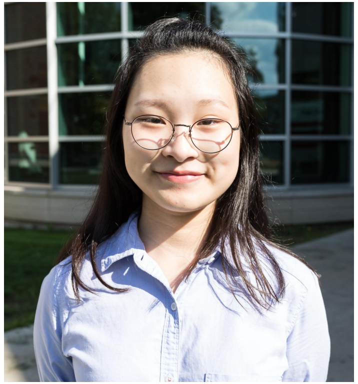
Impact of Drought on Agriculture

Team

Jaya Nagesh



Shuyi Zhu

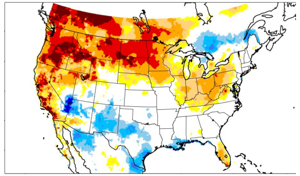
Shamika Kalwe

Selma Sentissi

Introduction and Background

Why Drought?

- Raising concern
- Affecting almost all US states
- Short-term nature (<6 months)
- Costs ~$9 billion a year
- Farmers are the worst affected
- Information asymmetry



Why New York & Corn?

- Leading agricultural state
- $5.75 billion in revenue and employing 200,000 (2017)
- Corn is one of the New York's top 10 agricultural products
- Generated a revenue of $256 million (2017)

Project Goal

Collect, Analyze and Disseminate the Drought data to bridge the information gap

- Correlate drought with Agricultural industry (Corn crop -New York state)
- Predictive analytics on Corn yield
- Dashboard to enable dissemination of the information and analysis to farmers

Methodology and Keywords

Methodology:

Drought Indices Used:

SPI *Standardised Precipitation Index*	SPEI *Standardised Precipitation-Evapotranspiration Index*	PDSI *Palmer Drought Severity Index*	EDDI *Evaporative Demand Drought Index*	USDM *US Drought Monitor*	_n *EDDI_6, SPI_180*
Precipitation based	**Precipitation and Temperature based**	**Precipitation, Temperature and Soil Moisture based**	**Atmospheric evaporative demand**	**Composite Drought Index**	**The time over which index is aggregated 6 months, 180 days**

Models

Predicting Yield Using Drought Indicators

Linear Methods:

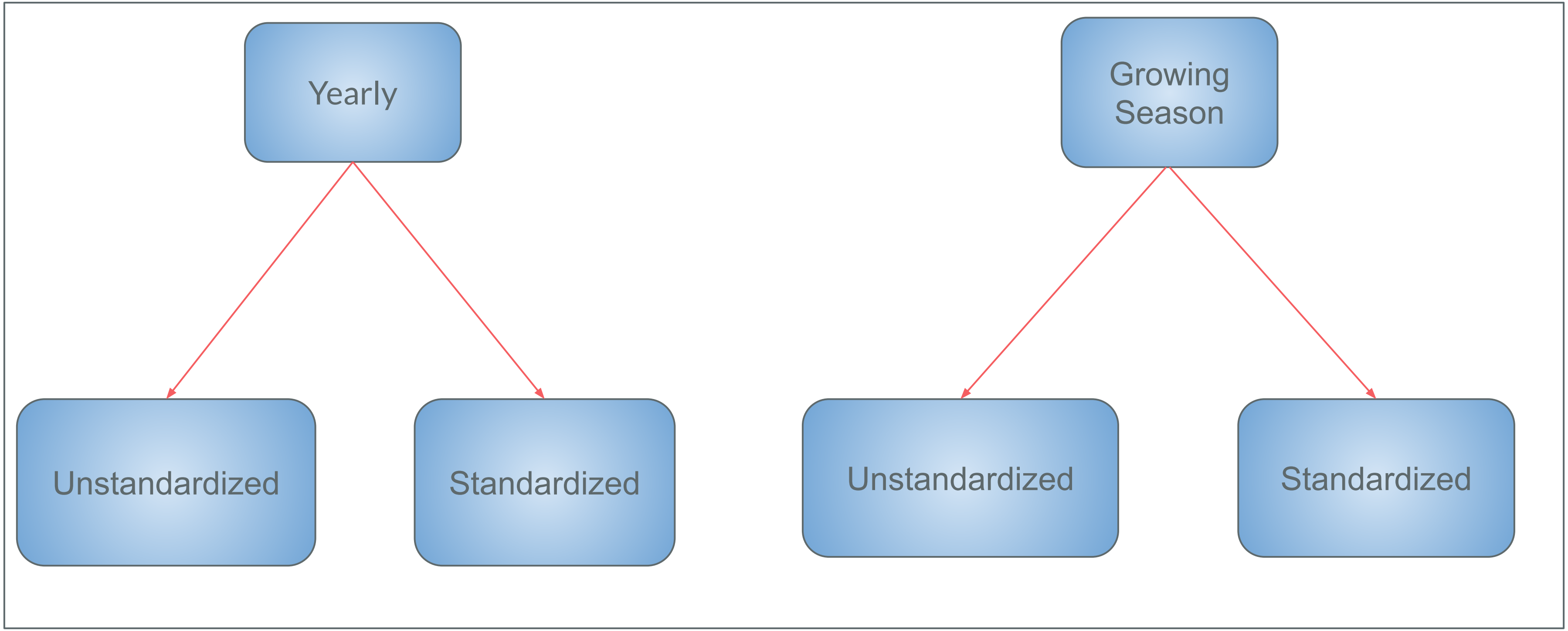
- Linear Regression
- Kernel Regression
- Polynomial Regression

Tree Based Methods:

- Boosting
  - Tuned with different base estimators, varied number of estimators, etc
- Random Forest Regressor
  - Tuned number of estimators, max numbers of features, max depth
  - *Overall, this model did best, indicating non non-linear patterns in the data*

Models

Splitting Options



Models

Results

Models on Yearly Values Unstandardized	R2 train	R2 validation	R2 test	Error rate
Linear Regression	0.434	0.350	0.444	0.986
Polynomial Regression	1.000	-0.170	-0.064	0.185
Random Forest	0.913	0.443	0.301	0.185
Boosting	0.005	0.490	0.419	0.138
Kernel	0.375	0.263	0.350	0.140

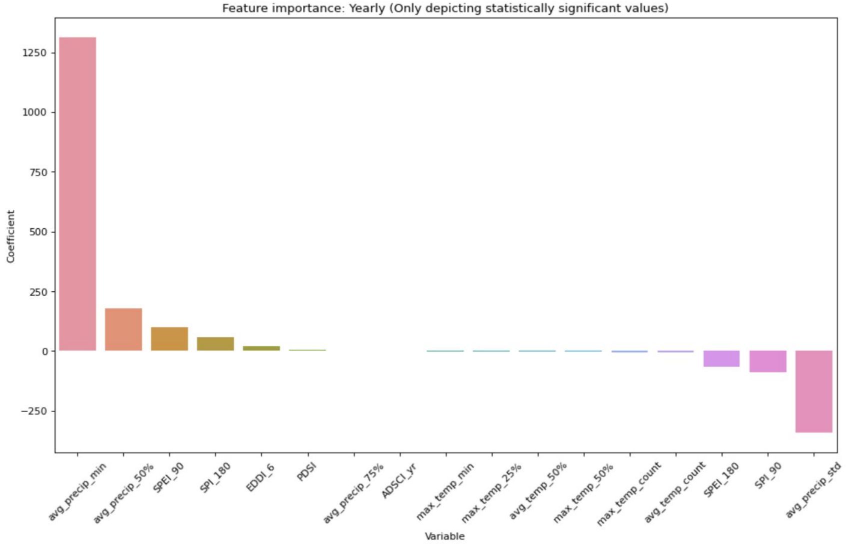
Models on Yearly Values Standardized	R2 train	R2 validation	R2 test	Error rate
Linear Regression	0.434	0.350	0.444	0.881
Polynomial Regression	0.836	-0.168	0.100	2.240
Random Forest	0.917	0.441	0.431	1.975
Boosting	0.583	0.473	0.362	1.429
Kernel	0.431	0.341	0.447	0.881

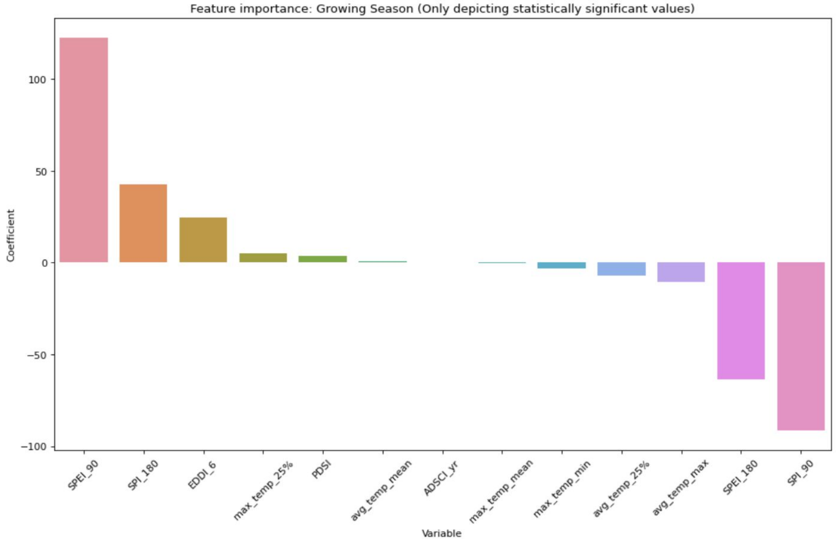
Models on Growing Season Unstandardized	R2 train	R2 validation	R2 test	Error rate
Linear Regression	0.379	0.421	0.333	0.145
Polynomial Regression	0.990	0.120	0.207	0.149
Random Forest	0.903	0.450	0.439	0.134
Boosting	0.005	0.436	0.409	0.139
Kernel	0.365	0.386	0.310	0.148

Models on Growing Season Standardized	R2 train	R2 validation	R2 test	Error rate
Linear Regression	0.379	0.421	0.328	0.713
Polynomial Regression	1.000	0.011	0.001	2.734
Random Forest	0.908	0.459	0.445	2.383
Boosting	0.561	0.477	0.372	2.038
Kernel	0.375	0.404	0.330	0.795

Feature Importance

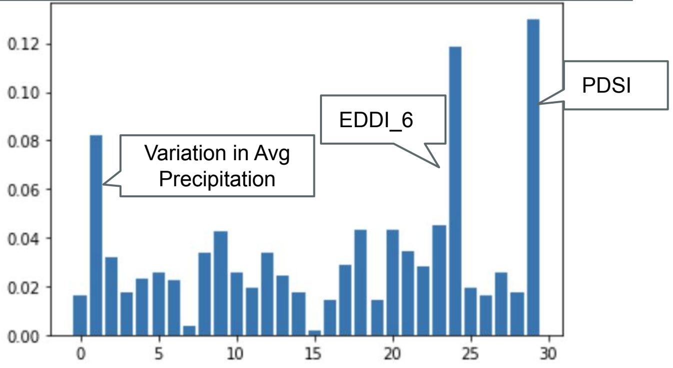
Ordinary Least Squares



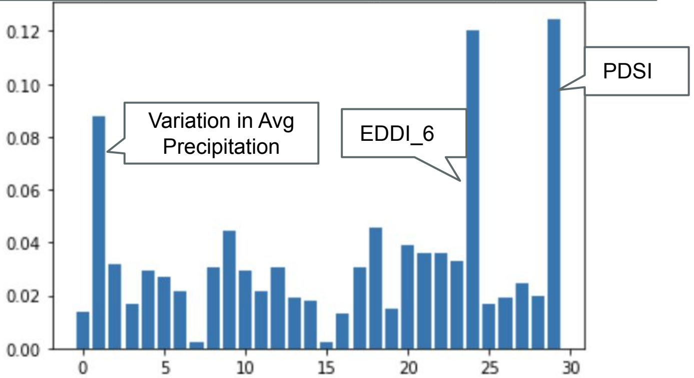


Feature Importance Random Forest

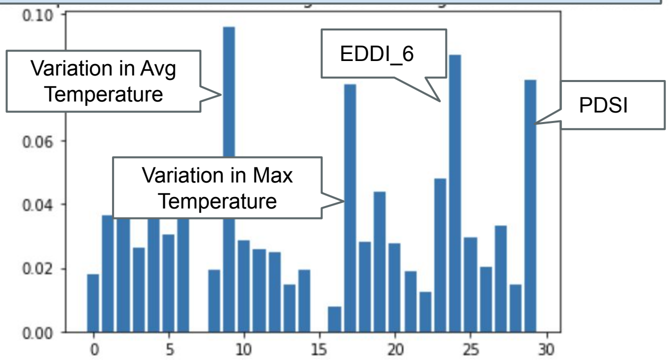
Yearly Unstandardized



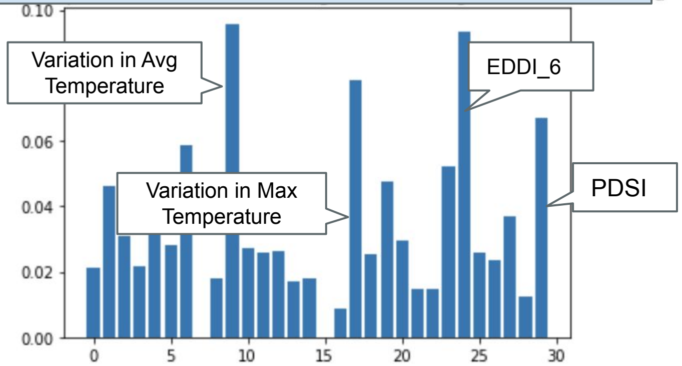
Yearly Standardized



Growing Season Unstandardized

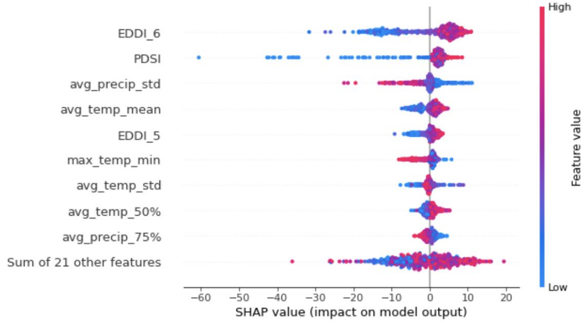


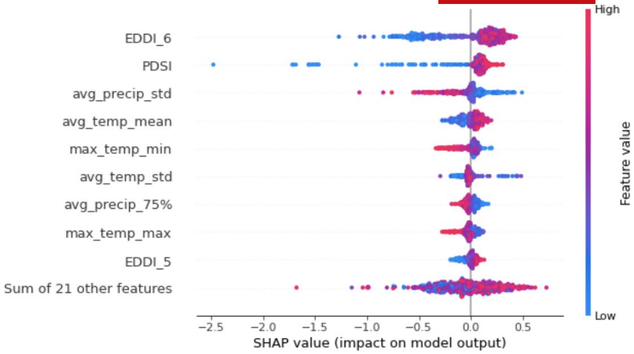
Growing Season Standardized



Feature Importance Random Forest

SHAP

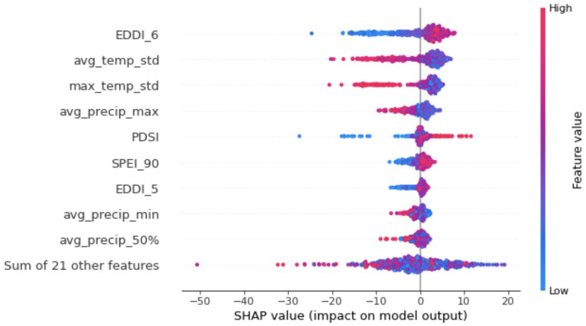




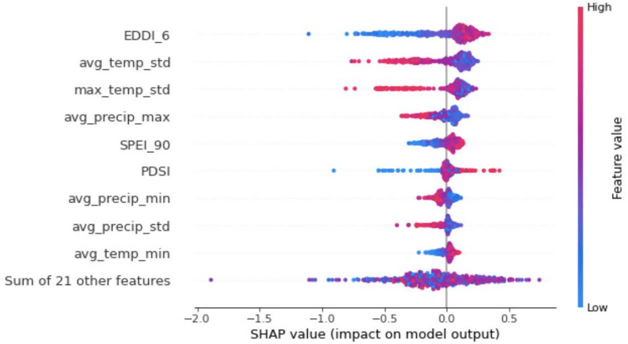
(1) Unstandardized Yearly

(2) Standardized Yearly

97%|=================== | 509/527 [00:15<00:00]



96%|=================== | 504/527 [00:14<00:00]



(3) Unstandardized Growing Season

(4) Standardized Growing Season

Dashboarding

Aim: Increase dissemination of drought and Corn related information to New York Corn growing farmers

Tools used:

Publicly accessible:

Part-01: https://public.tableau.com/app/profile/shamika.kalwe/viz/NOAA-USDA/Story1

Part-02: https://public.tableau.com/app/profile/shuyi.zhu/viz/NOAA-Part2_final/Story1

Conclusion

Models:

- Random Forest Regressor did the best generally
- EDDI 6, PDSI, temperature variations are the most important features across the feature importance methods

Dashboard:

- We built a publicly accessible Tableau dashboard to facilitate effective information dissemination
- We consulted with industry experts on the dashboard we built and incorporated their feedback to make it more understandable and useful for farmers

Thank you!