Summary Report of PH Predictive Model

Overview

Our team was tasked with understanding how the manufacturing process impacts PH levels in our beverages. We have used our historical manufacturing data to build a model and identified the variables which have the most impact on PH levels. PH levels in historical data vary between 7.88 and 9.36 with a mean value of 8.55. We will use Root Mean Square Error (RMSE) and R-Squared values to determine which model provides the best results.

Modeling

As a first step in our modeling, we pre-processed the historical data to reduce variability and remove outliers and filled in missing values. We have attempted to use a variety of models to identify the best one and used Root Mean Square Error (RMSE) and R-Squared values to make our final decision. Root Mean Square Error (RMSE) is a standard way to measure the error of a model and R-squared is a relative measure of fit. The Tree Models were performing better than non-linear models and linear models based on their R-squared and RMSE values. The best Model was Cubist, so that's what we used for our predictions. 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.

Highest Impact Variables

The following variables were determined to have the highest impact on PH levels in Cubist model.

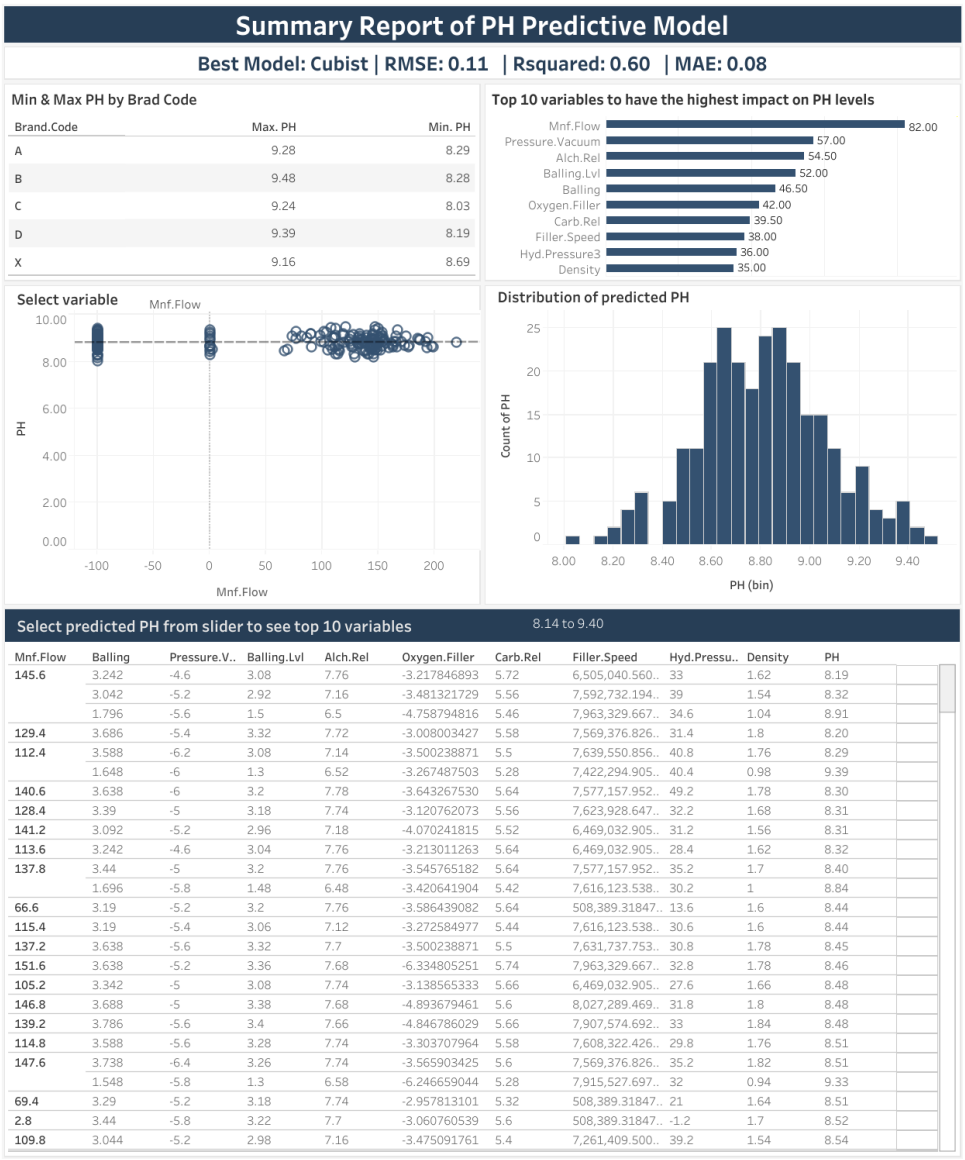
```
##                     Overall
## Mnf.Flow              82.0
## Pressure.Vacuum       57.0
## Alch.Rel              54.5
## Balling.Lvl           52.0
## Balling               46.5
## Oxygen.Filler         42.0
## Carb.Rel              39.5
## Filler.Speed          38.0
## Hyd.Pressure3         36.0
## Density               35.0
## Carb.Pressure1        32.5
## Temperature           30.5
## Filler.Level          29.0
## Hyd.Pressure2         28.5
## Brand.Code            22.5
## Carb.Flow             22.5
## Bowl.Setpoint         21.0
## MFR                   21.0
## Usage.cont            19.5
## Hyd.Pressure1         19.5
## Hyd.Pressure4         16.0
## Carb.Volume           13.5
## Carb.Pressure         11.5
## PC.Volume              9.5
## Fill.Pressure          9.0
## Carb.Temp              7.0
## Pressure.Setpoint      6.5
## PSC.Fill               6.0
## Air.Pressurer          4.5
## Fill.Ounces            2.5
## PSC                    1.5
## PSC.CO2                1.5
```

Tableau Dashboard

In order that both technical and nontechnical members of the management can make sense of our findings, we provided business friendly visualization of the summary in Tableau, right below. The dashboard can be accessed here:

https://public.tableau.com/views/SummaryReportofPHPredictiveModel/PHPrediction?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Image of the dashboard



Conclusions

Historical data allowed us to understand the manufacturing process, and which variables are most impactful predictive factors of PH values. We recommend using a Cubist model to predict PH values.