Progress Report - County Level Corn Yield Prediction

CSE 6242, Team Des Moines Power

Introduction — motivation

Yield prediction is an important part of any farming operation. It helps producers benchmark and understand their crop's potential for the year. Lots of farmers have difficulty predicting yields and understanding what their upcoming corn will bring. Our primary objective is to provide a scenario planning tool to forecast corn crop yields on a county level. This tool leverages state of the art models, leveraging highly researched covariates for prediction, such as days over 30 degrees celsius (Schlenker at al.). This tool will then allow users to adjust the covariates to develop scenarios and view the impacts to yield at a county level.

Problem definition

Today corn yields by county are estimated by several organizations including the USDA, public and private universities, and private organizations like commodity traders and grain elevators. Models and methodologies vary greatly. However, with producers closely guarding their operations data, it can be difficult for producers to flexibly benchmark for scenarios against their peers. County level forecasts have improved with more modern algorithms applied to county level forecasts. We improved upon these modern algorithms with the ability to build scenarios, enabling producers, government, and traders to flexibly benchmark early in the season.

All of the examples that we discovered from our research had one common theme - no prediction model let their users input their local data to get a quick and easy result. We believe that users need to be able to input their own data to make predictions based off of the current year. The ability to plan for contingencies, and understand the impact of key parameters is a key advantage to any user of our tool.

Literature Survey

As a group we read and reviewed a large variety of literature to prepare us for the development of our model and UI tool. We gain a lot of insights from other individuals who have many years of experience of predicting corn yield.

Rural grain elevators often rely on years of experience and insights, along with consideration of qualitative and quantitative insights to develop estimates (or educated guesses). Highly technical analyses like those performed by researchers at University of Madison-Wisconsin leveraging a Baysian neural network methodology are also created and made publicly available (Yuchi et al.). The USDA has developed a method of representative sampling and surveying farmers across the nation to generate estimates as well (Thessen).

This range of insights are often all considered by producers and others to get a good sense of how productive the upcoming year will be. These methodologies all seek to predict average yield on a county level while others seek to provide predictions on a granular field basis (Runge and Benci) (Cerrato and Blackmer). These methods have been enabled by satellite imagery and remote sensing (Guanyuan and Bruno). Seed, chemical, and equipment companies have used this space to differentiate their product offerings with yield mapping, as well as kernel counting image detection (Wu et al.).

The largest gap today is the ability to create robust scenario planning leveraging modern models. While fairly reliable models can be developed as soon as June 1st (Shahhosseini et al.) these models rely on future assumptions, or historical data. The ability to plan for contingencies is highly desirable for government, producers, and private organizations.

The impact of this tool on our users is measured by the accuracy of the tool and the model, and the impact of the decisions it makes. The decisions can vary highly based on the user, and in many cases can involve large sums of money, like in the case of agricultural subsidies by the government. As demonstrated by Fatima A.M. Tenorio producers highly leverage benchmarking to make decisions (Tenorio), these decisions would be suboptimal if leveraged by poor modeling.

Proposed method

The title of our final product is the Corn Yield Calculator. It consists of agronomic, economic, and weather data that feeds a model that predicts a corn yield by county. We will test a variety of supervised machine learning methods to identify the best model to generate predictions with. The product is housed in a DASH application. The users of the Corn Yield Calculator are able to input their own data to get a prediction for a county's anticipated yield.

The driver of the model is a set of features that we have gathered from open source data online. The features that we have found come from Census data, Meteostat, and USDA. The features extend back for 10 years (except the 2020 Census data) and are mapped by county. We did not focus just on the high corn growing states, but rather the entire United States. From an initial look at the features some states have more data than others. We found that the three most important categories for our features feeding the model were agronomic, economic, and weather.

The USDA and Meteostat data are pulling data via API from the open source data. We have set the calculator up in this way so that is refreshing the model with every user who is using the calculator. Cory on our team was able to clean the data we are receiving from the API and get it into one clean table. For the Meteostat data, he worked to match the counties with the closest weather station the weather is reporting from. One of the innovations for our product is the amount of robust weather data that is feeding our model. No other studies that our group researched had as much robust weather data we are featuring in ours.

Features:

- Min Temperature
- Max Temperature
- Too Cold Days
- Too Hot Days
- Total Rain
- Rain Days
- Corn Acres Planted
- Corn Bushel Price
- Land Rental Price
- County Population
- Latitude & Longitude

The model that is feeding our DASH application is an Extreme Gradient Boosting (XGBoost) Ensemble model. In the experiments and evaluation section it describes why we selected a XGBoost for our model. The model is run from all of the features that have been cleaned into one table. The model is looking by county and back 10 years to predict a current corn yield. We then used this information to predict corn yields by county for the 2022 harvest year.

The design and UI of our product is housed in a DASH application. The dashboard constructed allows the user to operate the tool like a calculator. The user is able to enter in their county, state, and a set of parameters from the list of agronomic, economic, and weather features. Once those parameters are entered the application pulls the associate value from the pre-scored dataset of model results and displays any of the updated features needed to keep the model up to date. Another reason we feel that our product is innovative, is because the user is providing parameters inputs like a calculator. None of the studies that the group looked at gave users a way to customize their corn yield prediction for the year.

The dashboard is broken into three sections - the county's yield prediction, historical yields barchart, and an interactive map. The county's yield prediction is at the middle of the page and color coded. This allows the user to easily see their results and the prediction for the year for their county. The historical yield barchart allows the user to dig into their county's data for year over year yields. The barchart gives insights into how their yield prediction from the calculator will stack up against previous years for that county. Lastly, the interactive map displays average yields for counties all across the nation. The user gains insights on how their county benchmarks against other counties in the US.

Innovations of our product:

1. User parameter inputs like a calculator
2. Robust use of weather data in our model
3. Quick predictions for corn yield at the touch of a button

Experiments/ Evaluation

The team gathered a long list of features that were experimented with and implemented into multiple models. The models that we tested for our Corn Yield Calculator were Linear Regression, Random Forest, LSTM, Ridge, Lasso, SVM, and XGBoost. How we evaluated these models was by looking at Mean Square Error, Root Mean Square Error, and Mean Absolute Error. We focused our evaluation on which models produced the least amount of error when tested with K-Fold Cross Validation using 10 folds. After evaluating each of the models we found that a simple XGBoost provided the best yield predictions. To further optimize model performance after selection applied hyperparameter tuning using GridSeachCV from the Scikit Learn package, this allowed us to arrive at the best performing parameters for the model.

Steps were also taken to select the optimal parameters for modeling. We evaluated the features by evaluating p-values for significance. We tested for a high R-squared value for the final output of the model to give the tool the best yield predictions. Highly correlated features were condensed, or one was removed. We also found through experimentation that scaling our features led to slightly worse results, for this reason we chose not to scale the features in our final model.

We also conducted a usability evaluation with our UI DASH dashboard to make sure the interface is user friendly for our target audience. We surveyed local farmers and took feedback on how we can improve their user experience with the UI. Since the group is located in central Iowa it was helpful to get feedback on the interface from local corn growers. The feedback we received from our test users drove how the completed DASH dashboard looked.

Questions the Corn Yield Calculator answers:

- What is my predicted corn yield based on this year's conditions?
- How will my corn yield stack up against my predicted county average?
- Historically, what were my county's corn yields?
- What does the United State's corn yield averages look like?

From our experiments and final product our team had many observations along in addition to our model experiments detailed above. We found that corn is not grown in every county in every state, and more importantly for our model, corn yields per bushel per acre on average also vary greatly by county by state. Even though not all counties have recorded corn yields for 2021, we were able to make some other insights as well.

We found that in 2021, the county with the best corn yield was in Iowa in Sac county, with over 230 bushels/acre of corn on average. In aggregate, we also saw that corn yields have been increasing year over year over the past 10 years, meaning we expect corn yields to grow steadily in most counties and states over the years to come.

In addition to these findings, as a farmer, our users are able to get new insights on their corn yields and better predict yields for the year. Using our yield lookup, a farmer can see historic corn yields for their county and understand how their individual yields on their farm compares to average. The user can also see how corn yields in their county compares to neighboring counties and understand how their yields may compare as well. The most important feature of our model is our corn yield calculator predictions where a user can select their county and local variables, such as weather, to determine what yields they would expect for their upcoming harvest.

Conclusion

Our team has built a strong yield prediction model based on agronomic, economic, and weather features. It is housed in an easy to use UI that allows users to see what yield is predicted for the selected county based on their own parameters. We conducted feedback from local farmers and received areas to improve the UI and parameters. One of the areas that we could expand in the future is to grow the calculator to also include soybeans or other crops. Lastly, one area of improvement is we could've explored more models than the three we originally looked at to better predict the yields for the corn. Overall, the corn yield calculator will make a local impact for farmers.

All team members have contributed a similar amount of effort to create the Corn Yield Calculator.

Works Cited

Runge, E. C. A., and John F. Benci. “User Friendly Corn Yield Prediction Model.” *Crop Management*, vol. 7, no. 1, Wiley, 2008, pp. 1–6. *Crossref*, https://doi.org/10.1094/cm-2008-0609-04-rs.

Cerrato, M. E., and A. M. Blackmer. “Comparison of Models for Describing; Corn Yield Response to Nitrogen Fertilizer.” *Agronomy Journal*, vol. 82, no. 1, Wiley, Jan. 1990, pp. 138–43. *Crossref*, https://doi.org/10.2134/agronj1990.00021962008200010030x.

Kaufmann, Robert K., and Seth E. Snell. “A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants.” *American Journal of Agricultural Economics*, vol. 79, no. 1, Wiley, Feb. 1997, pp. 178–90. *Crossref*, https://doi.org/10.2307/1243952.

Irwin, Scott, and Darrel Good. “The Historic Pattern of U.S Corn Yields, Any Implications for 2012? - Farmdoc Daily.” *Farmdoc Daily*, 2 Feb. 2012, farmdocdaily.illinois.edu/2012/02/the-historic-pattern-of-us-cor-1.html.

Schnitkey, Gary. “The Geography of High Corn Yields - Farmdoc Daily.” *Farmdoc Daily*, 4 Jan. 2019, farmdocdaily.illinois.edu/2019/01/the-geography-of-high-corn-yields.html.

Jiang, Zehiu. “DSpace Angular Universal.” *Predicting County Level Corn Yields Using Deep, Long, Short-term Memory Models in the Corn Belt*, 1 Jan. 2018, dr.lib.iastate.edu/entities/publication/f4794e87-8451-4275-a963-68fd060776d8.

Shahhosseini, Mohsen, et al. “Frontiers | Corn Yield Prediction With Ensemble CNN-DNN.” *Frontiers*, 1 Jan. 2001, www.frontiersin.org/articles/10.3389/fpls.2021.709008/full.

Guanyuan, Shuai, and Basso Bruno. “Subfield Maize Yield Prediction Improves When In-season Crop Water Deficit Is Included in Remote Sensing Imagery-based Models - ScienceDirect.” *Subfield Maize Yield Prediction Improves When In-season Crop Water Deficit Is Included in Remote Sensing Imagery-based Models - ScienceDirect*, 7 Feb. 2022, www.sciencedirect.com/science/article/pii/S0034425722000529.

Thessen, Greg. “DSpace Angular Universal.” *Predicting Yield Before Harvest: How Does the USDA Forecast Corn and Soybean Yield?*, 11 Dec. 2008, dr.lib.iastate.edu/entities/publication/0983ec3b-3ae9-4c82-a059-7e896e17f052.

Sun, Jie, et al. “Sensors | Free Full-Text | County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model | HTML.” *MDPI*, 9 Oct. 2019, www.mdpi.com/1424-8220/19/20/4363/htm.

Shahhosseini, Mohsen, et al. “Frontiers | Forecasting Corn Yield With Machine Learning Ensembles.” *Frontiers*, 1 Jan. 2001, www.frontiersin.org/articles/10.3389/fpls.2020.01120/full.

Yuchi, Ma, et al. “Corn Yield Prediction and Uncertainty Analysis Based on Remotely Sensed Variables Using a Bayesian Neural Network Approach - ScienceDirect.” *Corn Yield Prediction and Uncertainty Analysis Based on Remotely Sensed Variables Using a Bayesian Neural Network Approach - ScienceDirect*, 30 Mar. 2021, www.sciencedirect.com/science/article/abs/pii/S0034425721001267.

Ansarifar, Javad, et al. “An Interaction Regression Model for Crop Yield Prediction - Scientific Reports.” *Nature*, 7 Sept. 2021, www.nature.com/articles/s41598-021-97221-7.

Kim, Yang-Won; Nari;Lee. “Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State -Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography | Korea Science.” *Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State -Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography | Korea Science*, 1 Jan. 2016, koreascience.kr/article/JAKO201625752504104.page.

Schlenker, Wolfram, and Michael J. Roberts. “Nonlinear Effects of Weather on Corn Yields.” Review of Agricultural Economics, vol. 28, no. 3, 2006, pp. 391–98. JSTOR, http://www.jstor.org/stable/3877185. Accessed 14 Oct. 2022.

Wu, Di, et al. “Automatic Kernel Counting on Maize Ear Using RGB Images - Plant Methods.” *BioMed Central*, 1 June 2020, plantmethods.biomedcentral.com/articles/10.1186/s13007-020-00619-z.

Fatima A.M. Tenorio, et al. “Benchmarking impact of nitrogen inputs on grain yield and environmental performance of producer fields in the western US Corn Belt”, Agriculture, Ecosystems & Environment, Volume 294, 2020, 106865, ISSN 0167-8809, https://doi.org/10.1016/j.agee.2020.106865.