ABC Beverages – pH Level Prediction Model – Non Technical Report:

DATA 624 - Project2 Team:

- Soumya Ghosh
- Jose Mawyin
- Randy Thompson

Company Background:

ABC Beverages is a US based premier family owned beverage company with a broad global portfolio of widely popular brands in wine and spirits categories. Company's portfolio includes brands imported from some of the premium wineries from South America, Europe and Australia along with locally produced brands from the wineries based out of California and breweries located in different parts of US.

Project Background:

Senior leadership team at ABC Beverage was made aware of the new upcoming FDA regulations around more prescriptive pH level guidelines requiring the company to better understand it's internal manufacturing processes, the predictive factors etc. This requires company's head of productions to leverage firm's inhouse Data Science team to be able to generate insights and reports capturing predictions from an accurate and efficient predictive model for pH level.

Definition of the Target Predictor Variable pH Level:

A definition of pH is the measurement of the acidity or alkalinity of a solution commonly measured on a scale of 0 to 14. pH 7 is considered neutral, with lower pH values being acidic and higher values being alkaline or caustic. pH is the most common of all analytical measurements in industrial processing and since it is a direct measure of acid content [H+], it clearly plays an important role in food processing.

Importance of pH Measurement in Food and Beverage Production:

Measuring pH in food and beverage production provides essential insights into its quality, safety and consistency. Variability of pH in food and beverage production can lead to critical differences in taste, freshness and shelf-life of a final product, making the pH value one of the parameters most frequently measured during quality inspection before release.

Yet pH is also influential during food and beverage preparation. For instance, enzymatic activity is strongly pH-dependent; during cheese-making, monitoring this parameter both improves yields and enhances the characteristic flavor. Controlling pH also regulates the growth of desired and undesirable microorganisms alike.

Meat, dairy products, pasta, wine and fruit juice fall within the broad range of foods and beverages for which control of pH is important. With widely divergent properties, these samples' requirements for assessment vary considerably.

Primary reasons for measuring pH in food processing include:

- To produce products with consistent well defined properties
- To efficiently produce products at optimal cost
- To avoid causing health problems to consumers
- To meet regulatory requirements

Typical Measurement Considerations for pH Value:

Due to the logarithmic nature of the measurement, even small changes in pH are significant. The difference between pH 6 and pH 5 represents a ten-fold increase in acid concentration; a change of just 0.3 represents a doubling of acid concentration. Variations of pH can impact flavor, consistency, and shelf-life.

Beverage Industry Specific Considerations:

Beer Category:

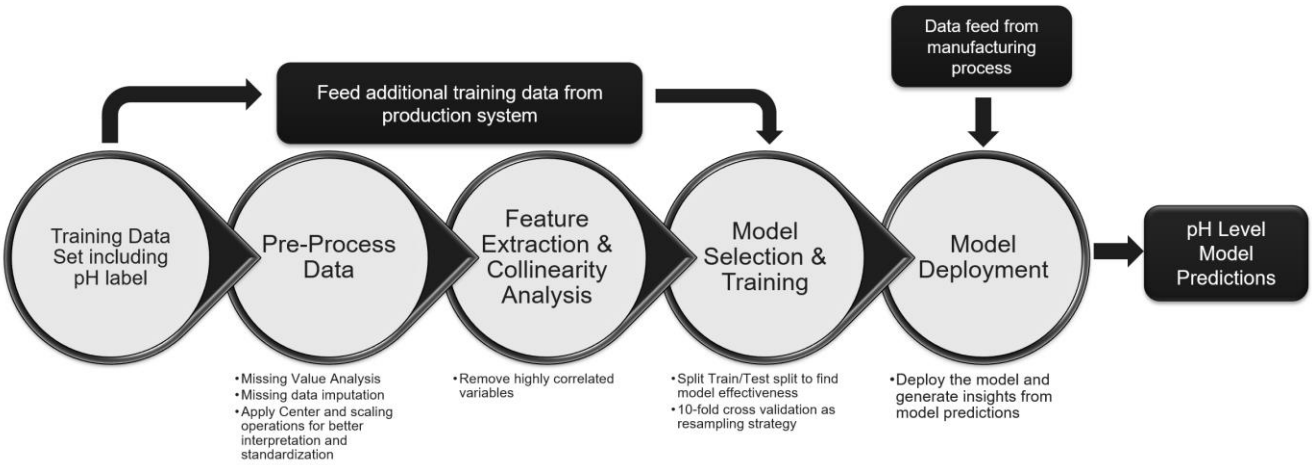
pH plays a crucial role in the production of beer. For example, the pH value of crushed malt is around 5.8 whereas its ideal value for protein decomposition is around 5.5. To ensure a consistent quality, the pH of brewed beer prior and after bottling is regularly monitored.

Wine Category:

Accurate and reliable pH measurement is required to maintain the characteristic texture and flavor of wine in the wine making process. pH of wine normally ranges from 2.8 to 3.8 with the pH influencing various stages of the process including fermentation and conservation. Because whites tend to have higher acidity, these will typically have lower pH than reds. With the pH exceeding 3.5, certain bacteria can attack the wine. However, taste of wine also depends largely on its pH value with acidic wines becoming dry.

Predictive ML Model Training Lifecycle:

Below is a high-level process flow to describe the pH level prediction model lifecycle –



Training Data Set Description:

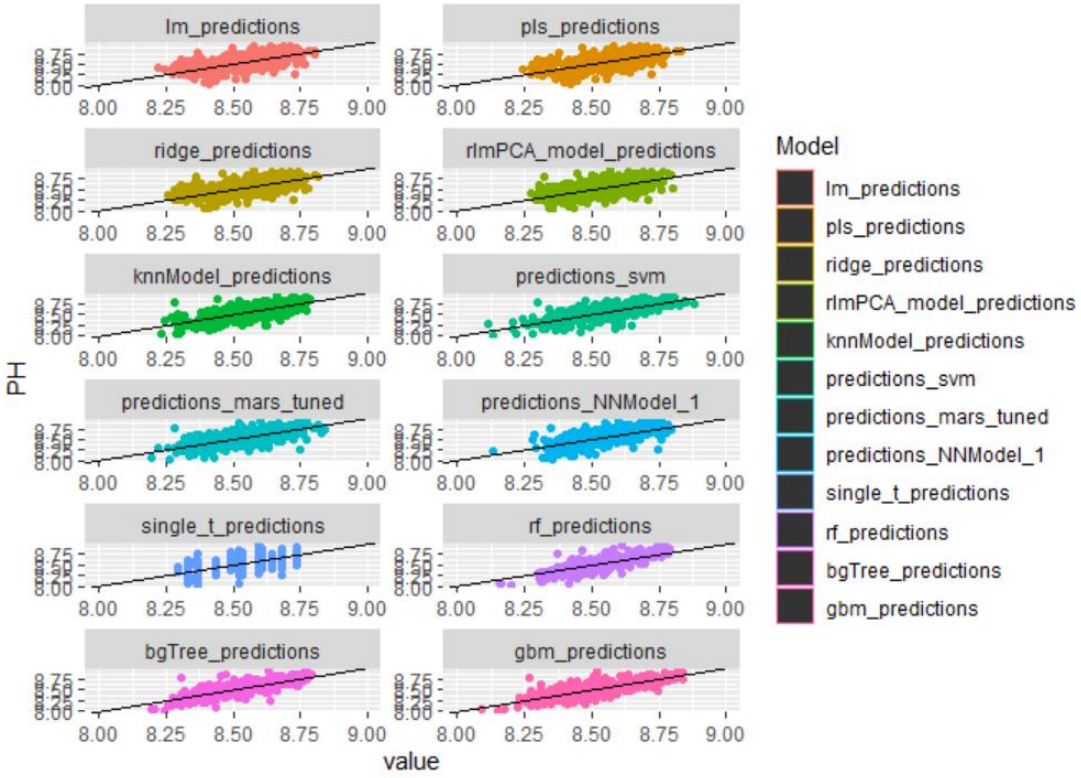
Data set used for model training purposes has following attributes -

```
##  [1] "Brand.Code"        "Carb.Volume"       "Fill.Ounces"
##  [4] "PC.Volume"         "Carb.Pressure"     "Carb.Temp"
##  [7] "PSC"               "PSC.Fill"          "PSC.CO2"
## [10] "Mnf.Flow"          "Carb.Pressure1"    "Fill.Pressure"
## [13] "Hyd.Pressure1"     "Hyd.Pressure2"     "Hyd.Pressure3"
## [16] "Hyd.Pressure4"     "Filler.Level"      "Filler.Speed"
## [19] "Temperature"       "Usage.cont"        "Carb.Flow"
## [22] "Density"           "MFR"               "Balling"
## [25] "Pressure.Vacuum"   "PH"                "Oxygen.Filler"
## [28] "Bowl.Setpoint"     "Pressure.Setpoint" "Air.Pressurer"
## [31] "Alch.Rel"          "Carb.Rel"          "Balling.Lvl"
```

Model Selection Summary:

Team has tried several linear, non-linear and tree-based models and used metrics like RMSE and R^2 to determine the best performing model to be able to predict pH level with highest degree of accuracy. Below is a summary of all such models –

Model#	Model Name	RMSE	R^2
1	Random Forest Grid	0.09215130	0.7294696
2	Bagged Tree	0.09583687	0.6893855
3	Gradient Boosted Machine	0.10113935	0.6482721
4	Support Vector Machine	0.10747655	0.6012682
5	Neural Network avNNet	0.11733185	0.5282685
6	MARS Tuned	0.11762364	0.5248174
7	k-Nearest Neighbors	0.11942558	0.5091825
8	Single Tree	0.12018700	0.5001648
9	Linear Regression	0.13127182	0.4080552
10	Ridge-regression	0.13254796	0.3955494
11	Partial Least Squares	0.13287128	0.3933813
12	Robust Linear Model	0.13673669	0.3558995



Conclusion:

Based on our analyses, the Random Forest model performed the best with an RMSE of 0.0937 and an R^2 of 0.729. Our recommendation is to productionize this model to leverage the pH level predictions generated by the model to ensure adherence to new regulatory guidelines as well as to ensure highest quality of the beverage products sold by the company to maximize the ROI for the investments made in this project.

References:

1) https://sperdirect.com/blogs/news/the-importance-of-ph-in-food-quality-and-production
2) https://www.mt.com/us/en/home/library/collections/lab-analytical-instruments/pH-in-food-and-beverage-production.html
3) https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=114&showFR=1