**Predicting Corn, Wheat and Soybean Yield**

Maura M. Tokay

DATA606 – Delivery-2

Dr. Ergun Simsek

Literature Review

Deep Neural Networks (DNN) was used to predict crop yield in the study conducted by Khaky and Wang [1] in response to the 2018 Syngenta Crop Challenge. DNN outperformed Lasso, Shallow Neural Networks (SNN) and Regression Tree. They also found that environmental factors (weather) had greater effect on crop yield than genotype and soil composition. For my study, genotype and soil composition is not a variable because all measurements were made with the same soil condition and seed variety. I would like to be able to quantify each weather variable importance on crop yield outcome. Crane-Droesch [2] also used DNN to predict crop yield and climate change impact assessment in agriculture. In his study he included growing degree days (GDD) as a feature in his model which ended up having a great impact on crop yield prediction. GDD "are used to estimate the growth and development of plants and insects during the growing season. The basic concept is that development will only occur if the temperature exceeds some minimum development threshold, or base temperature (TBASE). The base temperatures are determined experimentally and are different for each organism". [3]

GDD formula for corn and soybean in °C:

GDD = ((Daily Max Temp + Daily Min Temp) / 2) - 10

GDD formula for wheat in °C:

GDD = ((Daily Max Temp + Daily Min Temp) / 2) - 4.4

I will try to implement DNN in my study but I am not sure if it will work with my data because my labeled data probably is not large enough to produce good results. I signed up on the Syngenta Challenge website hoping to be able to access the 2018 data to use in my model but unfortunately the data is not available.

Preliminary Exploratory Data Analyses

This dataset is part of the Farming System Project (<https://www.ars.usda.gov/northeast-area/beltsville-md-barc/beltsville-agricultural-research-center/sustainable-agricultural-systems-laboratory/docs/farming-systems-project/>) at USDA, Beltsville MD. This data is not available online on the USDA website but can be found on my GitHub (<https://github.com/mmtokay/DATA606/tree/master/datasets>).

The data is split in two files, one that contains crop information and other with weather data.

Crop file:

* Crop - wheat, corn or soybean
* GrowingSeason - year crop was cultivated
* SystemName - crop management (traditional: NT, CT, 3 and 4; organic: Org2, Org3 and Org6)
* GrainYield - grain yield measured in kg/ha
* PlantingDate - date seeds were planted
* HarvestDate - date crop was harvested

Weather file:

* Year
* Julian Day
* Month
* Day
* Date
* avgtTempC - average temperature in C
* maxTempC - maximum temperature in C
* minTempC - minimum temperature in C
* maxHumPct - maximum humidity in %
* minHumPct - minimum humidity in %
* avgRadWm-2 - average radiation in w/m2
* meanWindMs-1 - mean wind in m/s
* PrecipitationMm - precipitation/snow melt in mm

After analyses, these are some finds about the data:

1. There is no crop data for 1999, this year we had a drought in Maryland and because the project didn’t use irrigation, crops never matured.
2. Wheat doesn’t have crop data for 1996, 1999, 2003, 2004, 2007, 2010.
3. Data was separated by crops: 407 labeled data for corn, 524 labeled data for soybean and 235 labeled data for wheat.
4. Average radiation will not be used because data is missing for years 2003-2008. I need to check the anomalies for maximum humidity and mean wind and how this can be corrected.

My next step is to work on feature engineering, I will combine crop management type in two categories traditional and organic. After I determine the average number of weeks that takes for each crop to mature I will calculate weather variables weekly average starting from planting date.

# References

1. Khaki, Saaed, and Lizhi Wang. “Crop Yield Prediction Using Deep Neural Networks.” *Frontiers*, Frontiers, 26 Apr. 2019. Retrieved February 25, 2020 from https://[www.frontiersin.org/articles/10.3389/fpls.2019.00621/full](http://www.frontiersin.org/articles/10.3389/fpls.2019.00621/full).
2. Crane-Droesch, Andrew. “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture.” *Environmental Research Letters, Vol 13, Num 11*, 26 Oct. 2018. Retrieved March 1, 2020 from <https://iopscience.iop.org/article/10.1088/1748-9326/aae159>
3. Explanation of Growing Degree Days, Midwestern Regional Climate Center. Retrieved March 1, 2020 from mrcc.illinois.edu/gismaps/info/gddinfo.htm.