=16)
## Termination condition: RSq changed by less than 0.001 at 49 terms
## Importance: Mnf.Flow, Air.Pressurer, Bowl.Setpoint, Oxygen.Filler, ...
## Number of terms at each degree of interaction: 1 4 6 5
## GCV 0.01142649    RSS 13.24559    GRSq 0.6001334    RSq 0.6240672
```

```
## Multivariate Adaptive Regression Spline
##
## 1235 samples
##   31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 1110, 1111, 1113, 1111, 1110, 1112, ...
## Resampling results across tuning parameters:
##
##   degree  nprune  RMSE       Rsquared   MAE
##   1        2      0.1376750  0.3384977  0.10864951
##   1        3      0.1336813  0.3767336  0.10554688
##   1        4      0.1334677  0.3791189  0.10462802
##   1        5      0.1295596  0.4140960  0.10074161
##   1        6      0.1276959  0.4311394  0.09772171
##   1        7      0.1254454  0.4507240  0.09491264
##   1        8      0.1225745  0.4752183  0.09241383
##   1        9      0.1223162  0.4781310  0.09178953
##   1       10      0.1213577  0.4864771  0.09057473
##   1       11      0.1199788  0.4980648  0.08946452
##   1       12      0.1178444  0.5149765  0.08776300
```

```
##   1       13      0.1172003  0.5202432  0.08715617
##   1       14      0.1165659  0.5256094  0.08670882
##   1       15      0.1164318  0.5267212  0.08649498
##   1       16      0.1163912  0.5270341  0.08640871
##   2        2      0.1367173  0.3475208  0.10664368
##   2        3      0.1305412  0.4055892  0.09938880
##   2        4      0.1280672  0.4281696  0.09695111
##   2        5      0.1261240  0.4462972  0.09537632
##   2        6      0.1226184  0.4762481  0.09230816
##   2        7      0.1211086  0.4890675  0.09105145
##   2        8      0.1193752  0.5036844  0.08957537
##   2        9      0.1173468  0.5202595  0.08779463
##   2       10      0.1171374  0.5237017  0.08724406
##   2       11      0.1161498  0.5314403  0.08598127
##   2       12      0.1147457  0.5420116  0.08502387
##   2       13      0.1140755  0.5473507  0.08460876
##   2       14      0.1131931  0.5541822  0.08414273
##   2       15      0.1115984  0.5660457  0.08302858
##   2       16      0.1121563  0.5637219  0.08291092
##   3        2      0.1367896  0.3474856  0.10708403
##   3        3      0.1302996  0.4078256  0.09921271
##   3        4      0.1272645  0.4350382  0.09593667
##   3        5      0.1247734  0.4570070  0.09391754
##   3        6      0.1216141  0.4845642  0.09114764
##   3        7      0.1188235  0.5067556  0.08945599
##   3        8      0.1161402  0.5288341  0.08718982
##   3        9      0.1148262  0.5398178  0.08586679
##   3       10      0.1137270  0.5487155  0.08497341
##   3       11      0.1126041  0.5582277  0.08361906
##   3       12      0.1125500  0.5586634  0.08347619
##   3       13      0.1111169  0.5690267  0.08229040
##   3       14      0.1093820  0.5821938  0.08089035
##   3       15      0.1094401  0.5823375  0.08079562
##   3       16      0.1084230  0.5900128  0.08005812
##   4        2      0.1364961  0.3499744  0.10657802
##   4        3      0.1300182  0.4097769  0.09884496
##   4        4      0.1273385  0.4350597  0.09590204
##   4        5      0.1240595  0.4644908  0.09295076
##   4        6      0.1212678  0.4880130  0.09086646
##   4        7      0.1194827  0.5033018  0.08979350
##   4        8      0.1177828  0.5169911  0.08817195
##   4        9      0.1157025  0.5341998  0.08639515
##   4       10      0.1145564  0.5433929  0.08570089
##   4       11      0.1132763  0.5534699  0.08451344
##   4       12      0.1122999  0.5613518  0.08405467
##   4       13      0.1114363  0.5681729  0.08331162
##   4       14      0.1113048  0.5699328  0.08275860
##   4       15      0.1103555  0.5769390  0.08188959
##   4       16      0.1151466  0.5588970  0.08236673
##
```

```
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were nprune = 16 and degree = 3.
```

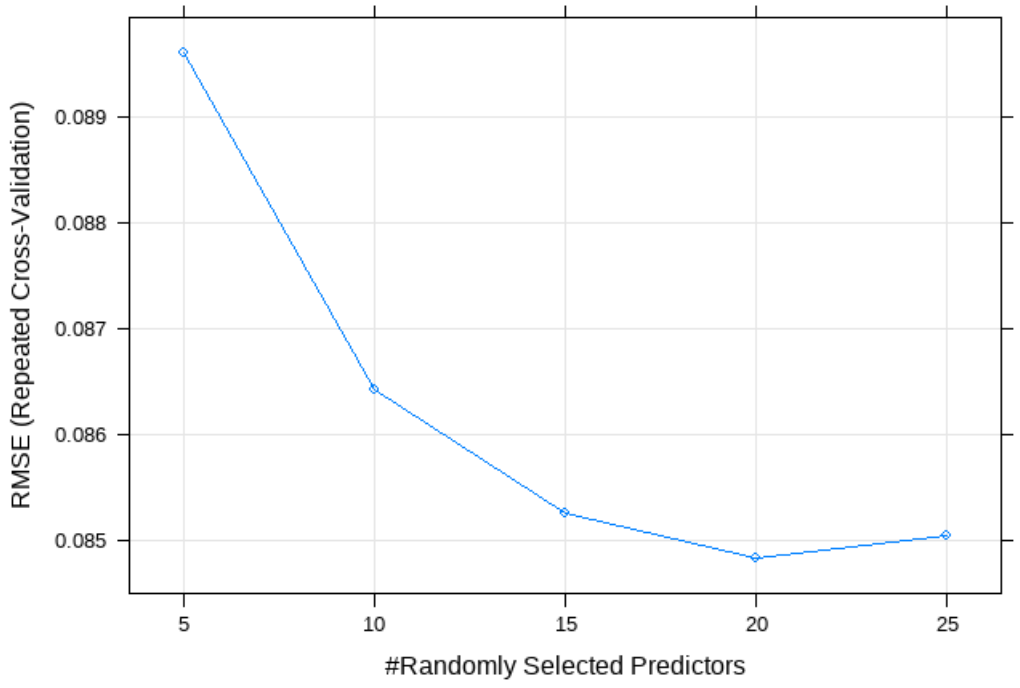
/// Support Vector Machines




```
## Support Vector Machines with Radial Basis Function Kernel
##
## 1235 samples
##   31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 1110, 1111, 1113, 1111, 1110, 1112, ...
## Resampling results across tuning parameters:
##
##   C       RMSE        Rsquared   MAE
##     0.25  0.11107590  0.5784956  0.08080718
##     0.50  0.10681096  0.6064676  0.07719877
##     1.00  0.10371133  0.6270447  0.07485768
##     2.00  0.10181062  0.6392437  0.07376590
##     4.00  0.10022285  0.6501767  0.07373164
##     8.00  0.09950006  0.6554830  0.07383945
##    16.00  0.10100885  0.6475533  0.07532603
##    32.00  0.10501883  0.6252896  0.07871522
##    64.00  0.11060292  0.5968492  0.08277670
##   128.00  0.11540417  0.5739673  0.08652947
##
```

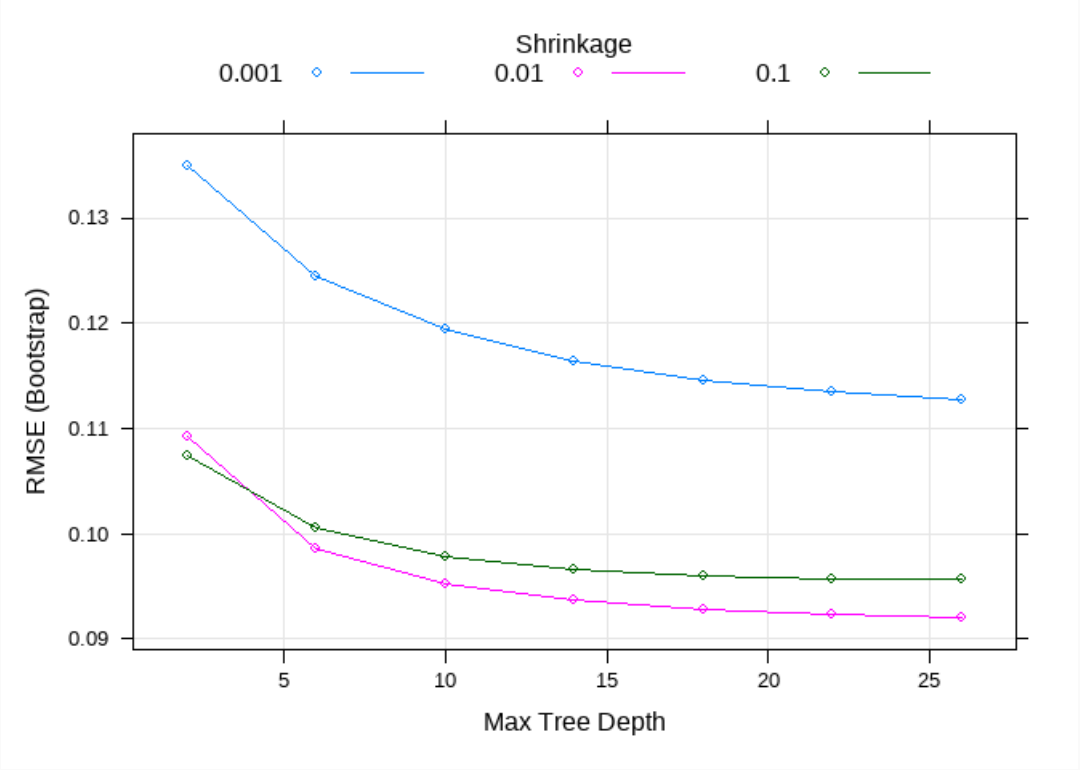
```
## Tuning parameter 'sigma' was held constant at a value of 0.02509541
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.02509541 and C = 8.
```

/// Random Forest




```
## Random Forest
##
## 1235 samples
##   31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 1110, 1111, 1113, 1111, 1110, 1112, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE        Rsquared   MAE
##    5    0.08960100  0.7362178  0.06567331
##   10    0.08641989  0.7494761  0.06272016
##   15    0.08525390  0.7533669  0.06141542
##   20    0.08483502  0.7538913  0.06079571
##   25    0.08505145  0.7509085  0.06065454
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 20.
```

/// Gradient Boosting Machines



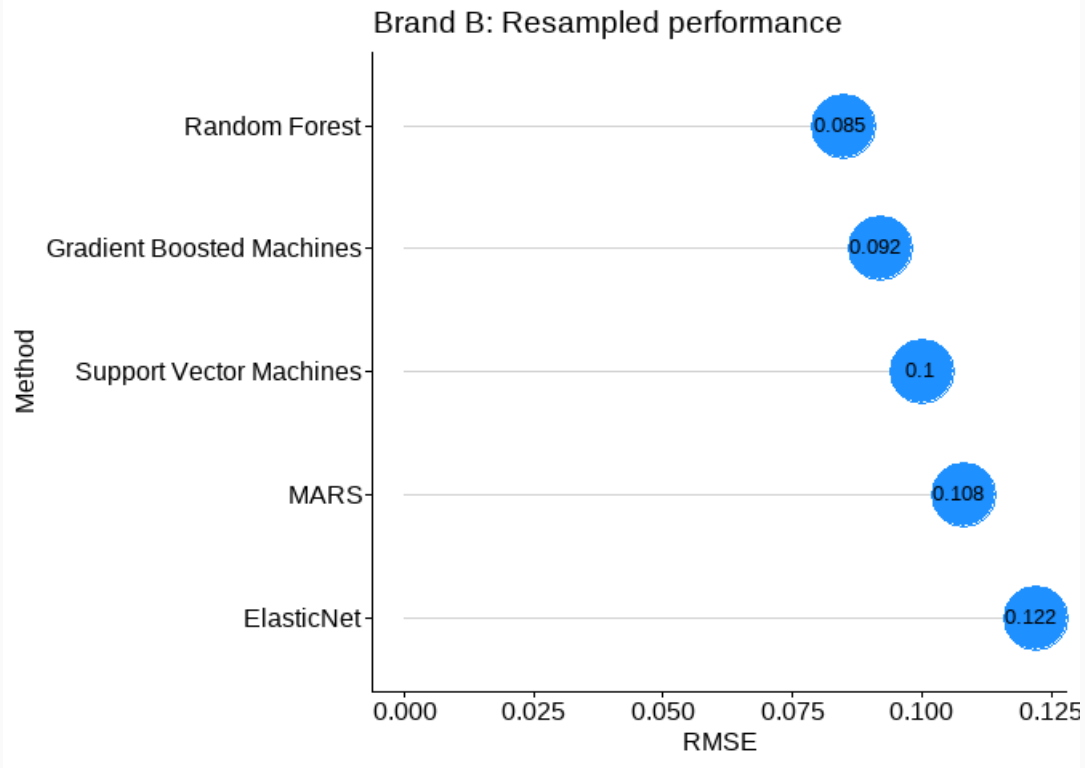

```
## Stochastic Gradient Boosting
##
## 1235 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 1235, 1235, 1235, 1235, 1235, 1235, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  RMSE        Rsquared   MAE
##   0.001       2                 0.13497501  0.4630725  0.10795725
##   0.001       6                 0.12447097  0.5774163  0.09870212
##   0.001      10                 0.11941045  0.6251654  0.09452001
##   0.001      14                 0.11644843  0.6509606  0.09186487
##   0.001      18                 0.11459034  0.6657442  0.09011714
##   0.001      22                 0.11342373  0.6745021  0.08895466
##   0.001      26                 0.11267435  0.6793795  0.08814625
##   0.010       2                 0.10917210  0.5940665  0.08264528
##   0.010       6                 0.09855487  0.6655718  0.07329143
##   0.010      10                 0.09524157  0.6872370  0.07045522
##   0.010      14                 0.09372454  0.6969449  0.06895925
##   0.010      18                 0.09277180  0.7030271  0.06784778
##   0.010      22                 0.09232587  0.7059601  0.06726007
##   0.010      26                 0.09200167  0.7079767  0.06678317
##   0.100       2                 0.10737738  0.6036566  0.08059131
##   0.100       6                 0.10057668  0.6504507  0.07488097
```

```
##   0.100      10                 0.09780341  0.6689041  0.07246936
##   0.100      14                 0.09653579  0.6770245  0.07122981
##   0.100      18                 0.09592440  0.6809402  0.07063402
##   0.100      22                 0.09563757  0.6831361  0.07028118
##   0.100      26                 0.09564170  0.6829185  0.06990467
##
## Tuning parameter 'n.trees' was held constant at a value of 1000
##
## Tuning parameter 'n.minobsinnode' was held constant at a value of 10
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 1000, interaction.depth
##  = 26, shrinkage = 0.01 and n.minobsinnode = 10.
```

// Brand B: Model Selection

The **Random Forest** model achieves the lowest resampled RMSE on brand **B** among the five models with score of **0.085**.

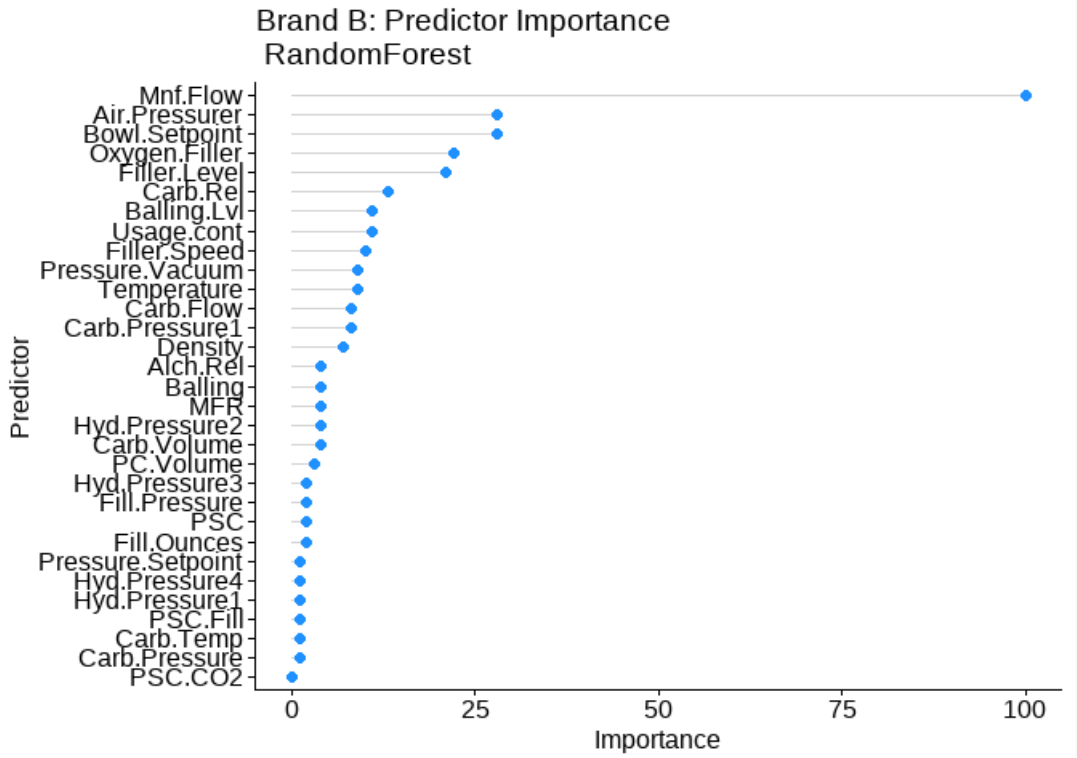
Brand B: Resampled performance	
Method	**RMSE**
Random Forest	0.085
Gradient Boosted Machines	0.092
Support Vector Machines	0.100
MARS	0.108
ElasticNet	0.122



// Brand B: Model Variable Importance

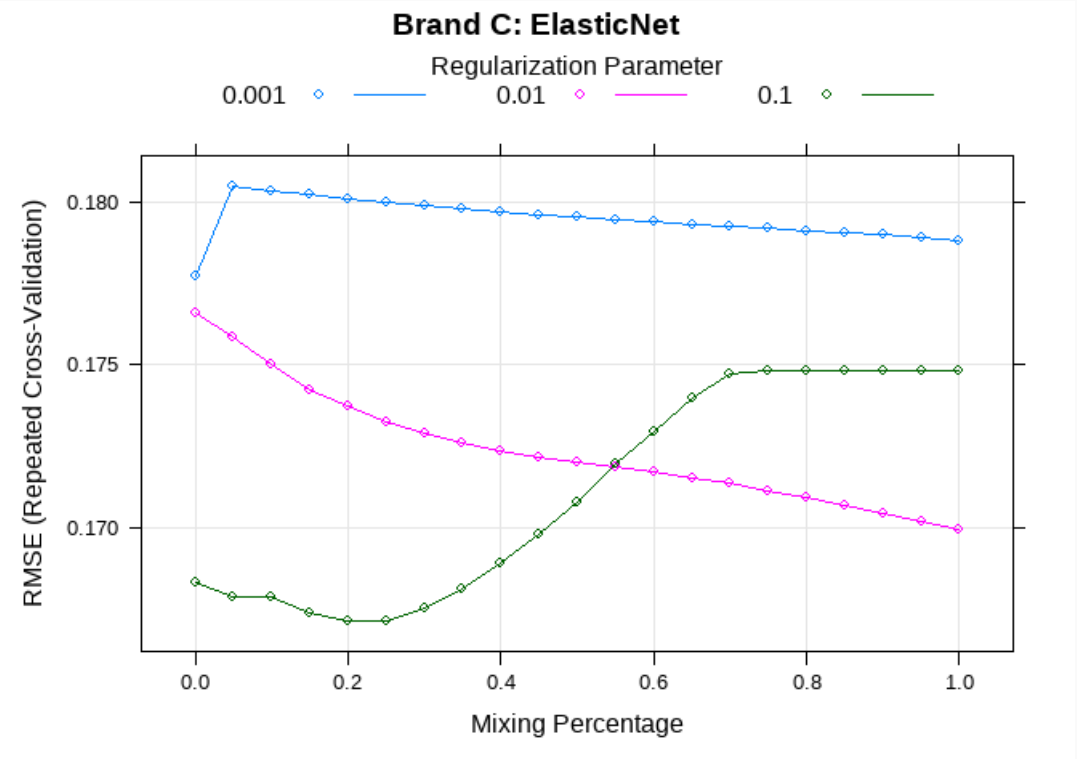
Brand B: Predictor Importance RandomForest

Predictor	Importance
Mnf.Flow	100
Bowl.Setpoint	28
Air.Pressurer	28
Oxygen.Filler	22
Filler.Level	21
Carb.Rel	13
Usage.cont	11
Balling.Lvl	11
Filler.Speed	10
Temperature	9
Pressure.Vacuum	9
Carb.Pressure1	8
Carb.Flow	8
Density	7
Carb.Volume	4
Hyd.Pressure2	4
MFR	4
Balling	4
Alch.Rel	4
PC.Volume	3
Fill.Ounces	2
PSC	2
Fill.Pressure	2
Hyd.Pressure3	2
Carb.Pressure	1
Carb.Temp	1
PSC.Fill	1
Hyd.Pressure1	1
Hyd.Pressure4	1
Pressure.Setpoint	1
PSC.CO2	0



// Brand C: Model Fitting

/// ElasticNet

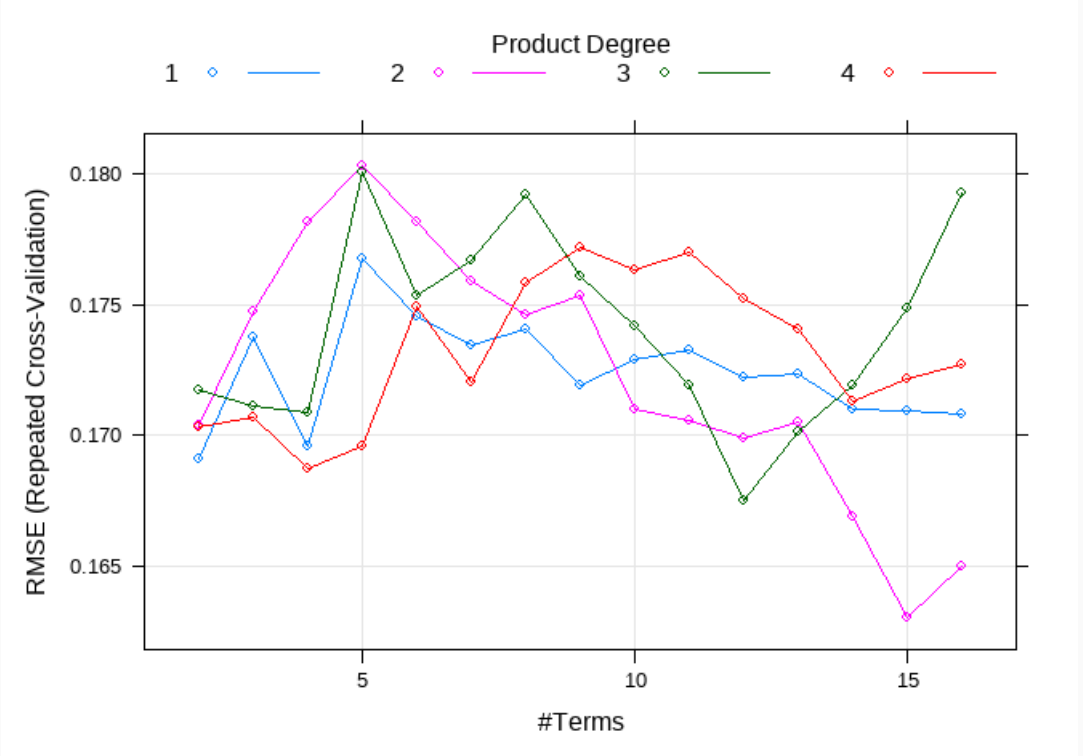



```
## glmnet
##
## 304 samples
```

```
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 274, 272, 272, 274, 275, 274, ...
## Resampling results across tuning parameters:
##
##   alpha  lambda  RMSE       Rsquared      MAE
##   0.00   0.001   0.1777149  0.1314626068  0.1315451
##   0.00   0.010   0.1765970  0.1336267288  0.1311131
##   0.00   0.100   0.1683537  0.1426561530  0.1289834
##   0.05   0.001   0.1804645  0.1272096871  0.1327622
##   0.05   0.010   0.1758561  0.1358828743  0.1307335
##   0.05   0.100   0.1678838  0.1454955345  0.1285634
##   0.10   0.001   0.1803310  0.1275659611  0.1326666
##   0.10   0.010   0.1750294  0.1386988585  0.1303396
##   0.10   0.100   0.1679007  0.1437538783  0.1289355
##   0.15   0.001   0.1802120  0.1278285845  0.1325830
##   0.15   0.010   0.1742566  0.1413192291  0.1299264
##   0.15   0.100   0.1674325  0.1485140350  0.1292390
##   0.20   0.001   0.1800983  0.1280592277  0.1325039
##   0.20   0.010   0.1737227  0.1424053011  0.1296710
##   0.20   0.100   0.1671645  0.1527540629  0.1293901
##   0.25   0.001   0.1799896  0.1282659273  0.1324319
##   0.25   0.010   0.1732579  0.1431253269  0.1294763
##   0.25   0.100   0.1671732  0.1577947121  0.1295252
##   0.30   0.001   0.1798852  0.1284603974  0.1323627
##   0.30   0.010   0.1728895  0.1435100789  0.1293415
##   0.30   0.100   0.1675712  0.1630925292  0.1299535
##   0.35   0.001   0.1797837  0.1286529013  0.1322996
##   0.35   0.010   0.1726067  0.1435672355  0.1292500
##   0.35   0.100   0.1681661  0.1653698595  0.1305408
##   0.40   0.001   0.1796850  0.1288344349  0.1322341
##   0.40   0.010   0.1723774  0.1436174913  0.1291945
##   0.40   0.100   0.1689247  0.1657526658  0.1312441
##   0.45   0.001   0.1795966  0.1289755029  0.1321780
##   0.45   0.010   0.1721830  0.1435617423  0.1292176
##   0.45   0.100   0.1698029  0.1661443474  0.1320246
##   0.50   0.001   0.1795259  0.1290585113  0.1321380
##   0.50   0.010   0.1720183  0.1433549069  0.1292788
##   0.50   0.100   0.1708207  0.1661918959  0.1329229
##   0.55   0.001   0.1794532  0.1291538590  0.1320987
##   0.55   0.010   0.1718822  0.1429696144  0.1293456
##   0.55   0.100   0.1719557  0.1661918959  0.1339457
##   0.60   0.001   0.1793831  0.1292393674  0.1320638
##   0.60   0.010   0.1717322  0.1425896679  0.1294171
##   0.60   0.100   0.1729413  0.1358615121  0.1349061
##   0.65   0.001   0.1793148  0.1293284790  0.1320306
##   0.65   0.010   0.1715614  0.1424105943  0.1294947
##   0.65   0.100   0.1739834  0.1358615121  0.1359330
##   0.70   0.001   0.1792502  0.1293934215  0.1319992
```

```
##   0.70   0.010   0.1713661  0.1425319164  0.1295488
##   0.70   0.100   0.1747245  0.0325725862  0.1366752
##   0.75   0.001   0.1791854  0.1294482026  0.1319680
##   0.75   0.010   0.1711609  0.1427829138  0.1295774
##   0.75   0.100   0.1748315  0.0008221073  0.1368255
##   0.80   0.001   0.1791218  0.1294788420  0.1319404
##   0.80   0.010   0.1709556  0.1431024991  0.1295707
##   0.80   0.100   0.1748325           NaN  0.1368281
##   0.85   0.001   0.1790608  0.1295039523  0.1319161
##   0.85   0.010   0.1707093  0.1438716399  0.1295045
##   0.85   0.100   0.1748325           NaN  0.1368281
##   0.90   0.001   0.1789883  0.1295745759  0.1318844
##   0.90   0.010   0.1704741  0.1443386881  0.1294482
##   0.90   0.100   0.1748325           NaN  0.1368281
##   0.95   0.001   0.1789080  0.1296867579  0.1318486
##   0.95   0.010   0.1702230  0.1448265384  0.1293806
##   0.95   0.100   0.1748325           NaN  0.1368281
##   1.00   0.001   0.1787976  0.1298868080  0.1318001
##   1.00   0.010   0.1699771  0.1451337499  0.1293262
##   1.00   0.100   0.1748325           NaN  0.1368281
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.2 and lambda = 0.1.
```

/// Multivariate Regression Adaptive Splines



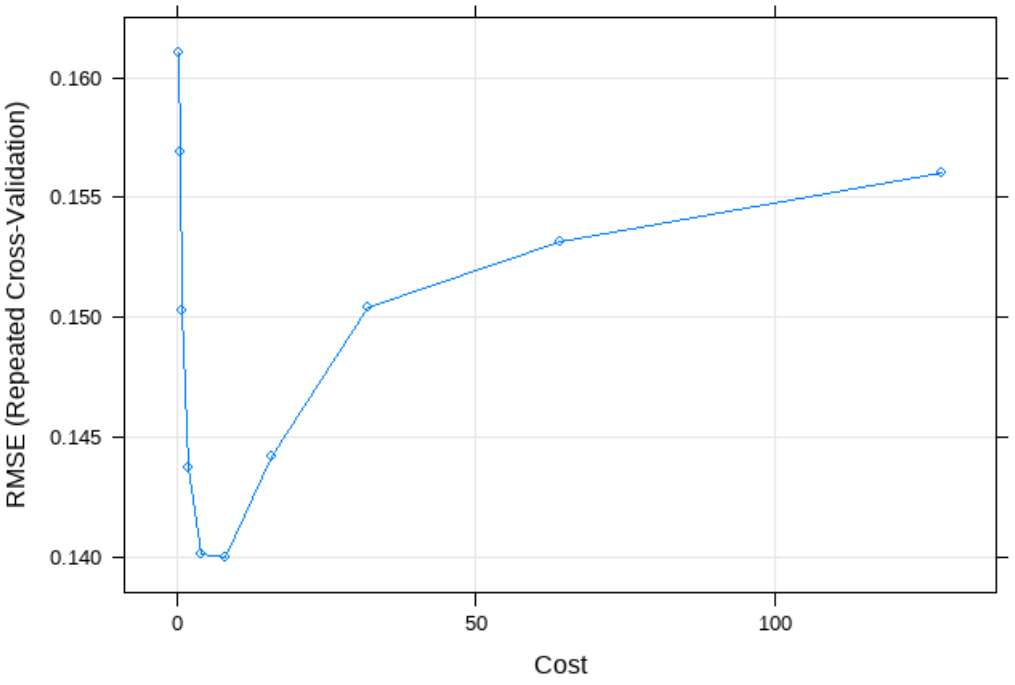
```
## Call: earth(x=data.frame[304,31], y=c(8.58,8.3,8.42...), keepxy=TRUE, degree=2,
##             nprune=15)
##
##                                                coefficients
## (Intercept)                                       8.4431657
## h(Mnf.Flow-150.4)                                -0.0038876
## h(Temperature-65.4)                               0.0337407
## h(Usage.cont-23.98)                              -0.4558221
## h(Carb.Rel-5.3)                                  -0.7058176
## h(1.58-Balling.Lvl)                               1.4007278
## PSC.Fill * h(65.4-Temperature)                   -0.9972436
## h(PC.Volume-0.268667) * h(37.6-Hyd.Pressure3)    -0.0361101
## h(Mnf.Flow-150.4) * h(12-Hyd.Pressure1)           0.0205018
## h(Hyd.Pressure1-13.6) * h(1.58-Balling.Lvl)      -0.0455243
## h(29.6-Hyd.Pressure2) * h(1.58-Balling.Lvl)      -0.0364174
## h(Temperature-65.4) * h(Pressure.Vacuum- -5.6)   -0.0831831
## h(Pressure.Vacuum- -5.4) * h(Carb.Rel-5.3)        1.8340741
## h(Pressure.Vacuum- -5.2) * h(1.58-Balling.Lvl)    1.3988651
## h(0.064-Oxygen.Filler) * h(1.58-Balling.Lvl)    -27.2095329
##
## Selected 15 of 45 terms, and 12 of 31 predictors (nprune=15)
## Termination condition: RSq changed by less than 0.001 at 45 terms
## Importance: Balling.Lvl, Oxygen.Filler, PSC.Fill, Temperature, ...
## Number of terms at each degree of interaction: 1 5 9
## GCV 0.01619871    RSS 3.827158    GRSq 0.4853858    RSq 0.5974071
```

```
## Multivariate Adaptive Regression Spline
##
## 304 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 274, 272, 272, 274, 275, 274, ...
## Resampling results across tuning parameters:
##
##   degree  nprune  RMSE       Rsquared    MAE
##   1        2      0.1690699  0.10136697  0.1309722
##   1        3      0.1737341  0.10755800  0.1326774
##   1        4      0.1695853  0.12547506  0.1284472
##   1        5      0.1767461  0.12777462  0.1305807
##   1        6      0.1745222  0.12634651  0.1293857
##   1        7      0.1734658  0.14701453  0.1297478
##   1        8      0.1740628  0.16053105  0.1298596
##   1        9      0.1718960  0.17460817  0.1284809
##   1       10      0.1728673  0.18044917  0.1285464
##   1       11      0.1732924  0.18224491  0.1285015
##   1       12      0.1722208  0.18922005  0.1285275
##   1       13      0.1723121  0.19491047  0.1282896
```

```
##   1       14      0.1709852  0.20277469  0.1269438
##   1       15      0.1709470  0.20395461  0.1270162
##   1       16      0.1708269  0.20581360  0.1264084
##   2        2      0.1703481  0.10374491  0.1321072
##   2        3      0.1747368  0.10750053  0.1347114
##   2        4      0.1781952  0.10150630  0.1356576
##   2        5      0.1803301  0.13411797  0.1317868
##   2        6      0.1781708  0.18924063  0.1278998
##   2        7      0.1758796  0.20462038  0.1266820
##   2        8      0.1746004  0.22158580  0.1246681
##   2        9      0.1753512  0.23039445  0.1240689
##   2       10      0.1710070  0.23340045  0.1232549
##   2       11      0.1705569  0.24216456  0.1238223
##   2       12      0.1698596  0.25267562  0.1234112
##   2       13      0.1705306  0.26599056  0.1217326
##   2       14      0.1668873  0.27795244  0.1201179
##   2       15      0.1629996  0.27932408  0.1186262
##   2       16      0.1649803  0.27834961  0.1192776
##   3        2      0.1717146  0.07687192  0.1327396
##   3        3      0.1710907  0.11855999  0.1323018
##   3        4      0.1708568  0.12099882  0.1304242
##   3        5      0.1801026  0.14088972  0.1301957
##   3        6      0.1753530  0.17469066  0.1276165
##   3        7      0.1766699  0.18622435  0.1274743
##   3        8      0.1792331  0.18527146  0.1281607
##   3        9      0.1760602  0.18838691  0.1279815
##   3       10      0.1742058  0.20671223  0.1257052
##   3       11      0.1719224  0.22777048  0.1243740
##   3       12      0.1675070  0.24314780  0.1233248
##   3       13      0.1701304  0.24185791  0.1231818
##   3       14      0.1718904  0.24577026  0.1234260
##   3       15      0.1748271  0.24530541  0.1230490
##   3       16      0.1792917  0.25137575  0.1249525
##   4        2      0.1703362  0.09498338  0.1320784
##   4        3      0.1706670  0.09807895  0.1327578
##   4        4      0.1687000  0.14323980  0.1291412
##   4        5      0.1695571  0.17460434  0.1284005
##   4        6      0.1749273  0.18058123  0.1286959
##   4        7      0.1720064  0.20102678  0.1264686
##   4        8      0.1758592  0.19468879  0.1278018
##   4        9      0.1772033  0.19842150  0.1285754
##   4       10      0.1763105  0.21029881  0.1286742
##   4       11      0.1769835  0.21019550  0.1277605
##   4       12      0.1752002  0.22639449  0.1268597
##   4       13      0.1740817  0.23384357  0.1251285
##   4       14      0.1713288  0.24445926  0.1236945
##   4       15      0.1721859  0.24614500  0.1232094
##   4       16      0.1726797  0.24636638  0.1236299
##
```

```
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were nprune = 15 and degree = 2.
```

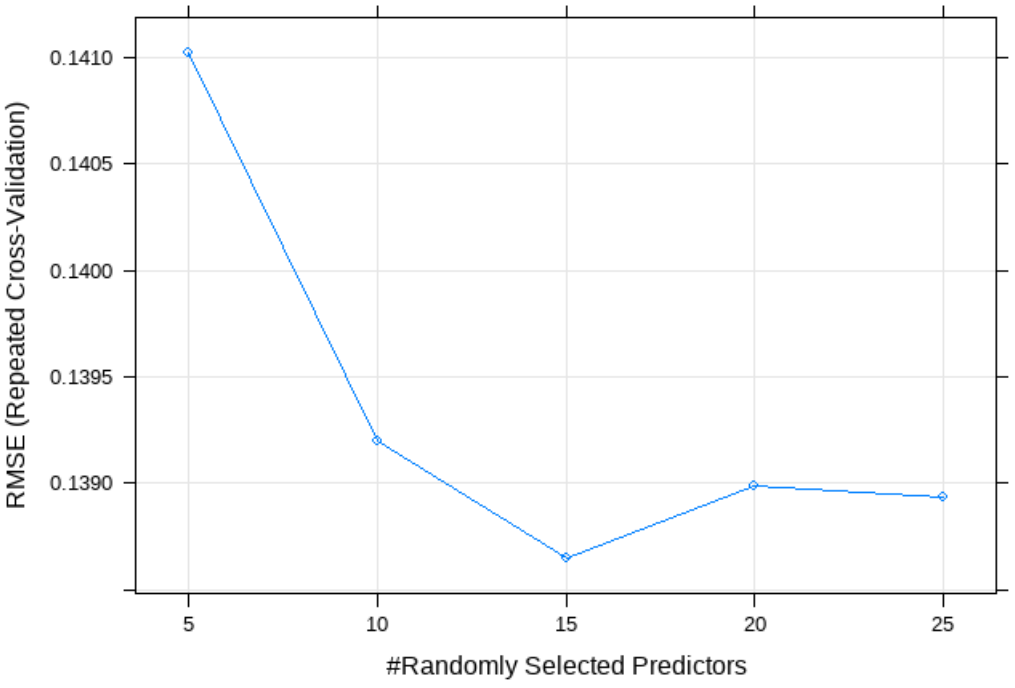
/// Support Vector Machines




```
## Support Vector Machines with Radial Basis Function Kernel
##
## 304 samples
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 274, 272, 272, 274, 275, 274, ...
## Resampling results across tuning parameters:
##
##   C       RMSE       Rsquared   MAE
##     0.25  0.1610282  0.1940700  0.1168427
##     0.50  0.1568823  0.2361124  0.1131864
##     1.00  0.1502603  0.2950607  0.1084254
##     2.00  0.1437626  0.3480926  0.1033758
##     4.00  0.1401129  0.3763227  0.1010702
##     8.00  0.1400015  0.3796044  0.1022203
##    16.00  0.1441912  0.3578199  0.1074816
##    32.00  0.1504019  0.3354010  0.1131090
##    64.00  0.1531805  0.3353311  0.1153474
##   128.00  0.1560646  0.3259696  0.1171056
##
```

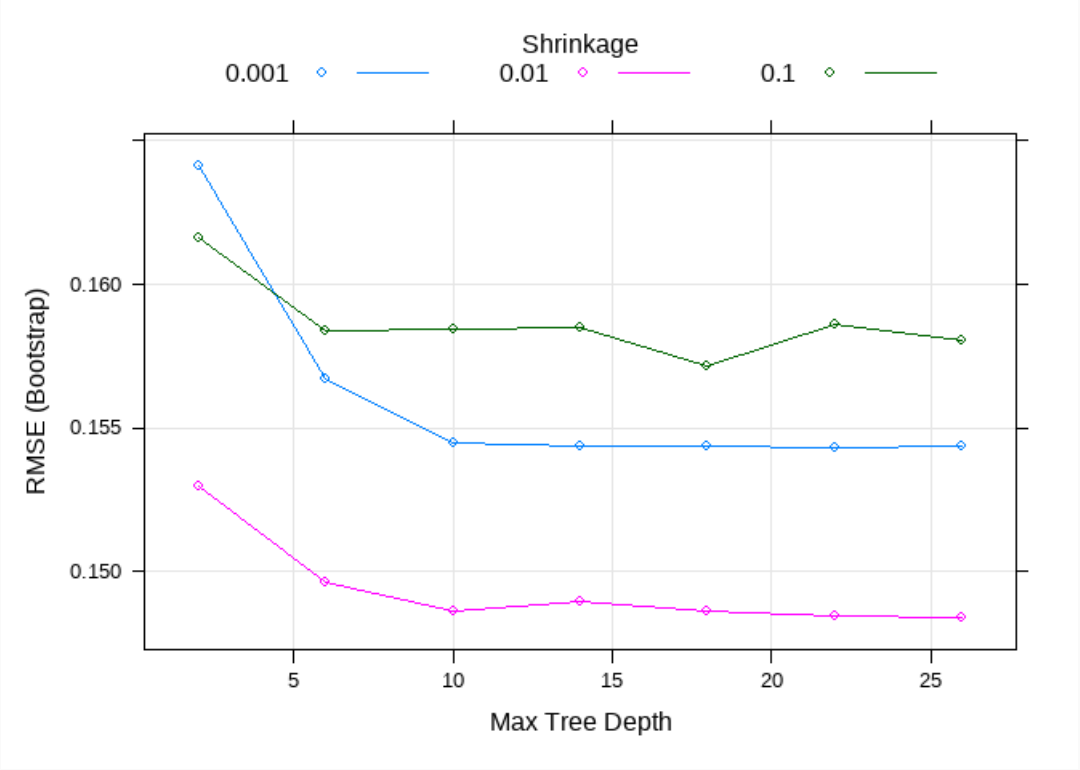
```
## Tuning parameter 'sigma' was held constant at a value of 0.02335804
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.02335804 and C = 8.
```

/// Random Forest




```
## Random Forest
##
## 304 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 274, 272, 272, 274, 275, 274, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE       Rsquared   MAE
##    5    0.1410206  0.4018532  0.10228889
##   10    0.1391973  0.4002163  0.09949196
##   15    0.1386487  0.3965829  0.09829419
##   20    0.1389875  0.3916452  0.09742373
##   25    0.1389379  0.3884283  0.09710201
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 15.
```

/// Gradient Boosting Machines




```
## Stochastic Gradient Boosting
##
## 304 samples
##  31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  RMSE       Rsquared   MAE
##   0.001       2                 0.1641125  0.1963360  0.1264797
##   0.001       6                 0.1567067  0.2771616  0.1193548
##   0.001      10                 0.1544554  0.3008623  0.1167014
##   0.001      14                 0.1543655  0.3015164  0.1165378
##   0.001      18                 0.1543489  0.3012733  0.1163894
##   0.001      22                 0.1542867  0.3027305  0.1164877
##   0.001      26                 0.1543503  0.3020659  0.1165086
##   0.010       2                 0.1529844  0.2686132  0.1133071
##   0.010       6                 0.1496585  0.3018668  0.1084808
##   0.010      10                 0.1486330  0.3103583  0.1071511
##   0.010      14                 0.1489864  0.3070315  0.1071087
##   0.010      18                 0.1486062  0.3105375  0.1069746
##   0.010      22                 0.1484654  0.3110123  0.1072851
##   0.010      26                 0.1483963  0.3123757  0.1068318
##   0.100       2                 0.1615922  0.2497526  0.1200204
##   0.100       6                 0.1583588  0.2584478  0.1151214
```

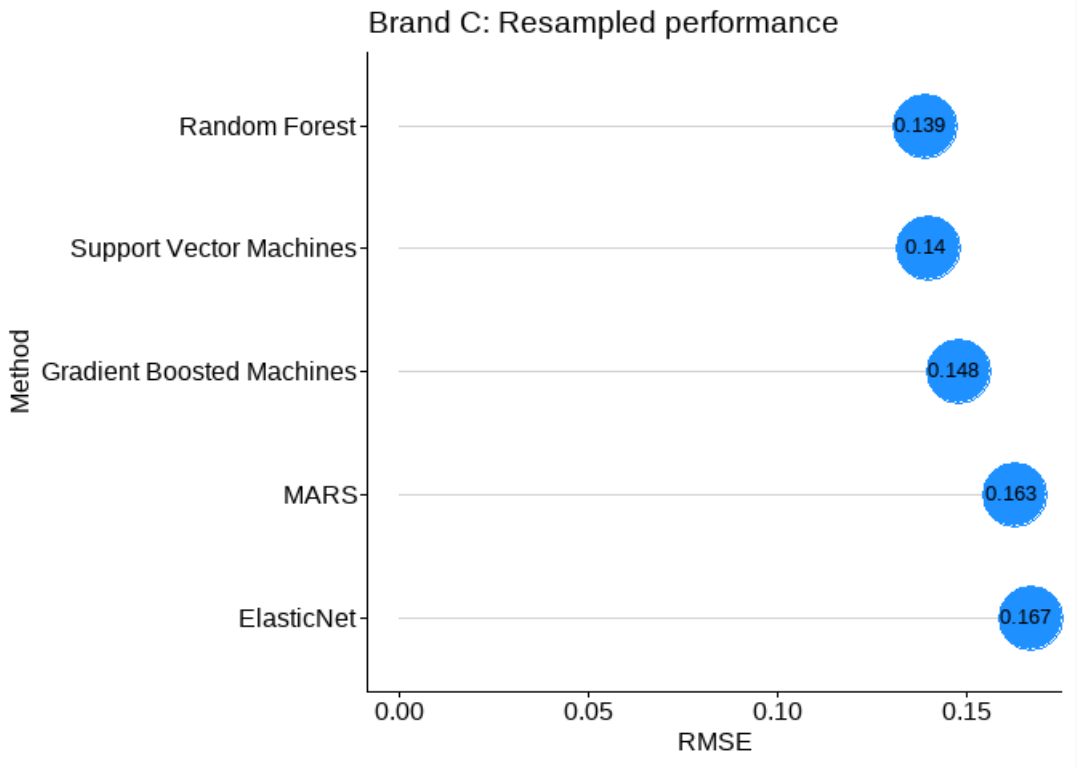
```
##   0.100      10                 0.1584473  0.2582170  0.1147787
##   0.100      14                 0.1584799  0.2622156  0.1148753
##   0.100      18                 0.1571537  0.2674185  0.1136614
##   0.100      22                 0.1585901  0.2585785  0.1152054
##   0.100      26                 0.1580184  0.2638790  0.1147484
##
## Tuning parameter 'n.trees' was held constant at a value of 1000
##
## Tuning parameter 'n.minobsinnode' was held constant at a value of 10
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 1000, interaction.depth
##  = 26, shrinkage = 0.01 and n.minobsinnode = 10.
```

// Brand C: Model Selection

The **Random Forest** model achieves the lowest resampled RMSE on brand **C** among the five models with score of **0.139**.

Brand C: Resampled performance

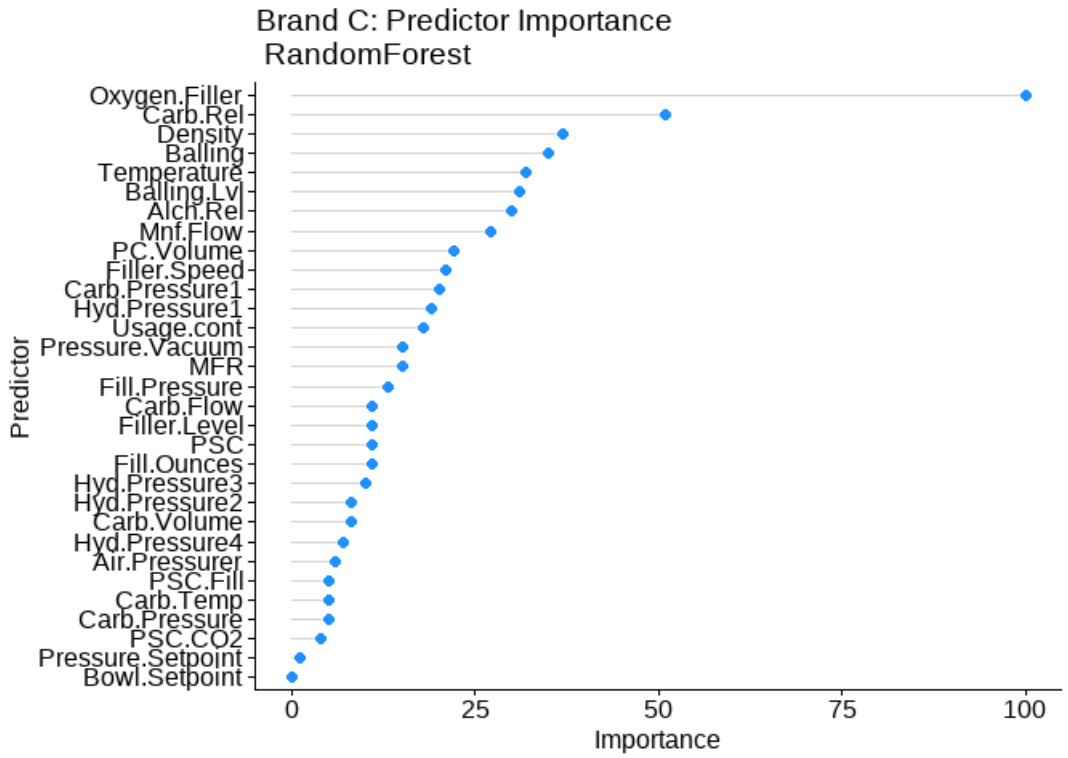
Method	RMSE
Random Forest	0.139
Support Vector Machines	0.140
Gradient Boosted Machines	0.148
MARS	0.163
ElasticNet	0.167



// Brand C: Model Variable Importance

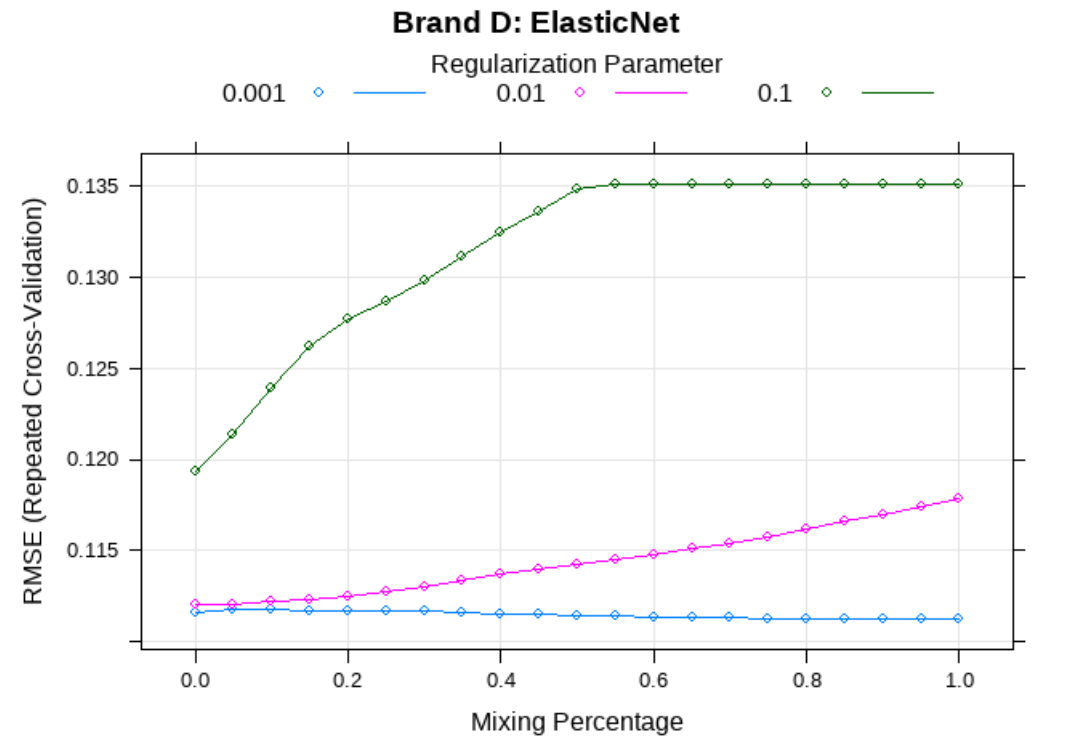
Brand C: Predictor Importance RandomForest

Predictor	Importance
Oxygen.Filler	100
Carb.Rel	51
Density	37
Balling	35
Temperature	32
Balling.Lvl	31
Alch.Rel	30
Mnf.Flow	27
PC.Volume	22
Filler.Speed	21
Carb.Pressure1	20
Hyd.Pressure1	19
Usage.cont	18
MFR	15
Pressure.Vacuum	15
Fill.Pressure	13
Fill.Ounces	11
PSC	11
Filler.Level	11
Carb.Flow	11
Hyd.Pressure3	10
Carb.Volume	8
Hyd.Pressure2	8
Hyd.Pressure4	7
Air.Pressurer	6
Carb.Pressure	5
Carb.Temp	5
PSC.Fill	5
PSC.CO2	4
Pressure.Setpoint	1
Bowl.Setpoint	0



// Brand D: Model Fitting

/// ElasticNet

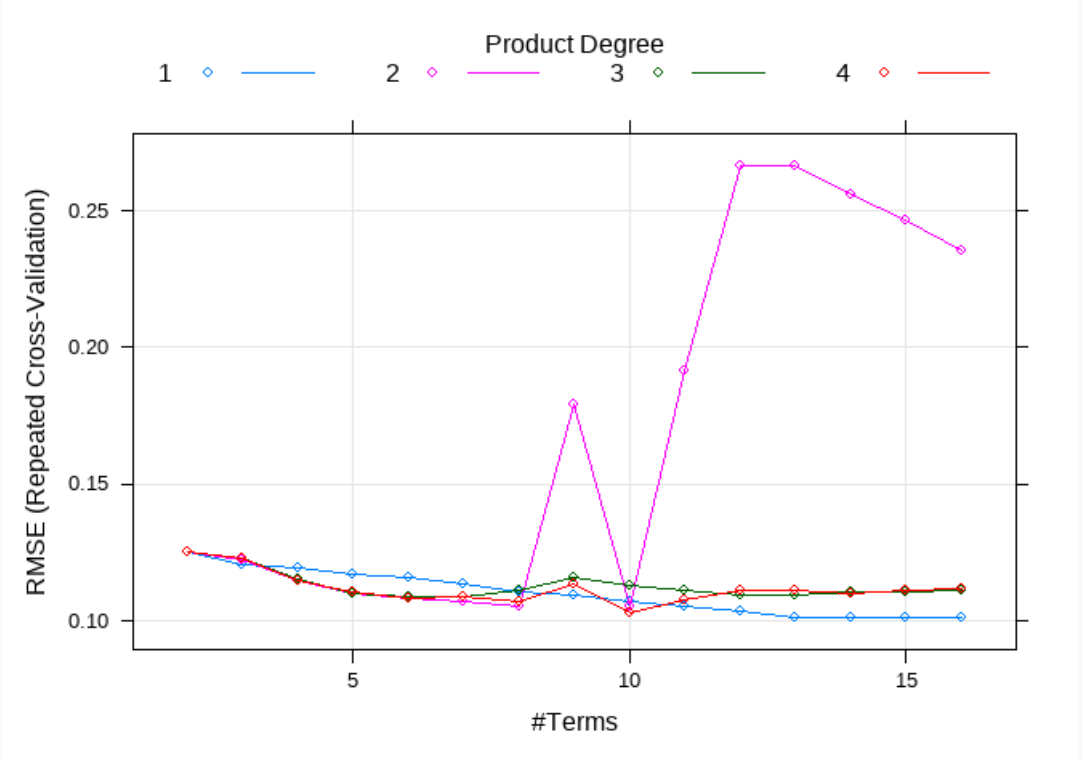



```
## glmnet
##
## 615 samples
```

```
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 553, 553, 553, 553, 554, 553, ...
## Resampling results across tuning parameters:
##
##   alpha  lambda  RMSE       Rsquared   MAE
##   0.00   0.001   0.1115903  0.3294118  0.08732405
##   0.00   0.010   0.1120615  0.3231983  0.08817757
##   0.00   0.100   0.1193450  0.2566299  0.09687483
##   0.05   0.001   0.1117816  0.3333903  0.08673835
##   0.05   0.010   0.1120568  0.3235726  0.08824305
##   0.05   0.100   0.1213380  0.2491277  0.09924697
##   0.10   0.001   0.1117416  0.3335068  0.08670507
##   0.10   0.010   0.1121609  0.3228632  0.08843069
##   0.10   0.100   0.1238971  0.2216320  0.10169799
##   0.15   0.001   0.1117061  0.3335814  0.08667499
##   0.15   0.010   0.1123011  0.3218494  0.08871429
##   0.15   0.100   0.1262628  0.1870275  0.10413101
##   0.20   0.001   0.1116827  0.3335596  0.08665274
##   0.20   0.010   0.1124860  0.3203600  0.08905067
##   0.20   0.100   0.1277203  0.1725962  0.10565207
##   0.25   0.001   0.1116571  0.3335720  0.08663570
##   0.25   0.010   0.1127043  0.3185933  0.08940633
##   0.25   0.100   0.1287120  0.1719946  0.10652082
##   0.30   0.001   0.1116320  0.3335829  0.08662337
##   0.30   0.010   0.1130094  0.3156209  0.08980289
##   0.30   0.100   0.1298558  0.1696831  0.10751979
##   0.35   0.001   0.1115949  0.3336949  0.08660910
##   0.35   0.010   0.1133755  0.3118316  0.09024099
##   0.35   0.100   0.1311561  0.1633928  0.10868223
##   0.40   0.001   0.1115406  0.3339901  0.08658746
##   0.40   0.010   0.1137080  0.3085108  0.09067988
##   0.40   0.100   0.1324739  0.1523446  0.10981883
##   0.45   0.001   0.1114854  0.3343156  0.08656559
##   0.45   0.010   0.1140018  0.3058569  0.09108813
##   0.45   0.100   0.1336384  0.1508827  0.11074819
##   0.50   0.001   0.1114321  0.3346156  0.08654198
##   0.50   0.010   0.1142605  0.3038479  0.09145969
##   0.50   0.100   0.1348271  0.1329855  0.11172731
##   0.55   0.001   0.1113875  0.3348356  0.08652466
##   0.55   0.010   0.1145156  0.3020561  0.09180930
##   0.55   0.100   0.1351046        NaN  0.11197099
##   0.60   0.001   0.1113510  0.3349556  0.08651781
##   0.60   0.010   0.1147797  0.3002196  0.09215319
##   0.60   0.100   0.1351046        NaN  0.11197099
##   0.65   0.001   0.1113206  0.3350427  0.08651177
##   0.65   0.010   0.1150794  0.2979658  0.09251448
##   0.65   0.100   0.1351046        NaN  0.11197099
##   0.70   0.001   0.1112947  0.3350906  0.08650809
```

```
##   0.70   0.010   0.1154106  0.2952800  0.09288960
##   0.70   0.100   0.1351046        NaN  0.11197099
##   0.75   0.001   0.1112716  0.3351087  0.08650669
##   0.75   0.010   0.1157668  0.2922167  0.09326982
##   0.75   0.100   0.1351046        NaN  0.11197099
##   0.80   0.001   0.1112534  0.3350657  0.08651167
##   0.80   0.010   0.1161572  0.2885666  0.09366859
##   0.80   0.100   0.1351046        NaN  0.11197099
##   0.85   0.001   0.1112392  0.3349831  0.08652202
##   0.85   0.010   0.1165709  0.2844433  0.09409113
##   0.85   0.100   0.1351046        NaN  0.11197099
##   0.90   0.001   0.1112320  0.3348256  0.08653683
##   0.90   0.010   0.1170004  0.2799952  0.09452949
##   0.90   0.100   0.1351046        NaN  0.11197099
##   0.95   0.001   0.1112295  0.3346233  0.08655419
##   0.95   0.010   0.1174326  0.2753906  0.09498174
##   0.95   0.100   0.1351046        NaN  0.11197099
##   1.00   0.001   0.1112347  0.3343360  0.08657926
##   1.00   0.010   0.1178573  0.2708526  0.09543194
##   1.00   0.100   0.1351046        NaN  0.11197099
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.95 and lambda = 0.001.
```

/// Multivariate Regression Adaptive Splines



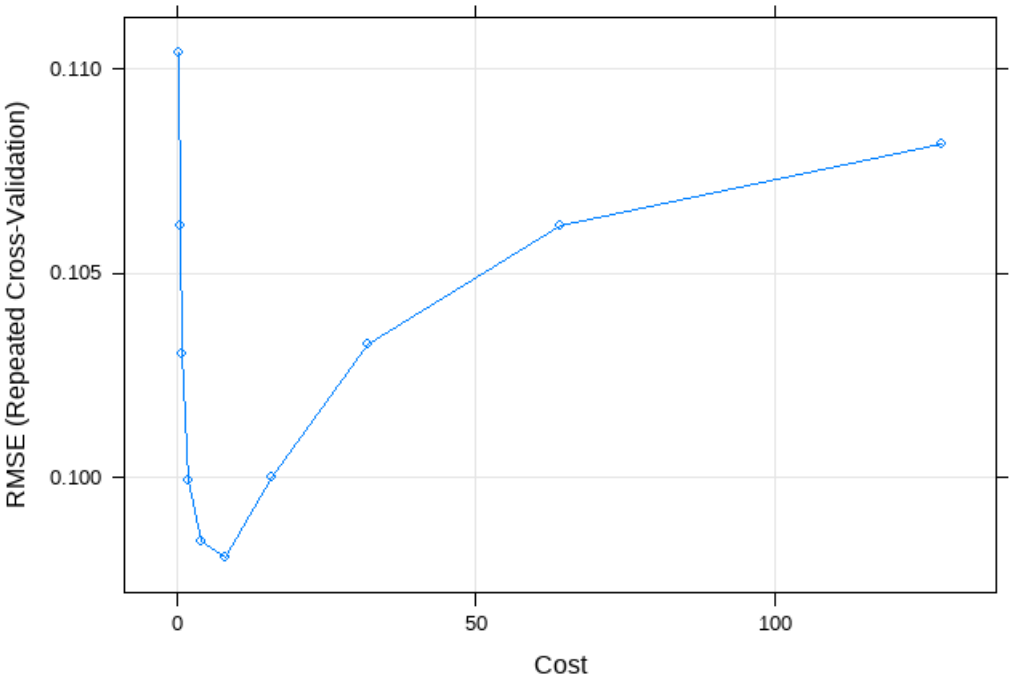
```
## Call: earth(x=data.frame[615,31], y=c(8.4,8.46,8.4,...), keepxy=TRUE, degree=1,
##             nprune=16)
##
##                           coefficients
## (Intercept)                  8.5641352
## h(Carb.Volume-5.55)          0.7721577
## h(4.8-Mnf.Flow)              0.0020432
## h(Mnf.Flow-4.8)              0.0008560
## h(Mnf.Flow-148)             -0.0041658
## h(121-Carb.Pressure1)       -0.0108457
## h(27.8733-Hyd.Pressure3)     0.0017559
## h(Hyd.Pressure3-27.8733)     0.0052085
## h(Filler.Speed-3900)         0.0004821
## h(Temperature-65)           -0.0441340
## h(Temperature-67.6)          0.0622956
## h(Usage.cont-21.22)         -0.0211761
## h(Density-1.54)             -0.5783653
## h(Pressure.Vacuum- -4.8)    -0.1704561
## h(Alch.Rel-7.74)             1.0964995
## h(5.56-Carb.Rel)            -0.5701609
##
## Selected 16 of 25 terms, and 11 of 31 predictors (nprune=16)
## Termination condition: RSq changed by less than 0.001 at 25 terms
## Importance: Mnf.Flow, Hyd.Pressure3, Pressure.Vacuum, Density, ...
## Number of terms at each degree of interaction: 1 15 (additive model)
## GCV 0.009284775    RSS 5.148989    GRSq 0.494088    RSq 0.5423179
```

```
## Multivariate Adaptive Regression Spline
##
## 615 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 553, 553, 553, 553, 554, 553, ...
## Resampling results across tuning parameters:
##
##   degree  nprune  RMSE       Rsquared   MAE
##   1        2      0.1249124  0.1609036  0.10224952
##   1        3      0.1206357  0.2220251  0.09688533
##   1        4      0.1195309  0.2322274  0.09687575
##   1        5      0.1171851  0.2806501  0.09335528
##   1        6      0.1158462  0.2982097  0.09128701
##   1        7      0.1131452  0.3146014  0.08989025
##   1        8      0.1103772  0.3501718  0.08784125
##   1        9      0.1090663  0.3683139  0.08612900
##   1       10      0.1068771  0.3949559  0.08453183
##   1       11      0.1054386  0.4119406  0.08289193
##   1       12      0.1034379  0.4301960  0.08127596
```

```
##   1       13      0.1009720  0.4546481  0.07982293
##   1       14      0.1009950  0.4588121  0.07921744
##   1       15      0.1010532  0.4578943  0.07912123
##   1       16      0.1009437  0.4619783  0.07892406
##   2        2      0.1250651  0.1585310  0.10234358
##   2        3      0.1222160  0.1961418  0.09958855
##   2        4      0.1143286  0.2930980  0.09189462
##   2        5      0.1100995  0.3471902  0.08692997
##   2        6      0.1080573  0.3830445  0.08462648
##   2        7      0.1072078  0.3964375  0.08348744
##   2        8      0.1050010  0.4178398  0.08155589
##   2        9      0.1789180  0.4181661  0.09069124
##   2       10      0.1052748  0.4217325  0.08125703
##   2       11      0.1913845  0.4223733  0.09169271
##   2       12      0.2664779  0.4231586  0.10064160
##   2       13      0.2665405  0.4248593  0.10028378
##   2       14      0.2560138  0.4230341  0.09904650
##   2       15      0.2465740  0.4350192  0.09684689
##   2       16      0.2351010  0.4397556  0.09527188
##   3        2      0.1250651  0.1585310  0.10234358
##   3        3      0.1228350  0.1881046  0.09971620
##   3        4      0.1151189  0.2866851  0.09315040
##   3        5      0.1099317  0.3500955  0.08709111
##   3        6      0.1086671  0.3772154  0.08513679
##   3        7      0.1088408  0.3789170  0.08448433
##   3        8      0.1109979  0.3866990  0.08350817
##   3        9      0.1154992  0.3989419  0.08333640
##   3       10      0.1128264  0.4226276  0.08184208
##   3       11      0.1111445  0.4406978  0.08049709
##   3       12      0.1095493  0.4527582  0.07953868
##   3       13      0.1095522  0.4516582  0.07943373
##   3       14      0.1102174  0.4604435  0.07876090
##   3       15      0.1103401  0.4651619  0.07865355
##   3       16      0.1107826  0.4672854  0.07831178
##   4        2      0.1250651  0.1585310  0.10234358
##   4        3      0.1227461  0.1897329  0.09951833
##   4        4      0.1145144  0.2892129  0.09282110
##   4        5      0.1102999  0.3439634  0.08765870
##   4        6      0.1083669  0.3719773  0.08532631
##   4        7      0.1084744  0.3777939  0.08473202
##   4        8      0.1069586  0.3967337  0.08327693
##   4        9      0.1136262  0.4117515  0.08351013
##   4       10      0.1028348  0.4430327  0.08002185
##   4       11      0.1076966  0.4338218  0.08046099
##   4       12      0.1108073  0.4378707  0.08043968
##   4       13      0.1108013  0.4434537  0.07983260
##   4       14      0.1098136  0.4510704  0.07908988
##   4       15      0.1109793  0.4501026  0.07880101
##   4       16      0.1117411  0.4508095  0.07869408
##
```

```
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were nprune = 16 and degree = 1.
```

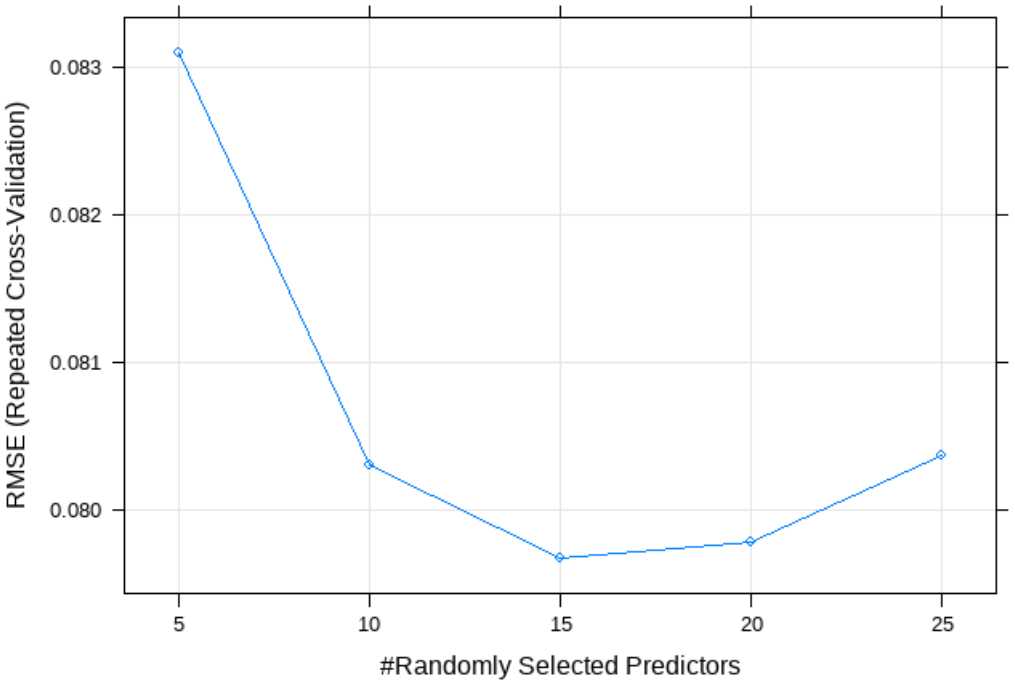
/// Support Vector Machines




```
## Support Vector Machines with Radial Basis Function Kernel
##
## 615 samples
##  31 predictor
##
## Pre-processing: centered (31), scaled (31)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 553, 553, 553, 553, 554, 553, ...
## Resampling results across tuning parameters:
##
##   C       RMSE        Rsquared   MAE
##     0.25  0.11039149  0.3556486  0.08448730
##     0.50  0.10616316  0.3972302  0.08020463
##     1.00  0.10300807  0.4288027  0.07733941
##     2.00  0.09992685  0.4607117  0.07429069
##     4.00  0.09841642  0.4767250  0.07242249
##     8.00  0.09803332  0.4851687  0.07220354
##    16.00  0.09999244  0.4742394  0.07392869
##    32.00  0.10325104  0.4562194  0.07672267
##    64.00  0.10616245  0.4428304  0.07925787
##   128.00  0.10817059  0.4336633  0.08096130
##
```

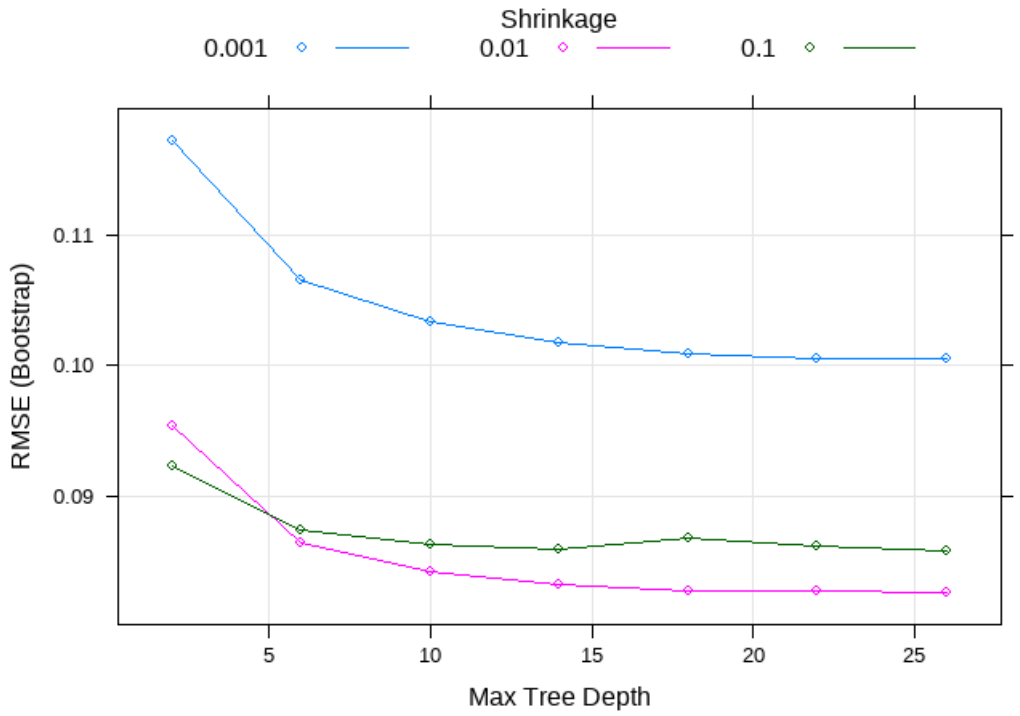
```
## Tuning parameter 'sigma' was held constant at a value of 0.02440234
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.02440234 and C = 8.
```

/// Random Forest




```
## Random Forest
##
## 615 samples
##  31 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 553, 553, 553, 553, 554, 553, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE        Rsquared   MAE
##    5    0.08309476  0.6783053  0.06353821
##   10    0.08029764  0.6848655  0.06109008
##   15    0.07967003  0.6816611  0.06049801
##   20    0.07978029  0.6755279  0.06042757
##   25    0.08036626  0.6656505  0.06065618
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 15.
```

/// Gradient Boosting Machines



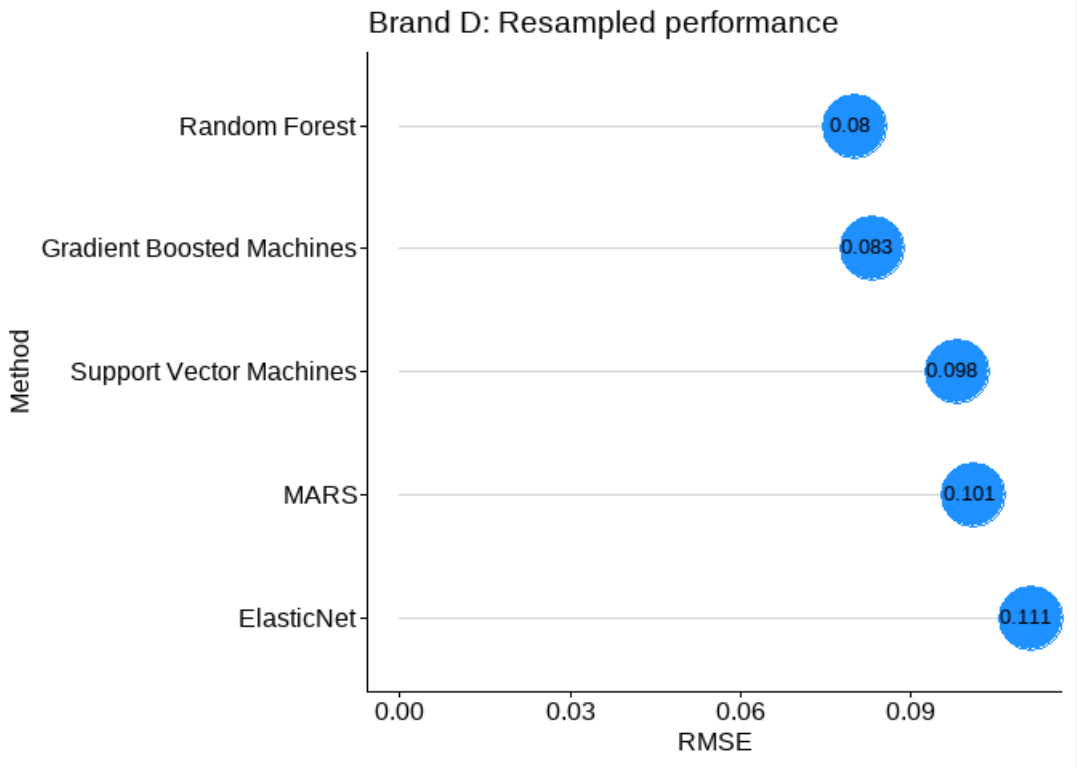

```
## Stochastic Gradient Boosting
##
## 615 samples
##  31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 615, 615, 615, 615, 615, 615, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  RMSE        Rsquared   MAE
##   0.001       2                 0.11732628  0.3710639  0.09616381
##   0.001       6                 0.10663463  0.5108503  0.08645115
##   0.001      10                 0.10338208  0.5426296  0.08333036
##   0.001      14                 0.10180350  0.5584888  0.08175284
##   0.001      18                 0.10095450  0.5657087  0.08089956
##   0.001      22                 0.10059878  0.5697051  0.08057121
##   0.001      26                 0.10057977  0.5697333  0.08054431
##   0.010       2                 0.09540718  0.5196177  0.07553338
##   0.010       6                 0.08635133  0.6019707  0.06650133
##   0.010      10                 0.08416954  0.6215710  0.06444668
##   0.010      14                 0.08316666  0.6308148  0.06333770
##   0.010      18                 0.08269362  0.6353771  0.06296925
##   0.010      22                 0.08272706  0.6347887  0.06286339
##   0.010      26                 0.08253585  0.6369264  0.06266448
##   0.100       2                 0.09223764  0.5434448  0.07150763
##   0.100       6                 0.08739584  0.5868367  0.06688458
```

```
##   0.100      10                 0.08626366  0.5974223  0.06592006
##   0.100      14                 0.08586126  0.6010813  0.06538293
##   0.100      18                 0.08673736  0.5934315  0.06599329
##   0.100      22                 0.08608257  0.5993577  0.06553960
##   0.100      26                 0.08573577  0.6030881  0.06533080
##
## Tuning parameter 'n.trees' was held constant at a value of 1000
##
## Tuning parameter 'n.minobsinnode' was held constant at a value of 10
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 1000, interaction.depth
##  = 26, shrinkage = 0.01 and n.minobsinnode = 10.
```

// Brand D: Model Selection

The **Random Forest** model achieves the lowest resampled RMSE on brand **D** among the five models with score of **0.079**.

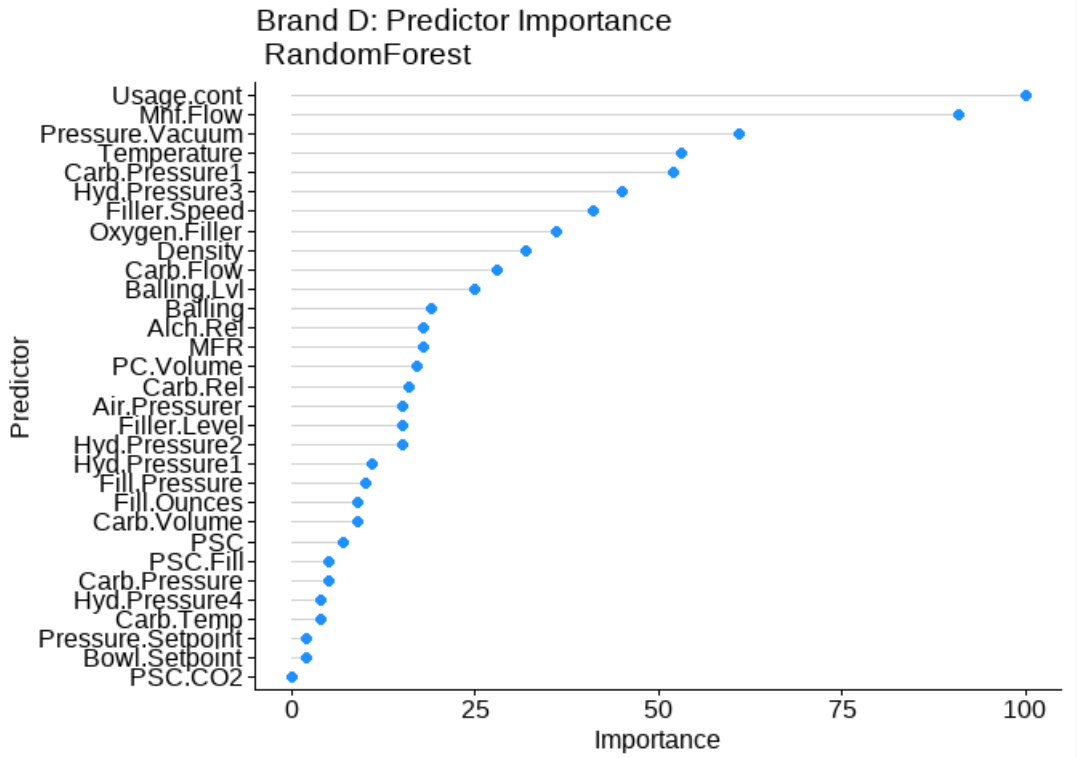
Brand D: Resampled performance	
Method	**RMSE**
Random Forest	0.080
Gradient Boosted Machines	0.083
Support Vector Machines	0.098
MARS	0.101
ElasticNet	0.111



// Brand D: Model Variable Importance

Brand D: Predictor Importance RandomForest

Predictor	Importance
Usage.cont	100
Mnf.Flow	91
Pressure.Vacuum	61
Temperature	53
Carb.Pressure1	52
Hyd.Pressure3	45
Filler.Speed	41
Oxygen.Filler	36
Density	32
Carb.Flow	28
Balling.Lvl	25
Balling	19
MFR	18
Alch.Rel	18
PC.Volume	17
Carb.Rel	16
Hyd.Pressure2	15
Filler.Level	15
Air.Pressurer	15
Hyd.Pressure1	11
Fill.Pressure	10
Carb.Volume	9
Fill.Ounces	9
PSC	7
Carb.Pressure	5
PSC.Fill	5
Carb.Temp	4
Hyd.Pressure4	4
Bowl.Setpoint	2
Pressure.Setpoint	2
PSC.CO2	0

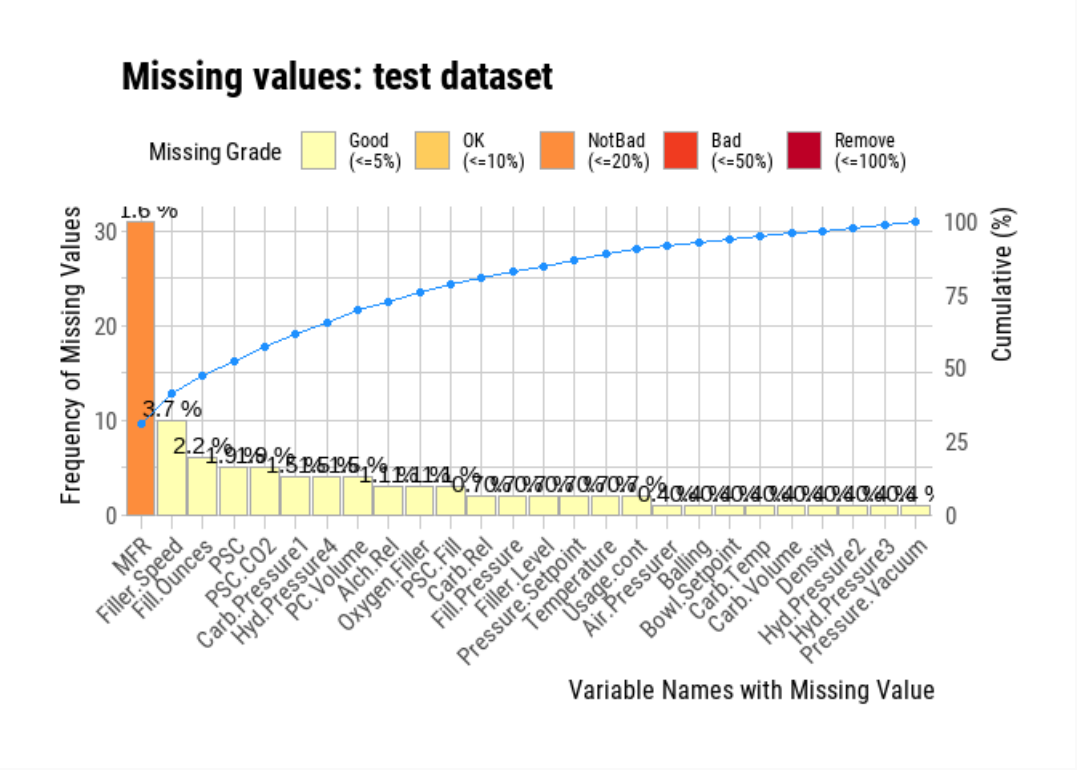


Model Prediction

The test data contains a fair number of missing values across several variables. As with the training data, we impute the missing values with the Bagged Trees method.

The final test data ready for prediction is summarized in the table below.

Predictions for all four brands are available for inspection and download.



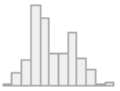
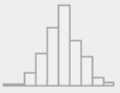
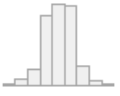
variables	missing_count	missing_percent
MFR	31	11.6
Filler.Speed	10	3.7
Fill.Ounces	6	2.2
PSC	5	1.9
PSC.CO2	5	1.9
PC.Volume	4	1.5
Carb.Pressure1	4	1.5
Hyd.Pressure4	4	1.5
PSC.Fill	3	1.1
Oxygen.Filler	3	1.1
Alch.Rel	3	1.1
Fill.Pressure	2	0.7
Filler.Level	2	0.7
Temperature	2	0.7
Usage.cont	2	0.7
Pressure.Setpoint	2	0.7
Carb.Rel	2	0.7
Carb.Volume	1	0.4
Carb.Temp	1	0.4
Hyd.Pressure2	1	0.4
Hyd.Pressure3	1	0.4
Density	1	0.4
Balling	1	0.4
Pressure.Vacuum	1	0.4
Bowl.Setpoint	1	0.4

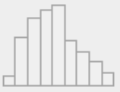
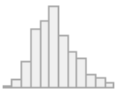
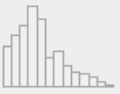
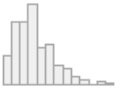
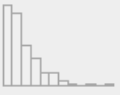
variables	missing_count	missing_percent
Air.Pressurer	1	0.4

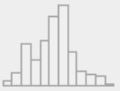
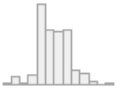
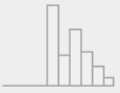
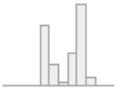
/// Data Frame Summary

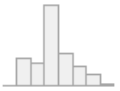
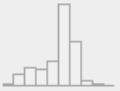
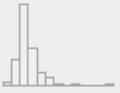
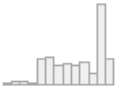
//// testData

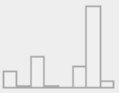
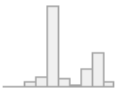
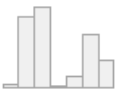
Dimensions: 267 x 32
Duplicates: 0

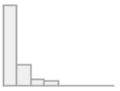
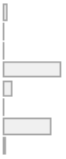
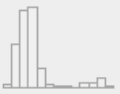
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Brand.Code [character]	1. (Empty string) 2. A 3. B 4. C 5. D	8 (3.0%) 35 (13.1%) 129 (48.3%) 31 (11.6%) 64 (24.0%)		267 (100.0%)	0 (0.0%)
2	Carb.Volume [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5.1 ≤ 5.3 ≤ 5.7 IQR (CV) : 0.2 (0)	72 distinct values		267 (100.0%)	0 (0.0%)
3	Fill.Ounces [numeric]	Mean (sd) : 24 (0.1) min ≤ med ≤ max: 23.7 ≤ 24 ≤ 24.2 IQR (CV) : 0.1 (0)	58 distinct values		267 (100.0%)	0 (0.0%)
4	PC.Volume [numeric]	Mean (sd) : 0.3 (0.1) min ≤ med ≤ max: 0.1 ≤ 0.3 ≤ 0.5 IQR (CV) : 0.1 (0.2)	187 distinct values		267 (100.0%)	0 (0.0%)

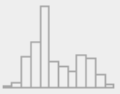
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
5	Carb.Pressure [numeric]	Mean (sd) : 68.3 (3.9) min ≤ med ≤ max: 60.2 ≤ 68 ≤ 77.6 IQR (CV) : 5.3 (0.1)	76 distinct values		267 (100.0%)	0 (0.0%)
6	Carb.Temp [numeric]	Mean (sd) : 141.3 (4.3) min ≤ med ≤ max: 130 ≤ 140.8 ≤ 154 IQR (CV) : 5.5 (0)	89 distinct values		267 (100.0%)	0 (0.0%)
7	PSC [numeric]	Mean (sd) : 0.1 (0.1) min ≤ med ≤ max: 0 ≤ 0.1 ≤ 0.2 IQR (CV) : 0.1 (0.6)	101 distinct values		267 (100.0%)	0 (0.0%)
8	PSC.Fill [numeric]	Mean (sd) : 0.2 (0.1) min ≤ med ≤ max: 0 ≤ 0.2 ≤ 0.6 IQR (CV) : 0.2 (0.6)	31 distinct values		267 (100.0%)	0 (0.0%)
9	PSC.CO2 [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.2 IQR (CV) : 0 (0.7)	16 distinct values		267 (100.0%)	0 (0.0%)
10	Mnf.Flow [numeric]	Mean (sd) : 21 (117.8)	114 distinct values		267 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		min ≤ med ≤ max: -100.2 ≤ 0.2 ≤ 220.4 IQR (CV) : 241.3 (5.6)				
11	Carb.Pressure1 [numeric]	Mean (sd) : 123 (4.4) min ≤ med ≤ max: 113 ≤ 123.4 ≤ 136 IQR (CV) : 5.3 (0)	91 distinct values		267 (100.0%)	0 (0.0%)
12	Fill.Pressure [numeric]	Mean (sd) : 48.1 (3.4) min ≤ med ≤ max: 37.8 ≤ 47.8 ≤ 60.2 IQR (CV) : 4.2 (0.1)	63 distinct values		267 (100.0%)	0 (0.0%)
13	Hyd.Pressure1 [numeric]	Mean (sd) : 12 (13.5) min ≤ med ≤ max: -50 ≤ 10.4 ≤ 50 IQR (CV) : 20.4 (1.1)	115 distinct values		267 (100.0%)	0 (0.0%)
14	Hyd.Pressure2 [numeric]	Mean (sd) : 20 (17.2) min ≤ med ≤ max: -50 ≤ 26.8 ≤ 61.4 IQR (CV) : 34.8 (0.9)	96 distinct values		267 (100.0%)	0 (0.0%)
15	Hyd.Pressure3 [numeric]	Mean (sd) : 19.5 (16.6)	90 distinct values		267 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		min ≤ med ≤ max: -50 ≤ 27.6 ≤ 49.2 IQR (CV) : 33 (0.8)				
16	Hyd.Pressure4 [numeric]	Mean (sd) : 97.9 (13.9) min ≤ med ≤ max: 68 ≤ 98 ≤ 140 IQR (CV) : 14 (0.1)	38 distinct values		267 (100.0%)	0 (0.0%)
17	Filler.Level [numeric]	Mean (sd) : 110.3 (15.4) min ≤ med ≤ max: 69.2 ≤ 118.4 ≤ 153.2 IQR (CV) : 19.6 (0.1)	109 distinct values		267 (100.0%)	0 (0.0%)
18	Filler.Speed [numeric]	Mean (sd) : 3581.2 (901.6) min ≤ med ≤ max: 1006 ≤ 3958.9 ≤ 4020 IQR (CV) : 189 (0.3)	89 distinct values		267 (100.0%)	0 (0.0%)
19	Temperature [numeric]	Mean (sd) : 66.2 (1.7) min ≤ med ≤ max: 63.8 ≤ 65.8 ≤ 75.4 IQR (CV) : 1.2 (0)	38 distinct values		267 (100.0%)	0 (0.0%)
20	Usage.cont [numeric]	Mean (sd) : 20.9 (3)	176 distinct values		267 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		min ≤ med ≤ max: 12.9 ≤ 21.4 ≤ 24.6 IQR (CV) : 5.6 (0.1)				
21	Carb.Flow [integer]	Mean (sd) : 2408.6 (1161.4) min ≤ med ≤ max: 0 ≤ 3038 ≤ 3858 IQR (CV) : 2132 (0.5)	178 distinct values		267 (100.0%)	0 (0.0%)
22	Density [numeric]	Mean (sd) : 1.2 (0.4) min ≤ med ≤ max: 0.1 ≤ 1 ≤ 1.8 IQR (CV) : 0.7 (0.3)	54 distinct values		267 (100.0%)	0 (0.0%)
23	MFR [numeric]	Mean (sd) : 675.2 (117.8) min ≤ med ≤ max: 15.6 ≤ 722.4 ≤ 784.8 IQR (CV) : 39.4 (0.2)	169 distinct values		267 (100.0%)	0 (0.0%)
24	Balling [numeric]	Mean (sd) : 2.2 (0.9) min ≤ med ≤ max: 0.9 ≤ 1.6 ≤ 3.8 IQR (CV) : 1.7 (0.4)	83 distinct values		267 (100.0%)	0 (0.0%)
25	Pressure.Vacuum [numeric]	Mean (sd) : -5.2 (0.6)	16 distinct values		267 (100.0%)	0 (0.0%)

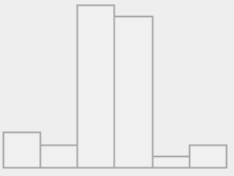
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		min ≤ med ≤ max: -6.4 ≤ -5.2 ≤ -3.6 IQR (CV) : 0.8 (-0.1)				
26	Oxygen.Filler [numeric]	Mean (sd) : 0 (0) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.4 IQR (CV) : 0 (1.1)	152 distinct values		267 (100.0%)	0 (0.0%)
27	Bowl.Setpoint [numeric]	Mean (sd) : 109.6 (15) min ≤ med ≤ max: 70 ≤ 120 ≤ 130 IQR (CV) : 20 (0.1)	70.00 : 9 (3.4%) 80.00 : 15 (5.6%) 90.00 : 35 (13.1%) 100.00 : 12 (4.5%) 106.06 ! : 1 (0.4%) 110.00 : 49 (18.4%) 120.00 : 139 (52.1%) 130.00 : 7 (2.6%) ! rounded		267 (100.0%)	0 (0.0%)
28	Pressure.Setpoint [numeric]	Mean (sd) : 47.7 (2.1) min ≤ med ≤ max: 44 ≤ 46 ≤ 52 IQR (CV) : 4 (0)	44.00 : 9 (3.4%) 45.20 : 1 (0.4%) 45.36 ! : 1 (0.4%) 46.00 : 128 (47.9%) 48.00 : 18 (6.7%) 48.78 ! : 1 (0.4%) 50.00 : 106 (39.7%) 52.00 : 3 (1.1%) ! rounded		267 (100.0%)	0 (0.0%)
29	Air.Pressurer [numeric]	Mean (sd) : 142.8 (1.2) min ≤ med ≤ max: 141.2 ≤ 142.6 ≤ 147.2 IQR (CV) : 0.6 (0)	25 distinct values		267 (100.0%)	0 (0.0%)
30	Alch.Rel [numeric]	Mean (sd) : 6.9 (0.5)	32 distinct values		267 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
		min ≤ med ≤ max: 6.4 ≤ 6.6 ≤ 7.8 IQR (CV) : 0.6 (0.1)				
31	Carb.Rel [numeric]	Mean (sd) : 5.4 (0.1) min ≤ med ≤ max: 5.2 ≤ 5.4 ≤ 5.7 IQR (CV) : 0.2 (0)	30 distinct values		267 (100.0%)	0 (0.0%)
32	Balling.Lvl [numeric]	Mean (sd) : 2.1 (0.9) min ≤ med ≤ max: 0 ≤ 1.5 ≤ 3.4 IQR (CV) : 1.7 (0.4)	53 distinct values		267 (100.0%)	0 (0.0%)

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

// Brand A

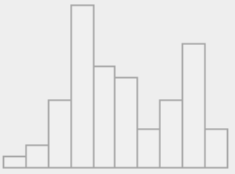
Test set prediction summary statistics for beverage brand A are provided below.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid
1	Prediction [numeric]	Mean (sd) : 8.5 (0.1) min ≤ med ≤ max: 8.2 ≤ 8.5 ≤ 8.8 IQR (CV) : 0.1 (0)	22 distinct values		35 (100.0%)

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

// Brand B

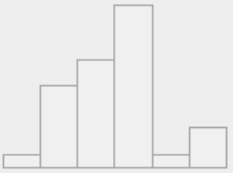
Test set prediction summary statistics for beverage brand B are provided below.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid
1	Prediction [numeric]	Mean (sd) : 8.6 (0.1) min ≤ med ≤ max: 8.3 ≤ 8.6 ≤ 8.8 IQR (CV) : 0.2 (0)	42 distinct values		129 (100.0%)

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

// Brand C

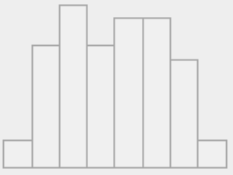
Test set prediction summary statistics for beverage brand C are provided below.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid
1	Prediction [numeric]	Mean (sd) : 8.4 (0.1) min ≤ med ≤ max: 8.1 ≤ 8.4 ≤ 8.7 IQR (CV) : 0.2 (0)	23 distinct values		31 (100.0%)

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10

// Brand D

Test set prediction summary statistics for beverage brand D are provided below.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid
1	Prediction [numeric]	Mean (sd) : 8.6 (0.1) min ≤ med ≤ max: 8.4 ≤ 8.6 ≤ 8.8 IQR (CV) : 0.1 (0)	29 distinct values		64 (100.0%)

Generated by summarytools 1.0.1 (R version 4.1.0)
2023-04-10