Modeling Crop Gene-Environment Interactions under Climate Change

PURDUE UNIVERSITY | The Data Mine

BAYER

Project Overview and Motivations

Overview:

The human population is growing rapidly and the demand for food continues to increase. In the figure to the right, the corn yield has increased dramatically over the years. The increase began in the 1930s due to the double-cross hybrid corn. The demand for corn will continue to increase which is a driving force in the motivation for this project.

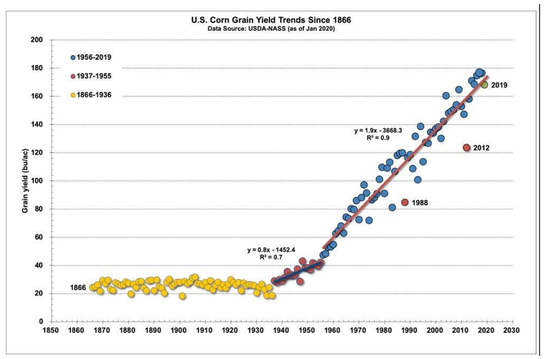


Figure: Annual Corn Grain Yields since 1866. The data was derived from annual USDA-NASS crop production reports.

Motivations:

- Cope with climate change
- Reduce efforts in predictions
- Resource management
- Competitive advantages in crop growing

MODELING APPROACHES

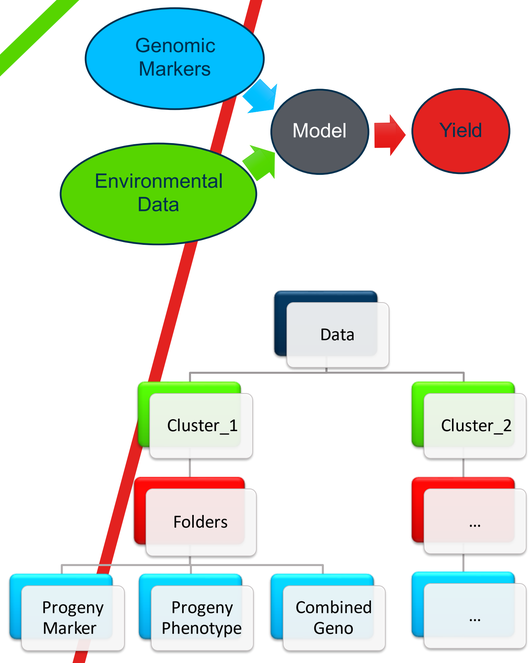
	RMSE	MAPE	R^2
Light Gradient Boost	17.825	.0824	0.7288
Extreme Gradient Boost	18.954	.0824	0.7288
Support Vector Machine	29.2623	0.1322	0.3574
Bayesian Ridge	28.690	0.1302	.36068

Metrics:

- RMSE: mean squared error
- MAPE: mean absolute percentage error
- R^2: percentage of the variation in yield explained by the model

Model comparison:

Comparing the models, Light gradient Boost has the best scores. For RMSE, the root mean square error, lower scores represents a better fit to the data. Additionally, for MAPE, the mean absolute percentage error, a lower score is also preferred. The higher the R^2 , the better the model fits the data. When looking at LGBoost we can see its RMSE, MAPE, R^2 scores are most suitable for our goals.



DATA OVERVIEW

Genotype Data:

- Around 3000 genetic markers for parent and progeny
- 1 represents dominant homozygous, 0 represents heterozygous, and -1 represents recessive homozygous

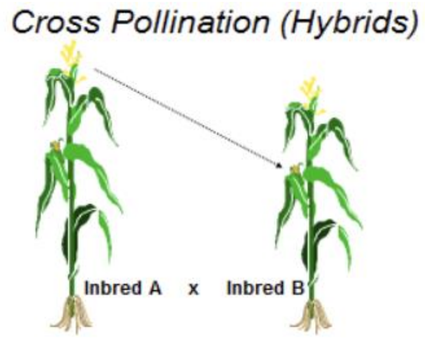
Post-growing season phenotype

Non-Bayer Data Sources for Weather:

- Copernicus ECMWF Reanalysis 5 (ERA5)
- 2030, 2040, 2050 benchmarks Coupled Model Intercomparison Projects (CMIP6)

Weather Variables:

- Water Evaporation From Canopy
- Surface Upward Heat Flux
- Total and Surface Runoff Rate
- Precipitation Rate
- Snow Fall/Snow Melt Rates and Snow Thickness
- Air and Surface Temperature
- Directional Wind Speeds



Inbred hybrid Inbred

CONCLUSIONS and PREDICTIONS

- We have created a new data frame that implements genomic and climate data
- We used new models and worked on creating more accurate predictions
- Integrating a user interface that allows us to look at predictions based on different variables

REFERENCES

1. https://passel2.unl.edu/view/lesson/c3ded390efbf/9
2. https://blogs.cornell.edu/agsci-interns/2014/08/15/pollinating-season-at-pioneer/
3. "Historical Corn Grain Yields In The U.S." Purdue Entomology

ACKNOWLEDGEMENTS

Jack Zhang
Nima Hamidi Ghalehjedh
David Glass
Entire Data Mine Staff

Allison Kramp, Jianxiang Tao, Victoria White, Xiaojin Wang, Chelsea Medina, Crystal Gutierrez, Jack Hansen, Justin Wynn, Adian Acosta, Jorge Mendoza, Ayishe Jin, Ernesto Nevarez, Luis Columbie