Final Project Proposal

2023-07-20

Team members

1. Cameron Greenwalt
2. Mark Bauer
3. Ben Carman

Project Title

Predicting Grain Crop Yields in India for Years 1997 - 2020

Data Description

The **Agricultural Crop Yield in Indian States Dataset** contains information about crop type, inputs, weather, time (year), and yield for various crops grown in India. It has 9 predictors which we will use to predict `Yield`.

The data set was pulled from Kaggle. The author aggregated data from the following sources to create the dataset:

1. https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics-0
2. https://www.fao.org/faostat/en/#data
3. https://data.gov.in/catalog/rainfall-india
4. https://environicsindia.in/
5. https://www.imdpune.gov.in/library/public/e-book110.pdf

In addition to being well curated, the data is of interest because one of our team members (Mark Bauer) has an agricultural background. We are curious to know more about efficient land usage as well as how weather and inputs interact to determine yield for the several crops we will examine. We will limit the types of crops under observation to a subset of grains (e.g., wheat, barley, rye, etc.).

Evidence that the data can be loaded into R

```
library(tidyverse)
library(knitr)

crops = read_csv("crop_yield.csv")

# Other grain types may be added later
```

```r
crops = crops[crops$Crop %in% c("Wheat", "Barley", "Maize", "Rye", "Rice"),]

# Yield is our response
kable(
  head(crops, 5)
)
```

Crop	Crop_Year	Season	State	Area	Production	Annual_Rainfall	Fertilizer	Pesticide	Yield
Maize	1997	Kharif	Assam	19216	14721	2051.4	1828787	5956.96	0.6156522
Rice	1997	Autumn	Assam	607358	398311	2051.4	57802261	188280.98	0.7808696
Rice	1997	Summer	Assam	174974	209623	2051.4	16652276	54241.94	1.0604348
Rice	1997	Winter	Assam	1743321	1647296	2051.4	165911860	540429.51	0.9413043
Wheat	1997	Rabi	Assam	84698	110054	2051.4	8060709	26256.38	1.2595238