# PH Prediction in Beverage Manufacturing

Prepared for: ABC Beverage Production Team

Prepared by: Michael Robinson and Steven Gonzalez

Date: 18th, May 2025

## Introduction

The accurate prediction of pH levels is critical in beverage manufacturing, as it directly influences product quality, taste, and shelf life. Consistently maintaining the correct pH ensures customer satisfaction and compliance with regulatory standards. This report provides an overview of our approach to predicting pH levels using historical production data, highlighting key factors that impact pH and the effectiveness of different predictive models.

## Objective

The primary objective of this analysis is to develop a reliable predictive model for pH levels in beverage production, leveraging historical data to identify the most accurate approach for real-time quality control.

## Data Summary

The dataset used for this analysis includes 2,571 records with 33 features representing various production parameters, including:  
- Brand Code  
- Carb Volume  
- Fill Ounces  
- Carb Pressure  
- Temperature  
- Oxygen Filler  
- Bowl Setpoint  
- Pressure Setpoint

## Data Exploration

To ensure a comprehensive understanding of the data, initial exploration was conducted to identify key characteristics, data distributions. This included reviewing basic statistics, identifying missing values, and analyzing the relationships between production parameters and pH levels.

## Data Preparation

To prepare the data for modeling, the following steps were taken:  
1. Missing values were imputed using column means for numeric features and the most common category for categorical features.  
2. Categorical features, such as the Brand Code, were encoded numerically to enable machine learning analysis.  
3. Features were standardized to ensure consistent scaling across variables.

## Hyperparameter Tuning Setup

For the Decision Tree model, a search was performed to identify the optimal tree depth and minimum split size. This approach ensures the model captures the complex relationships between production parameters without overfitting.

## Assessing Models

Several machine learning models were tested to identify the most accurate approach for pH prediction, including:  
- Linear Regression  
- Decision Tree  
- Random Forest  
- Neural Network

## Model Performance Evaluation and Visualization

The performance of each model was evaluated based on Mean Squared Error (MSE) and R-squared (R²) values. Visualizations were generated to compare model accuracy and highlight the most effective predictors.

## Key Findings

After testing multiple models, the Decision Tree model provided the best overall performance based on its interpretability, simplicity, and ability to effectively capture the decision-making process in beverage production. Key advantages include:  
- Simple Structure: Easy to interpret and visualize, making it ideal for production settings.  
- Speed: Fast training and prediction, suitable for real-time quality control.  
- Robustness: Handles non-linear relationships without requiring extensive preprocessing.

## Conclusion

The Decision Tree model proved to be the most accurate in predicting pH levels in beverage manufacturing, making it the recommended choice for real-time quality control. This model effectively captures the complex interactions between production parameters, ensuring more consistent product quality.

## Next Steps

- Implement the Decision Tree model in the production environment for real-time pH prediction.  
- Continuously monitor model performance and retrain as necessary to adapt to changes in production processes.