ABC BEVERAGE CO pH ANALYSIS

Report Prepared By:
Banu Boopalan
Jonathan Burns
Molly Siebecker
Daria Dubovskaia
Marley Myrianthopoulos

Table of Contents

Executive Summary

The purpose of this analysis was to better understand the beverage manufacturing process at ABC Beverage Company by evaluating key factors that influence pH levels. Maintaining proper pH is essential for product quality, safety, and consistency.

Key Insights:

- **Data overview:** During the initial review, we discovered that important variables were incomplete, which required careful data cleaning and preparation before analysis.
- **Model selection:** After testing five advanced predictive models, one model emerged as the best-performing due to its superior accuracy and reliability.
- **Summary of findings:** The most influential factor in predicting pH levels was the manufacturing flow (a measure of the speed of the production line,) showing a negative correlation with pH levels.
- **Recommendations:** We recommend careful monitoring of manufacturing procedures to control pH, improved record keeping to prevent loss of data, and standardization of data cleaning procedures to ensure consistency in future reports.

These insights provide actionable data for optimizing production and maintaining beverage quality.

ABC BEVERAGES

Introduction

Beverage pH plays a crucial role in making products safe, tasty, and long-lasting. pH measures how acidic or basic a substance is on a scale from 0 to 14. Beverages like soda and coffee are acidic. Keeping the right pH level helps prevent harmful bacteria from growing and keeps the product fresh. This ensures consumers get a safe and consistent product every time.

Beverage companies must regularly check pH levels to meet industry standards and follow government regulations. Failure to do so can result in product recalls, legal troubles, and damages to the company's reputation.

The FDA provides helpful guidelines to support manufacturers in maintaining safe pH levels.

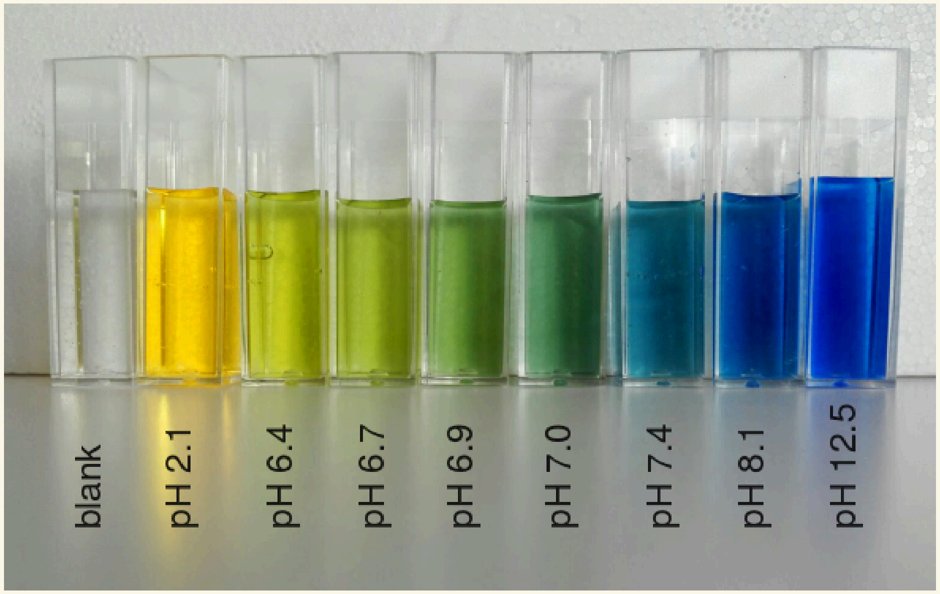

Figure 1

Data Overview

To ensure compliance with regulatory agencies, ABC Beverage Co. records information on 33 variables involved in the production process, including the pH of the resulting product. These variables include information on the nutrition information, carbonation process, bottling process, machinery operation, and quality control of 4 different brands of beverage. Each of these variables contained incomplete information, ranging from 0.03% of data missing to 8.25% of data missing, with approximately 1% of data missing overall. Missing values were imputed using industry-standard best practices for data management.

From this information, advanced statistical techniques and machine learning were used to create predictive models* that were able to approximate the pH of a product batch with a high degree of accuracy based on the other 32 variables ("predictor variables"). These predictive models were then analyzed using key performance metrics to identify the most accurate model. This model was then further studied to identify the manufacturing process elements that have the strongest influence on pH.

*Full list of machine learning models that were tuned:

- ➔ MARS Model (Multivariate Adaptive Regression Spline)
- ➔ SVM Model (Support Vector Machine)
- ➔ NNet Model (Neural Net)
- ➔ GBM Model (Gradient Boosting)
- ➔ RF Model (Random Forest)

Predictive Modeling

Five different machine learning models were tuned using the 32 predictor variables. Model accuracy was evaluated using key performance metrics including root-mean-squared-error (RMSE), percentage of variance explained (R^2), and mean-absolute-error (MAE). These are industry-standard metrics for Machine Learning model evaluation. Final model selection was determined according to these criteria. The results are shown in figure 2 below. The Random Forest model produced the lowest errors and explained the largest share of variance in resulting pH, so this model was selected for analysis.

	Resample Metrics			Test Metrics		
Model	Resample RMSE	Resample R^2	Resample MAE	Test RMSE	Test R^2	Test MAE
MARS	0.1269369	0.4591680	0.0960094	0.1220811	0.5103157	0.0921927
SVM	0.1193689	0.5235771	0.0877122	0.1178720	0.5449518	0.0844574
NNet	0.1145209	0.5585798	0.0855949	0.1107385	0.5978824	0.0826644
GBM	0.1101819	0.5955535	0.0830438	0.1054839	0.6444825	0.0792804
RF	0.1007928	0.6741332	0.0727108	0.0963002	0.7130397	0.0698903

Figure 2

ABC BEVERAGES

Summary of Findings

Analysis of the Random Forest model was performed to identify the predictor variables with the strongest influence on batch pH. The impact of these predictor variables is summarized in figure 3 below. By far, the factor with the strongest influence was Manufacturing Flow Rate (mnf_flow), a metric related to the speed of the production line. The remaining high-impact predictors are predominantly predictors associated with the manufacturing process, with two predictors instead being associated with the recipe (Brand Code and Balling Level).

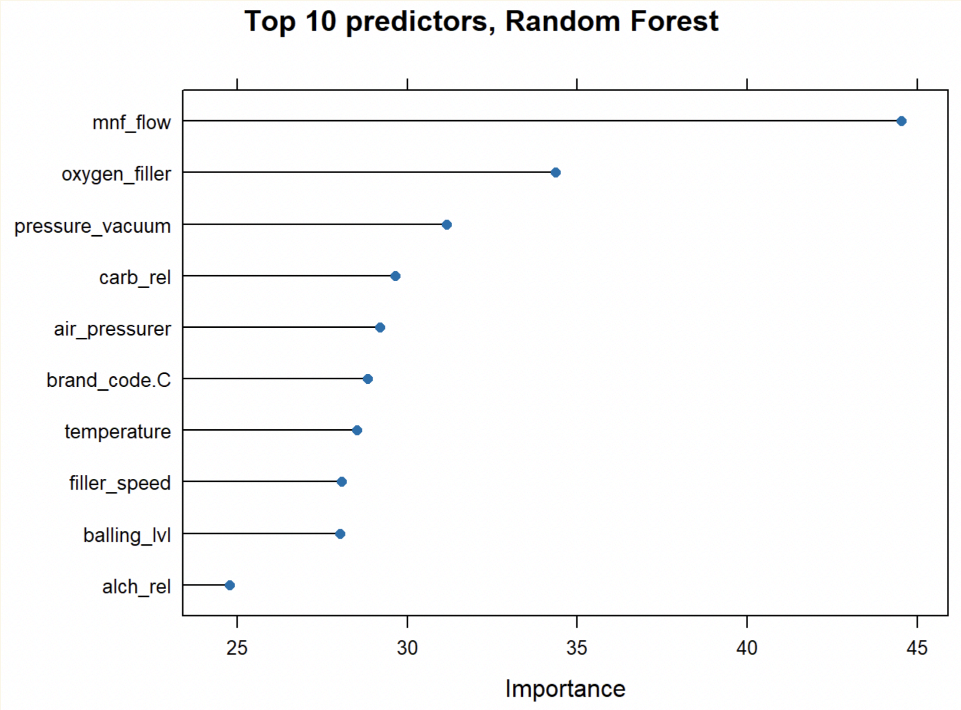

Figure 3

Recommendations

1) **Careful Monitoring of Manufacturing Procedures:** Our analysis indicates that the largest impact on beverage pH comes from variables associated with the production process rather than the recipe. We recommend that manufacturing procedures remain under careful scrutiny and strict quality control, as errors or variation in the manufacturing process may have a large impact on batch pH. Specifically:

    a) Take action to reduce variability in key predictors by using stronger monitoring and automated controls to stabilize flow rates and ensure consistent pH levels.

    b) Optimize key predictors by prioritizing improvements in high-impact factors through focused process improvements.

    c) Use real-time analytics to proactively monitor and alter important factors during manufacturing, lowering the possibility of deviations.

    d) Create dashboards and alerts that use these models to tell stakeholders about potential problems before they affect production.

Recommendations (Continued)

2) **Improved Record Keeping:** Some variables were not recorded for all batches, resulting in 1% of data missing overall. The percentage of data that was missing was higher for some variables, up to a maximum of 8.2% of data missing. This required corrective action before the data could be analyzed, and limited the accuracy of the results. We recommend that resources be put toward investigating and bolstering record-keeping procedures to enable more rapid and accurate analyses in the future.

3) **Regular Updates of Operational Data:** Data collection of these predictor variables and resulting batch pH should continue, so that we can regularly assess the predictive power of these models and take steps to improve them.

4) **Standardization of Data Cleaning Procedures for Analysis:** Our technical report outlines in detail the steps that were taken to prepare the data for analysis. We recommend that these steps be codified and standardized to ensure that the results of future analyses are consistent and properly contextualized.

ABC BEVERAGES