Impact of Drought on Agriculture

NOAA - NIDIS Capstone Project




MSBA Team 3: Selma Sentissi, Jaya Nagesh, Shuyi Zhu, Shamika Kalwe

Introduction

The purpose of this project is to increase dissemination of drought related information to farmers in a way that allows them to develop a multi-seasonal drought mitigation strategy. We have the following goals in mind:

- Correlating drought with Agricultural industry (starting from a crop-state pair and expanding from there, e.g. Corn-New York)
- Predicting parameters related to Agriculture with reference to drought (e.g. yield etc.)
- A front-end/dashboard to enable dissemination of the above analysis to farmers
- A database that stores relevant subset of data from public repositories, with a backend to update/merge routinely or as needed.

Methodology

Data Collection and Analysis

We chose to focus on Agricultural Industry and started with New York state and the corn crop. We chose drought indicators like drought indices (USDM, EDDI, SPI)* and independent variables like Temperature and Precipitation as predictors. We analysed these variable against Corn yield and acres planted under corn.

Data Collection and Analysis

Data Collection:

- We gathered the data from multiple government sponsored sources like USDA*, USDM*, ACIS* etc.
- Automated process by utilizing APIs for dynamic download

Complete Dataset Summary:

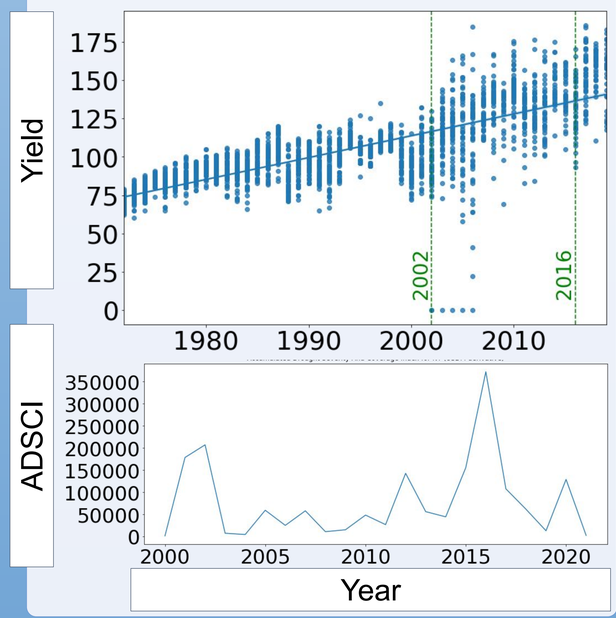
- Data Timescope: 2000 - 2020
- 1% of missing values in USDA were removed
- 52 New York Counties Values across datasets

Data Manipulation:

- USDA data is recorded on an annual basis while the drought indicators are recorded on more granular levels (weekly/daily). Thus we brought all the data on same time scale (annual) by either aggregating it over the time (USDM), or by computing annual quartiles (temperature, precipitation)

Exploratory Data Analysis (EDA):

- Plot 1 (top): Corn yield (bushels/acre) regressed on year. As years increase, corn yield is also increasing.
- Plot 2 (bottom): year vs. Accumulated Drought Severity and Coverage Index*. Variable drought through the years (peaking in 2002 & 2016.)



Machine Learning Models

We used the dataset to train a supervised machine learning model to help predict the corn yield (complete year and growing season) in New York State. The model was trained in various advanced methods, such as linear regression, random forest regressor.

- The inputs variables: ADSCI_yr, Q25%, 50%, 75% precipitation and temperature values
- The predicted outcome: corn yield
- The dataset was divided into 75% training set, 15% validation set, and 10% testing set.
- Out of all models, random forest is the best with the lowest RMSE and highest R2 value.

Variable: Corn Yield (Bushels/ acre); Predictors: Temperature, Precipitation and USDM			
Linear Regression		Random Forest	
Predictors measured Annually	Predictors measured over Growing Season	Predictors measured Annually	Predictors measured over Growing Season
r2: 0.1885	r2: 0.2357	r2: 0.3202	r2: 0.2645
RMSE: 20.6058	RMSE: 19.9977	RMSE: 18.8602	RMSE: 19.6169
After Standardizing Data			
r2: 0.1885	r2: 0.2357	**r2: 0.3396**	r2: 0.274
RMSE: 0.7876	*RMSE: 0.7643*	*RMSE: 0.7105*	*RMSE: 0.745*

Results

Correlations	USDM (complete year)	USDM (growing season)	Temperature	Precipitation
County - Yield	-0.07	-0.103	0.162	-0.098
County - Acres Planted	0.067	0.142	0.115	-0.303

Modeling:

Overall when we were looking at parameters individually our models did not have good predictive value. But after combining all the parameters our models had decent results. The best model was the Random Forest with R2 of ~0.34 (shown in the table under Machine Learning Models tab).

Limitations and Challenges

- Lacking adequate domain knowledge
- Gathering proper data - data (agricultural vs. weather related) very dispersed
- Relying on others for expertise (propor indices to predict corn production)

Next Steps

- Try out other machine learning models such as kernel, bagging, boosting, and polynomial regression.
- Design dashboard and create a database to store and update related data for this project.
- Create an appendix document where we list all the data sources, API applications, and indices explanation.
- Expand drought impact analysis to other crops, e.g. apples

Appendix (*)

USDM - U.S. Drought Monitor
DSCI - Drought Severity Coverage Index (consolidated measure of USDM across drought levels)
ADSCI - Accumulated DSCI (aggregated DSCI over time of interest)
EDDI - Evaporative Demand Drought Index
SPI - Standardized Precipitation Index
USDA - U.S. Department of Agriculture
ACIS: Applied Climate Information System

Acknowledgement

We also appreciate Sylvia Reeves, Mike Hobbins, Keith Eggleston, Brian N. Belcher, Professor Arthur DeGaetano, who helped us a lot with this project. Their professional expertise in agricultural field and drought field help us navigate through this project.

Github Link

https://github.com/BU-NOAA-Capstone/Capstone

LinkedIn Profile

https://www.linkedin.com/in/jaya-nagesh
https://www.linkedin.com/in/shuyizhu9712
https://www.linkedin.com/in/shamika-kalwe
https://www.linkedin.com/in/selma-sentissi