



PREDICTIVE MODELS OF PH

Data Science Department, ABC Beverage

ABSTRACT

Impact of PH levels in our beverages with the manufacturing process

Abdelmalek Hajjam, Monu Chacko,
Paul Perez
DATA 624

Contents

Overview.....	2
Modelling.....	2
High Impact Variables	3
Conclusion.....	4

Predictive model of PH

Overview

In this report we highlight the impact of PH levels in our beverages with the manufacturing process. We use the past manufacturing data to predict the future and stay ahead in our market segment.

PH levels historically vary from 7.88 and 9.36. It has a mean value of 8.55. In order to determine the best results we use the industry standard indicators – Root Mean Square Error (RMSE) and R-Squared values.

Modelling

We used various modelling techniques and evaluated their performance. We used models like - Generalized Linear model (GLM), Partial Least Squares model (PSL), Ridge model, ENET model, MARS Model, SVM Model, KNN Model, Random Forest, Boosted trees, Cubist and Bagged Trees.

We used historical data and transformed them for a meaningful analysis. The data was then evaluated using all the models mentioned above to evaluate the best performance. We used *Root Mean Square Error (RMSE)* and *R-Squared values* to compare the model performance.

We found that the tree models performed better than the linear and non-linear models. The best model was Cubist. Cubist is a rule-based model that is an extension of Quinlan's M5 model tree. Random Forest model also had very good performance compared to all the other models we tuned and evaluated, so it might be a valid alternative to Cubist if for any reason an alternative is needed. The Cubist R-Squared value is 0.61, meaning that the model explains 61% of variability in the data. RMSE for Cubist is 0.11. The Random

Forest R-Squared value is 0.60, meaning that the model explains 60% of variability in the data. RMSE for Random Forest is 0.11.

High Impact Variables

Using the *Cubist* model we found the following variables to have highest impact on PH levels.

	Overall
Mnf.Flow	84.00
Balling	53.50
Pressure.Vacuum	49.00
Balling.Lvl	47.50
Alch.Rel	44.00
Oxygen.Filler	42.50
Density	40.50
Carb.Rel	34.00
Filler.Speed	31.00
Temperature	29.50
Hyd.Pressure3	29.00
Carb.Pressure1	28.00
Brand.Code	27.00
Bowl.Setpoint	25.50
Hyd.Pressure1	25.00
Usage.cont	24.00
MFR	24.00
Carb.Flow	22.50
Filler.Level	21.00
Hyd.Pressure2	20.00
Carb.Volume	12.50
Pressure.Setpoint	11.50
Fill.Pressure	11.50
PC.Volume	10.50
Hyd.Pressure4	8.50
Carb.Pressure	7.00
PSC.Fill	6.50
Air.Pressurer	6.00
Carb.Temp	6.00
PSC.CO2	3.50
PSC	2.00
Fill.Ounces	1.50

