**TEAM NAME**: Team Miracle Grow

**MEMBERS**: Ryan Rizzo, Cole McKinley Derek Colombo, Colby Chambers

**PURPOSE**: This repository was made specifically for the MCLA2021 competition. This group participated in MCLA2021 for the purpose of their midterm project in an ISTA class titles “Applied Concepts Cyberinfrastructure.” The contributors of this project include members of this class.

Additional information on this Machine Learning competition can be found here: <https://eval.ai/web/challenges/challenge-page/1251/overview>

Prize Amounts: 1st: $2000, 2nd $1500, 3rd $1000

**RESOURCES:**

*Inputs\_weather\_train.npy:*

For each record, daily weather data - a total of 214 days spanning the crop growing season (defined April 1 through October 31). Daily weather records were compiled based on the nearest grid point from a gridded 30km product. Each day is represented by the following 7 weather variables -

* Average Direct Normal Irradiance (ADNI)
* Average Precipitation (AP)
* Average Relative Humidity (ARH)
* Maximum Direct Normal Irradiance (MDNI)
* Maximum Surface Temperature (MaxSur)
* Minimum Surface Temperature (MinSur)
* Average Surface Temperature (AvgSur)

*Inputs\_others\_train.npy:*

* Maturity Group (MG)
* Genotype ID
* State
* Year
* Location for each performance record.

*Yield\_train.npy*

* Yearly crop yield value for each record.

*Inputs\_weather\_test.npy*

Daily weather data for each performance record for a total of 214 days (time-steps).

Each day is represented by 7 weather variables –

* ADNI
* AP
* ARH
* MDNI
* MaxSur
* MinSur
* AvgSur.

Inputs\_others\_test.npy

* Maturity Group (MG)
* Genotype ID
* State, Year
* Location for each performance record.

**REPOSITORY CONTENT:**

* DataFrames.ipynb

This file is one of this groups early attempts at figuring out data-related issues for the project. The purpose of this notebook it to take the data which we are given, then clean and shape it in a way that can be useful later. More specifically, the “input\_weather” files were given to us in a 3-Dimentional array. To be able to use this information we were required to alter the way this data was structured.

* RoughDraftModel.ipynb

This file contains this project original model during the rough draft phase of our collaboration. During this time our group was focused on using a linear-regression model to make the crop predictions. This file is no longer used, however has been left in for reference purposes.

* Finalmodel.py

This file is a python file rather than a notebook. This is the groups final model for the competition. This is the file that is mainly responsible for making our predictions.

* Results.npy

This is a numpy file which contains our results. The result is an array of all our crop predictions. This file is the one which will be uploaded to the competition’s website.

* rowbycol.csv

This csv file is no longer used in this project. The purpose of this csv file was so that this group could get a better understanding of the data and do some of our testing in R rather than python.

**ADDITIONAL INFORMATION**

* More information can be found in a report from last-year’s winners
* <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0252402#sec01>