Project 2 - Non-Technical Report

Douglas Barley, Rachel Greenlee, Sean Connin

5/9/2022

Table of Contents

[Executive Summary 1](#_Toc103201463)

[The Issue 1](#_Toc103201464)

[Understanding manufacturing factors that impact pH 1](#_Toc103201465)

[Predictive Factors 2](#_Toc103201466)

[Predictive Model of pH 3](#_Toc103201467)

[Conclusions and Recommendations 5](#_Toc103201468)

[Next Steps 6](#_Toc103201469)

## Executive Summary

The FDA issued new regulations related to the pH levels in beverage manufacturing that require companies to understand and monitor manufacturing factors that influence pH levels in their products. The Data Science team at ABC Beverage was tasked to develop a model that will assist Manufacturing Operations to better understand and control pH levels in our products. The team’s results show that there are four discrete manufacturing processes that contribute to pH levels, and the formula for controlling pH is included in this report, along with next steps for each of the teams involved.

## The Issue

ABC Beverage has become aware of new regulations that require us to understand our manufacturing process in terms of how pH variance is affected. This includes the predictive factors that impact pH, as well as a clear understanding of our predictive model of pH.

## Understanding manufacturing factors that impact pH

Based on the metrics that our Manufacturing Operations teams regularly collect throughout the manufacturing process, our team of data scientists evaluated a set of machine learning models to predict pH. The models were selected due to their ability to handle complex data with little preprocessing requirements.

The following predictive models were trained and evaluated:

* Basic Decision Tree
* Random Forest
* Support Vector Machine (SVM)
* Neural Net (single layer, forward) (NNET)
* Multivariate Adaptive Regression Splines (MARS)

For those who may be interested in knowing more about the modeling process that was utilized, additional details are available in the Technical Report. Please contact the data science team members directly to request a copy of it.

The data science team focused its efforts around model optimization and selection using industry standards, and as a result they identified the manufacturing components that can help us as a company better understand and identify issues that impact pH levels.

## Predictive Factors

Our Data Science team used the RMSE (root mean squared error) metric to compare the models’performance and select the model with the lowest predictive error. A low RMSE indicates that a model is able to predict pH accurately, so according to the Estimate column in Table 1 the MARS model is clearly the most accurate predictor of pH.

**Table** **1**: Table 1. Selected Models Based on Training Fit

| Model | Metric | Estimate | CV-Fold |
| --- | --- | --- | --- |
| MARS | RMSE | 0.1398617 | Preprocessor1\_Model2 |
| Random Forest | RMSE | 0.5177498 | Preprocessor1\_Model17 |
| Decision Tree | RMSE | 0.6534581 | Preprocessor1\_Model09 |
| SVM | RMSE | 0.6799684 | Preprocessor1\_Model18 |
| NNET | RMSE | 0.7011618 | Preprocessor1\_Model12 |

This was confirmed with further tests that confirmed MARS as the model of choice. Table 2. shows the results of our model evaluation on test data. It’s clear, based on RMSE scores, that the MARS model outperformed the others. Based on further tests with additional data, the team is confident the model will generalize well for production purposes.

**Table** **2**: Table 2. Comparison of Model Fit on Test Data

| Model | Metric | Estimate |
| --- | --- | --- |
| MARS | rmse | 0.1398370 |
| Neural Network | rmse | 0.8170020 |
| Random Forest | rmse | 0.8347546 |
| SVM | rmse | 0.8579408 |
| Decision Tree | rmse | 0.9389812 |

## Predictive Model of pH

Using the MARS model, Table 3 shows the variables selected for the final model along with their respective coefficients.

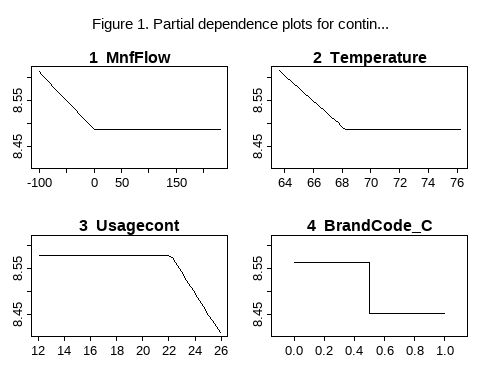
**Table** **3**: Table 3. Final MARS Model: Features &amp; Coefficients

| Model Components | Coefficient |
| --- | --- |
| (Intercept) | 8.465801028 |
| h(0.2-`Mnf Flow`) | 0.001256308 |
| BrandCode\_C | -0.112952249 |
| h(`Usage cont`-22.14) | -0.044961352 |
| h(68.2-Temperature) | 0.027763745 |

Therefore, the final model for manufacturing factors that influence pH levels is:

where *h* = the model hinge as shown below graphically in Figure 1.

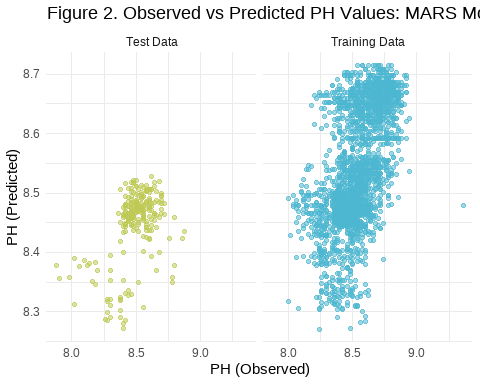
The hinge is the point where each horizontal line meets a non-horizontal line:



Each plot also reveals the hinge that is characteristic of MARS such that:

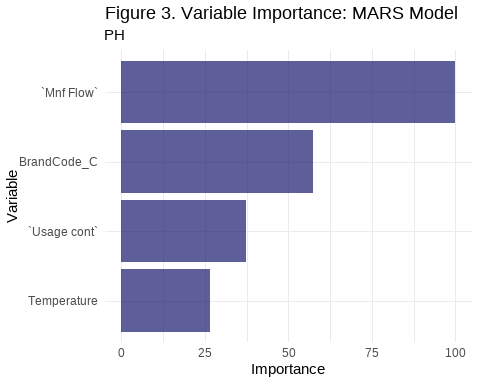
* Mnf Flow = 0 above 0.2
* Usage Cont = 0 below 22.14
* Temperature = 0 above 68.2
* BrandCode\_C = 0 for all values other than 0.45

Additionally, as seen in Figure 2. we can see how predictions of pH levels tend to converge with actual measurements around the 8.5 level when using the MARS model formula.



## Conclusions and Recommendations

The data science team concluded that there are four main manufacturing factors in play in terms of controlling pH in our products. They are shown below with the relative importance of each in Figure 3.



It is recommended that any investments to manipulate pH as part of the manufacturing process should focus on adjustments to these four manufacturing factors, and that a cost/benefit analysis be conducted before committing company resources for any modifications to the manufacturing processes.

## Next Steps

* The Manufacturing Operations team will work with the Finance team to complete a Cost/Benefit analysis focusing on control of the four components of the manufacturing process identified above by the Data Science team.
* The Legal and Government Affairs teams will coordinate reporting of all of our findings to the regulating authority.
* This committee will reconvene within the next 30 days to evaluate the further findings of the Manufacturing Operations and Finance teams.