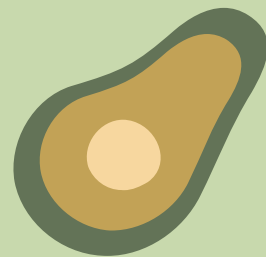
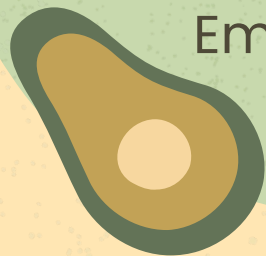


Avocado Classification and Price Prediction

Emily Ubbelohde



Avocado Advocates

A stylized illustration of an avocado, showing the green outer skin, the brown pit, and the light-colored flesh. It is positioned on the left side of the image, partially overlapping a light orange abstract shape.

01

Gather Sales from Grocers

Quantities of each individual PLU and Bags sold

02

Classify Avocados

Organic and Conventional

03

Predict Prices

For each type of avocado in each region

04

Partner with Local Farmers

To entice them to grow avocados

Data and Workflow



01

Gather Data from
Hass Avocado Board

02

Build and Tune
Classification Model

03

Use Classification Model
to populate Type

04

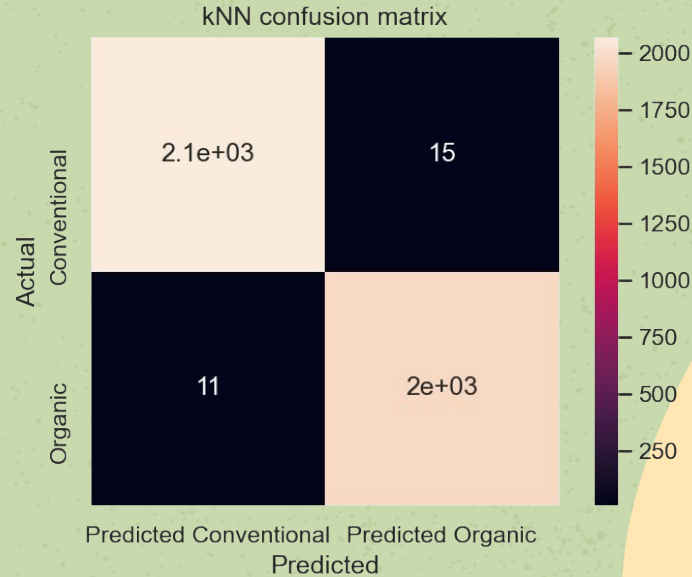
Build and Tune Price
Prediction Model

Avocado Classification Baseline

**KNN Confusion Matrix
for N = 1**

**Score on
Validation Set:**

99.36%



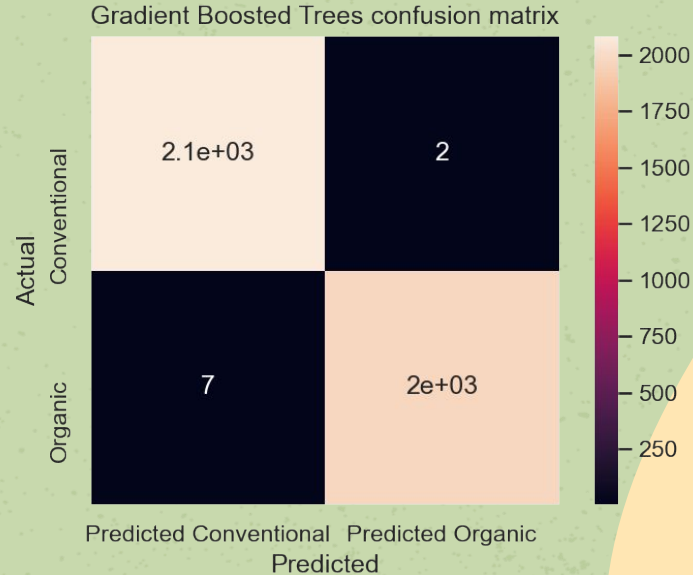
Avocado Classification

Final Model

Modeling using XGBoost

Score on
Validation Set:

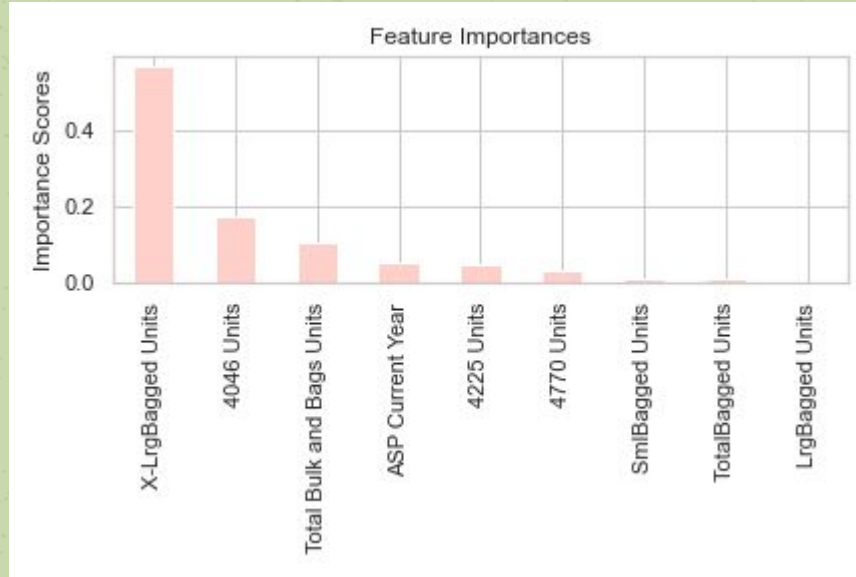
99.78%



Avocado Classification Using XGBoost

Most Important Features

- Qty. Sold XL Bags
- Qty. Sold 4046 Units
- Total Units Sold



Data and Workflow



01

Gather Data from
Hass Avocado Board

02

Build and Tune
Classification Model

03

Use Classification Model
to populate Type

04

Build and Tune Price
Prediction Model

Avocado Price Prediction



Random Forest

The best model was a Random Forest that contained all features

Most Important Features

- Type Prediction
- Qty. 4225 Units Sold
- Qty. 4046 Units Sold

Score on Validation Data

$R^2 = .848$



Conclusions

**XGBoost was the best
Classification Model
available**

**Random Forests was
the best model for
Price Prediction**

Most Important Features

- Qty. Sold XL Bags
- Qty. Sold 4046 Units

- Type Prediction
- Qty. 4225 Units Sold
- Qty. 4046 Units Sold

**Final Model Accuracy
on Hold Out Data**

$$R^2 = .927$$



Future Work

Continue to gather more data and reevaluate both models for accuracy

Add coordinates to location data

Build an interactive dashboard for price prediction

