VANDERBILT UNIVERSITY®
Data Science Institute

# Recommending Crops To Farmers

Case 2 - Team 1:
Donnie Sengstack, Preston Abraham, Meena Muthusubramanian, Yilin Ye

VANDERBILT UNIVERSITY®
Data Science Institute

# Elevator Pitch

- Help farmers determine the best crops to grow in their fields based on climate and soil conditions
- Help farmers make decisions based on the financial returns of growing different crops
- Develop a simple dashboard that farmers can use to make these decisions
    - Dashboard let's farmers know the exact dollar amount they could expect from growing each crop

VANDERBILT UNIVERSITY®
Data Science Institute

# Our Data

- Collected by the Indian Chamber of Food and Agriculture
- 2200 rows, 8 columns (last updated 5 months ago)
- Soil Data
  - pH
  - Nitrogen, Phosphorus, Potassium Levels
- Local Weather
  - Temperature
  - Humidity
  - Rainfall
- Crop Suitable
  - 22 different types of crops that are suitable given the conditions
  - Examples: Rice, Corn, Chickpeas, Kidney beans, etc.

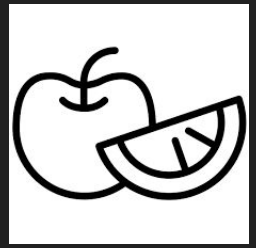
VANDERBILT UNIVERSITY®
Data Science Institute

# Additional Data

- 2020 Research Paper “Factors Affecting Yield of Crops” by Lilane and Mutengwa
- 22 Rows (one per crop), 7 Columns
- Helped us determine the financial and environmental benefits farmers would get from growing each crop by allowing us to explore:
  - Spacing
  - Yield
  - Cost
  - Profit

| Crop | Spacing | Yield(tonne/acre) [1 tonne=1000kg] | Cost of cultivation (in INR) | Profit (in INR) |
|---|---|---|---|---|
| chickpea | 30 cm x 10 cm | 0.8-1 | 9957 | 55043 |
| apple | 5-8m x 5-8m | 30 | 65000 | 460000 |
| papaya | 2m x 2m | 30 | 210000 | 279880 |
| kidney beans | 90cm x 90cm | 1 | 15000 | 85000 |
| pigeon peas | 60-75cm x 30 cm | 0.6 to 0.8 | 14047 | 30000 |

# Profit Analysis



Fruits

Pulses

Cereals

Fibre Crops

Plantation Crops

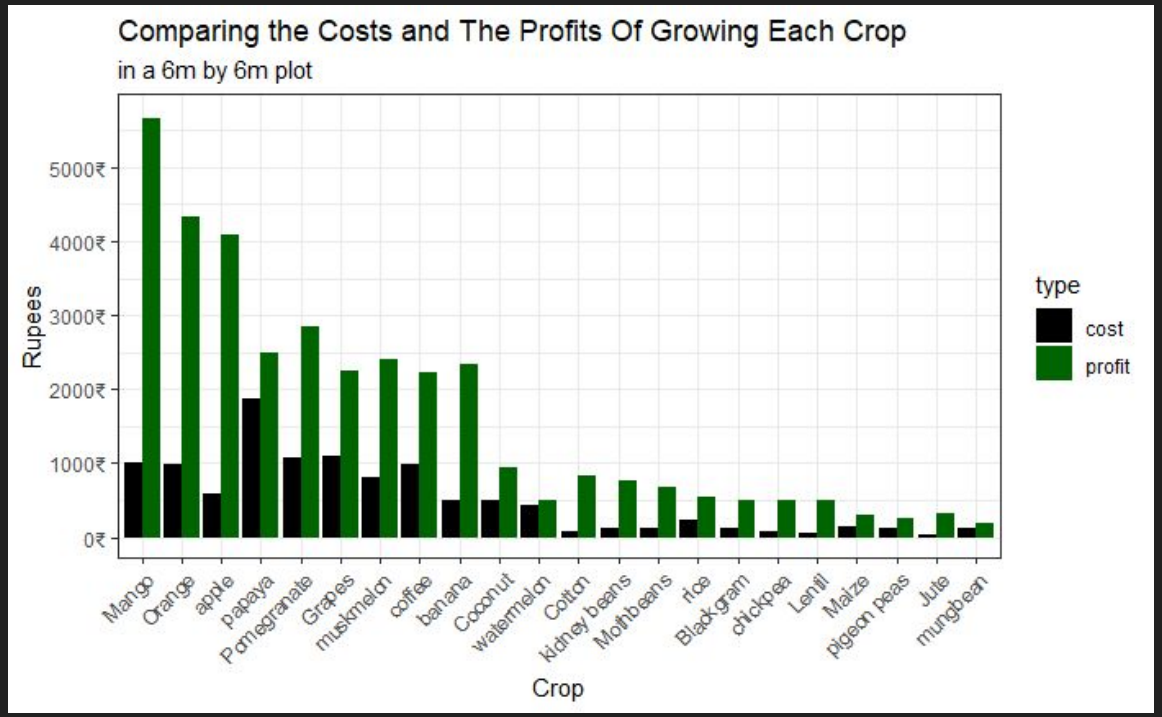
Total cost of 1 acre rice farming ~27,940
Yield : 23 quintals
Gross Returns : ~78,315
Farmers income - ~50,975

*Costs are represented in INR

VANDERBILT UNIVERSITY®
Data Science Institute

# Profit Analysis

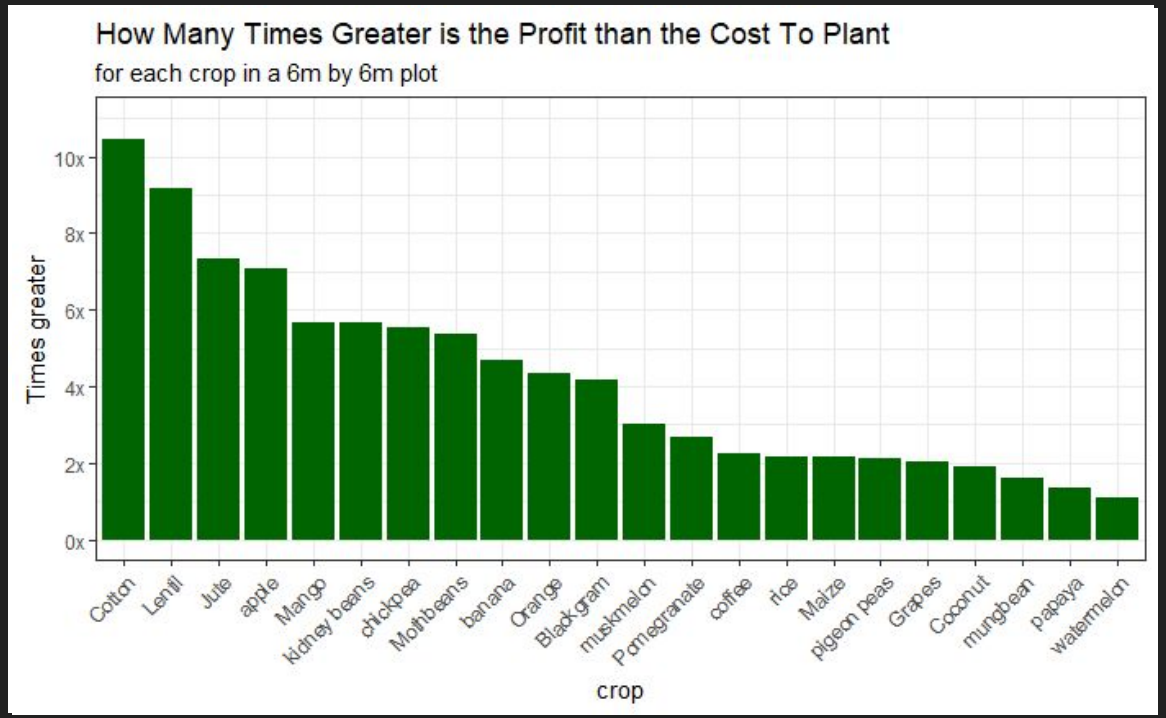
- Comparing Cost to Plant and Profits



VANDERBILT UNIVERSITY®
Data Science Institute

# Profit Analysis

- How much greater is the profit than the cost?



VANDERBILT UNIVERSITY®
Data Science Institute

# Our Assumptions

- Data on each crop represented good conditions for each crop
- Standardized crop data before splitting into a train set and a test set

VANDERBILT UNIVERSITY®
Data Science Institute

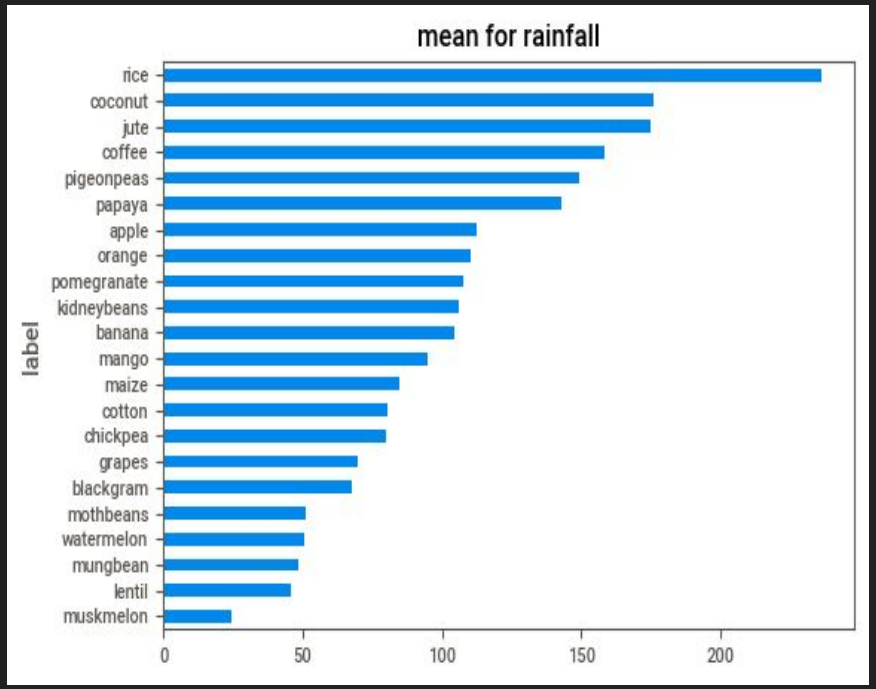
# Weather and Soil Influence

- Humidity & Rainfall
  - Moisture or drought stress accounts for about 30–70% loss of productivity of field crops during crop growth period
- Temperature
  - Heat stress contributed about 40% to overall yield loss of crops, ex: 1.0–1.7% yield loss per day in maize for every raise in temperature above 30°C
- Soil Minerals (N, K, P)
  - The lack of these nutrients in the soil causes deficiencies in plants, and their excess leads to toxicities, which have negative impacts on crop yields.
- Soil Ph Level
  - Low soil pH significantly affects crop growth and therefore decreases yield. In maize for instance, soil acidity causes yield loss of up to 69%.

VANDERBILT UNIVERSITY®
Data Science Institute

# Weather EDA

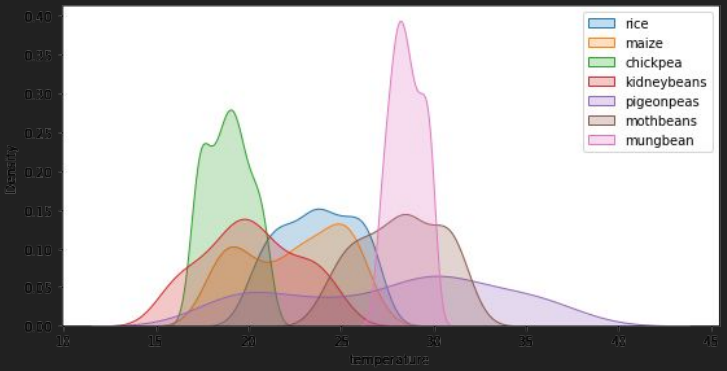
- Weather Variables

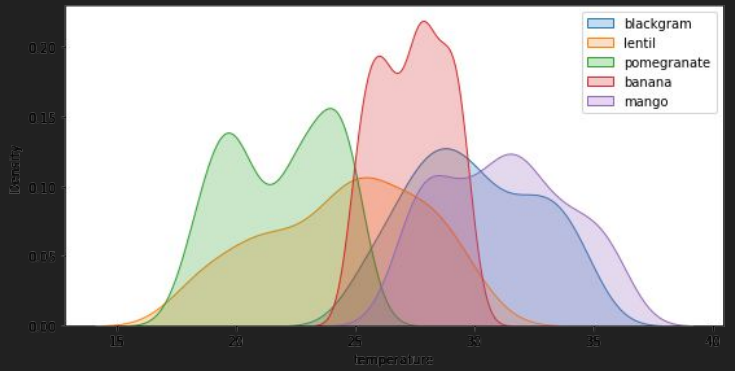


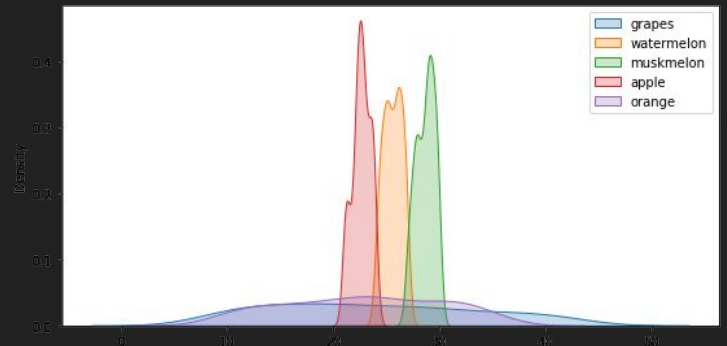
VANDERBILT UNIVERSITY®
Data Science Institute

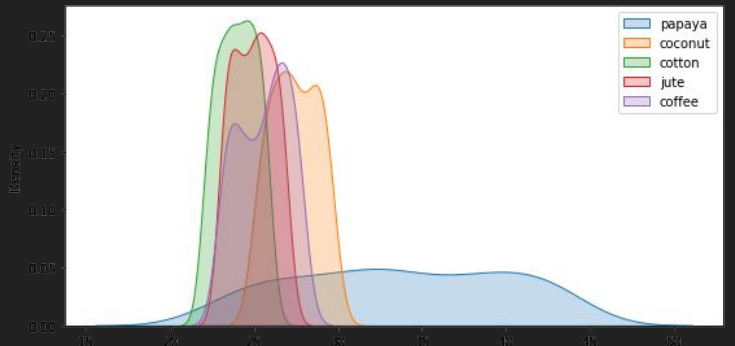
# Weather EDA

- Weather Variables: Distributions for Temperature



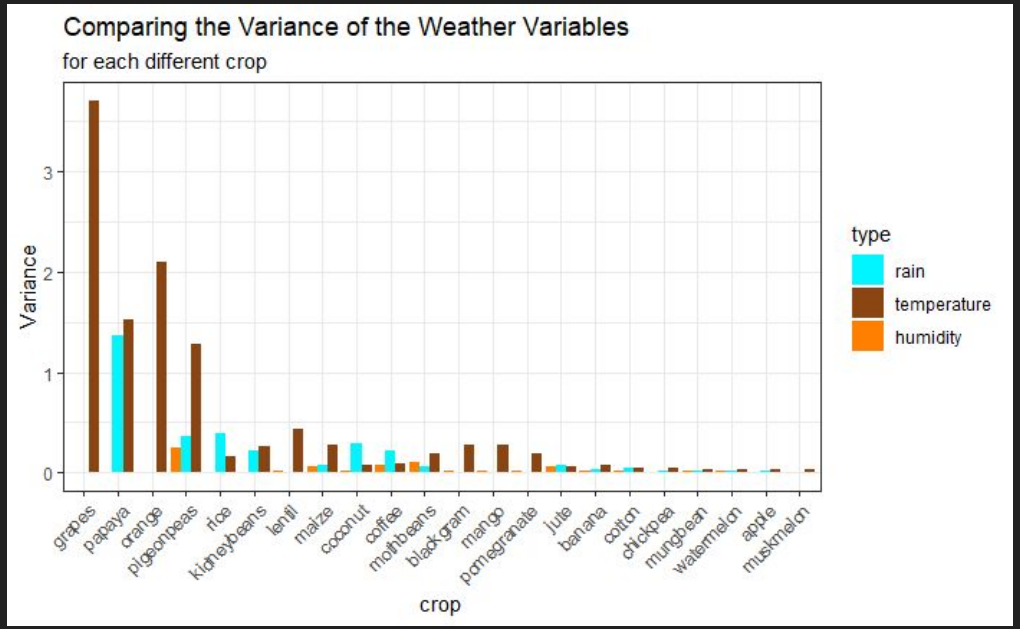



VANDERBILT UNIVERSITY®
Data Science Institute

# Weather EDA

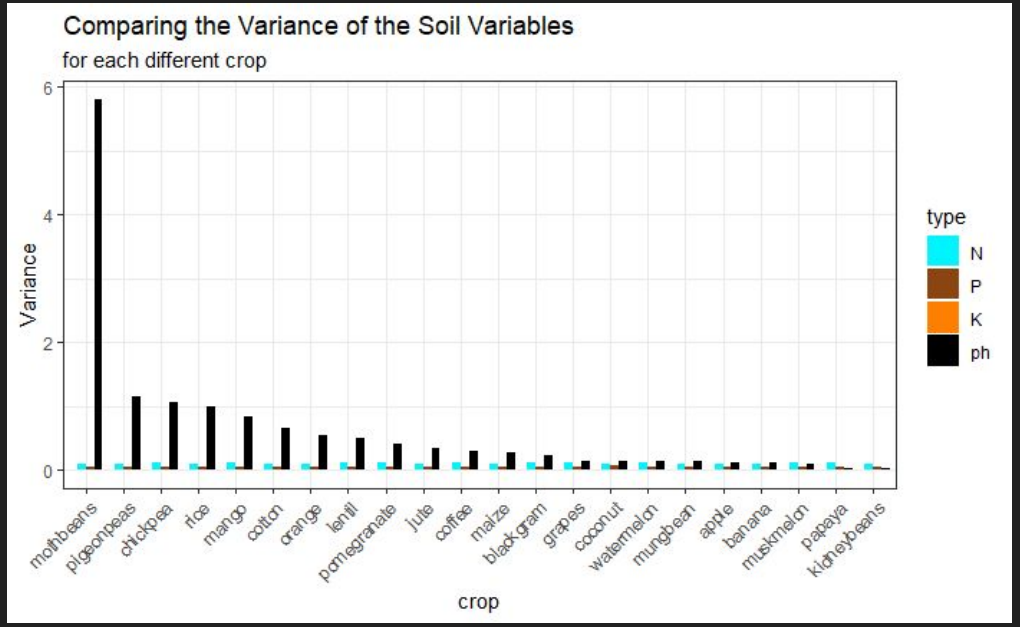
- Weather Variables



VANDERBILT UNIVERSITY®
Data Science Institute

# Soil EDA

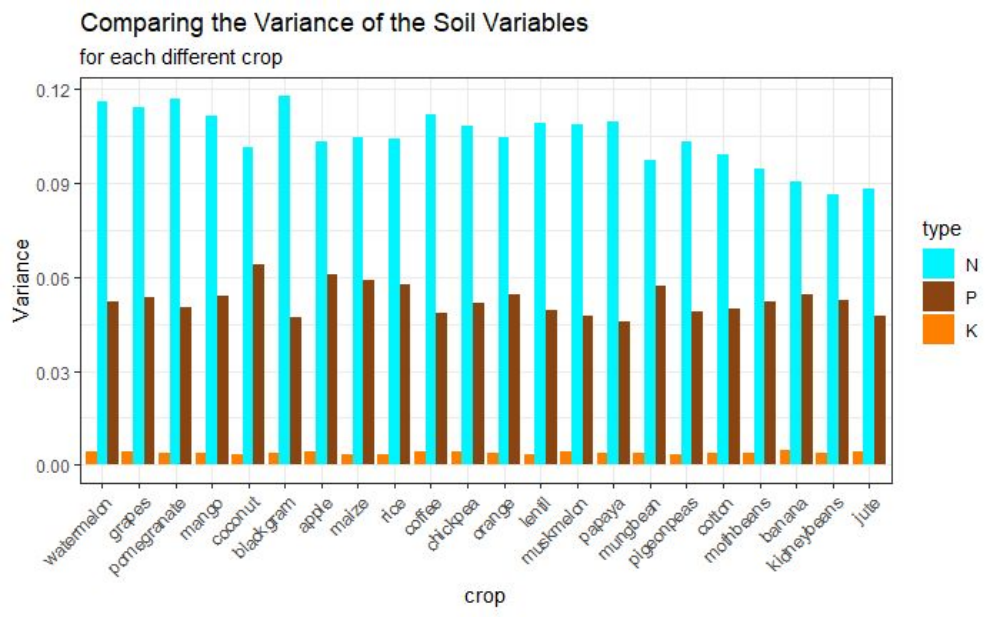
- Soil Variables



VANDERBILT UNIVERSITY®
Data Science Institute

# Soil EDA

- Soil Variables



VANDERBILT UNIVERSITY®
Data Science Institute

# Modeling Process

1. Determine the best crop with classification
2. Determine percent yield that each crop will have in the given conditions
3. Combine percent yield with field size, crop value, and crop cultivation cost variables to determine expected profit for each crop

VANDERBILT UNIVERSITY®
Data Science Institute

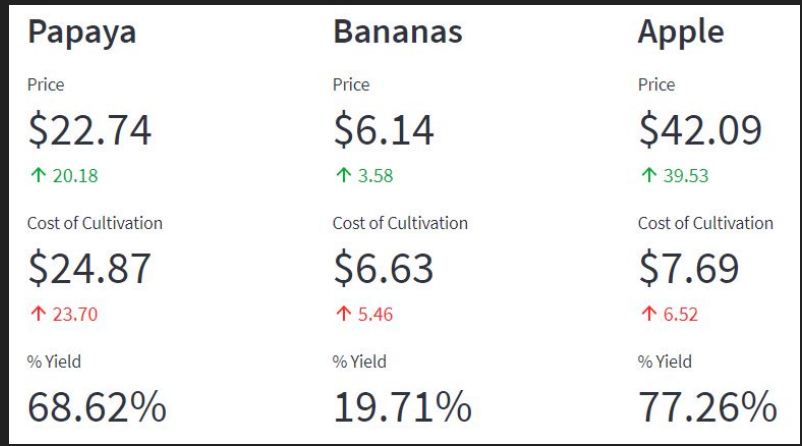
# Model Info

## Inputs:

- Soil Variables
- Climate Variables
- Desired Currency
- Field Area

## Outputs:

- Recommendations for crops to plant
- Percentage yield for crop in each field
- Cost to grow each crop
- Profit that can be gain from growing each crop



VANDERBILT UNIVERSITY®
Data Science Institute

# Model Info

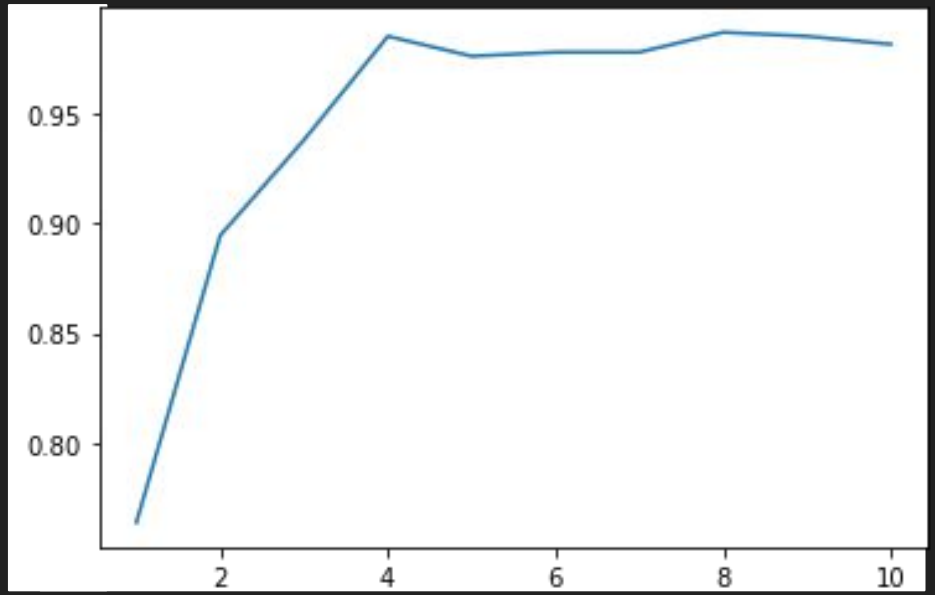
- Random Forest classifier
  - Selected so that it would treat picking the right crop for a set of conditions as a classification problem.
  - The high eventual accuracy gives us confidence that the model is able to distinguish the small changes in conditions that vary between all 22 crops.
- Parameters
  - Number of trees = 100
  - Max Depth = 4
  - Number of Features per Tree = 3

VANDERBILT UNIVERSITY®
Data Science Institute

# Our Model

- The first parameter chosen for tuning was Max Tree Depth
- The accuracy was considered acceptable after this so no further tuning was done

## Model Tuning: Max Tree Depth



VANDERBILT UNIVERSITY®
Data Science Institute

# Our Dashboard

[Crop Recommender System Dashboard]

Try it yourself!

bit.ly/3p7RX5G