Predict Grain Production

Team Progress [Train model with small portion of data] - Team 11
First Semester 2023, Artificial Intelligence Project (SWE3032-41)

박혜림 (2019315581), 박정훈 (2016313844), 임동준 (2017310326),
정홍준 (2017311985), 이성준 (2017312233), 김병무 (2018313395)

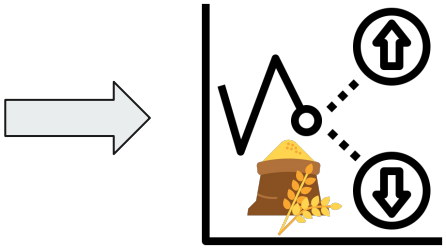
Goal

We will predict the **yearly grain production**.

We collect as many types of data as possible, and then see which ones have a significant impact on grain yield

While food prices are fluctuating due to climate crisis and wars, We hope that our model help people prepare ahead of time by accurately predicting grain production

climate, economics, consumption



predict grain production

Prepare for the food crisis

Dataset

- **Input** (data for prediction)

USDA United States Department of Agriculture

- Annual grain production
- Annual grain production area
- Annual consumption for grain & food

- Grain price
- Fertilizer price

NASA - Daylight hours

NATIONAL WEATHER SERVICE

- Precipitation
- Temperature

FAO FIAT PANIS FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS - Population

bp - Oil price

- **Output** (data to predict)

Wheat

Rice

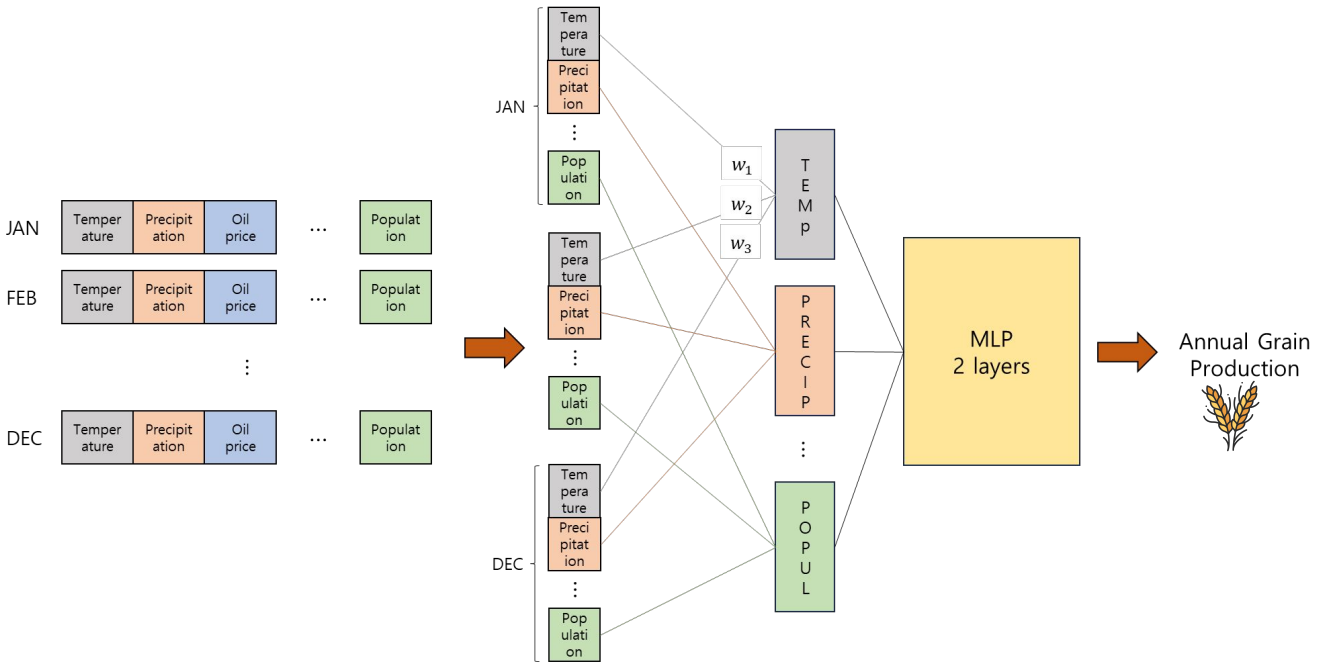
Corn

Dataset

- **Input** (data for prediction)
  - Used only part of the dataset
  - Climate-related data (i.e., sunlight hours, precipitation, temperature) are monthly basis, and other data (e.g., consumption, oil price) are annual basis.
  - Different weights were given depending on how important that month is to growing the grain.
    ex) For wheat, data from August to October is weighted as this period determines wheat production for that year.
  - All input data were flattened to fed into the model.

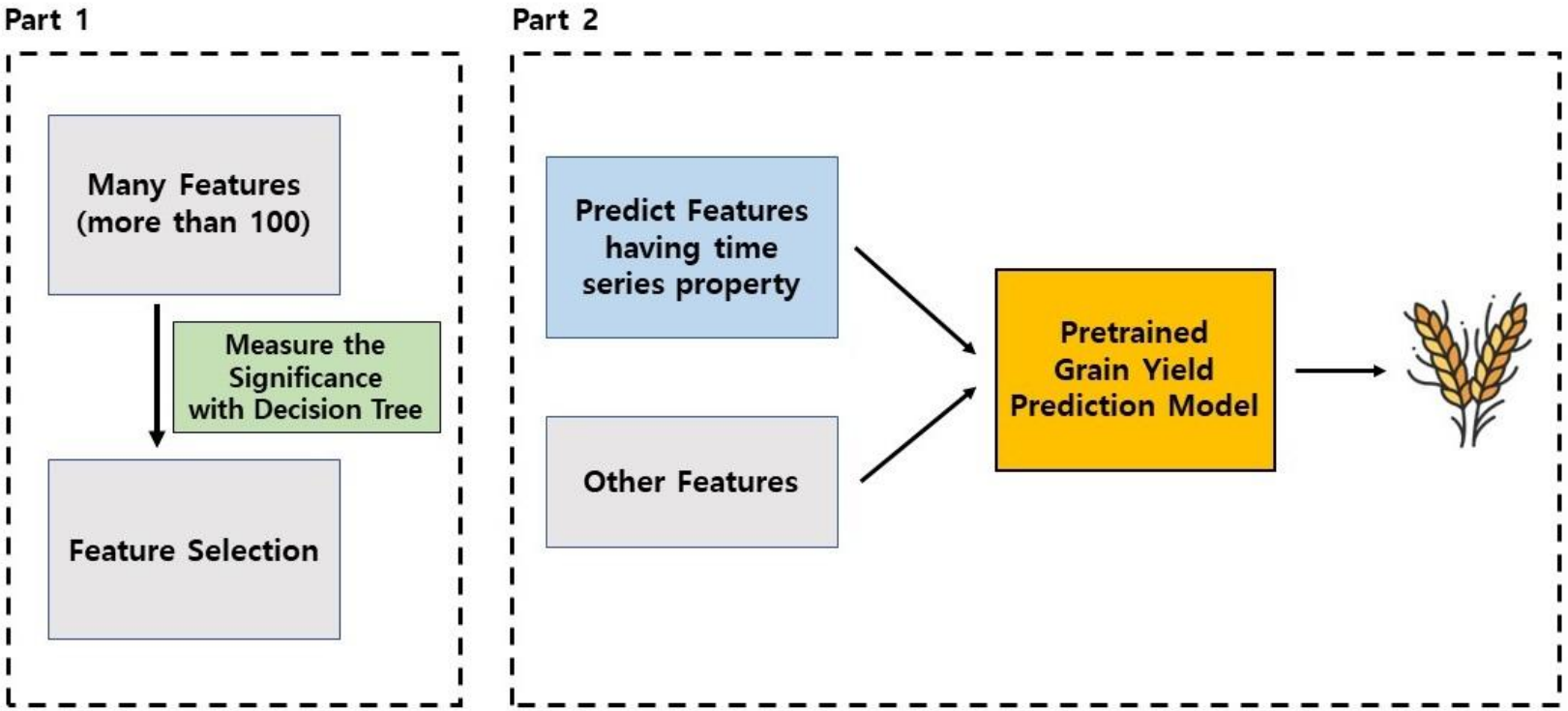
- **Output** (data to predict)
  - Three models predict three grain productions each.

Model Structure

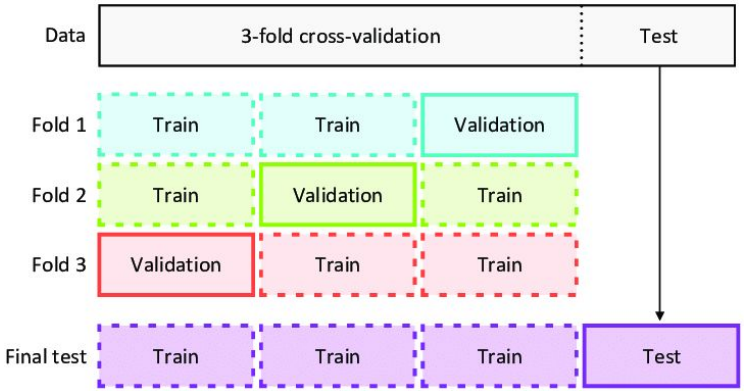


- Flatten all input data
- All features are weighted based on the impact to grain production.
- Model learns interactions between features non-linearly.

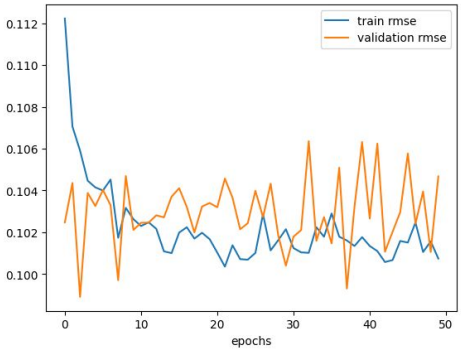
Whole view of Project

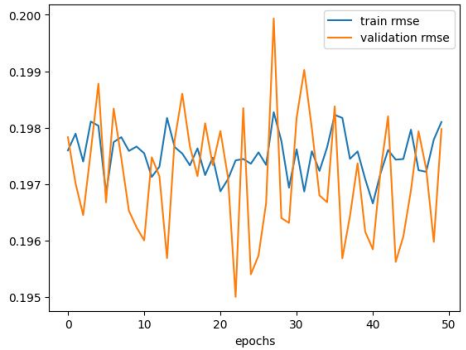


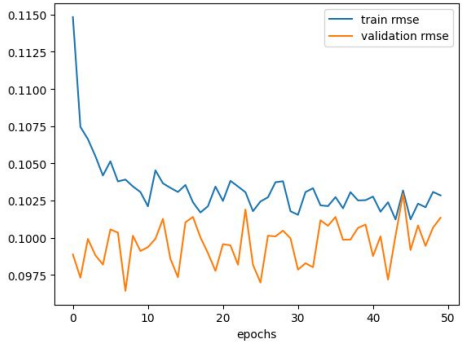
Cross Validation



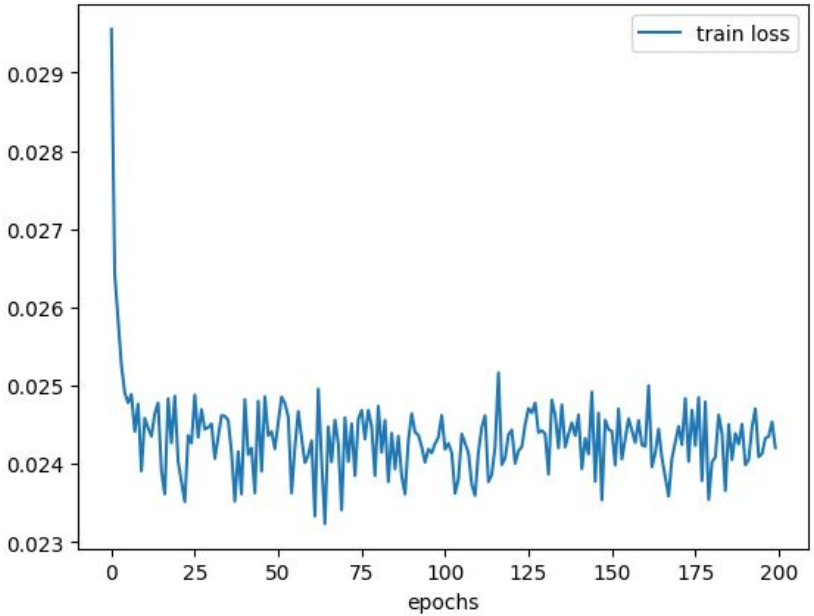
- K Fold is used due to the small size of the dataset.
- Doesn't show a significant improvement.

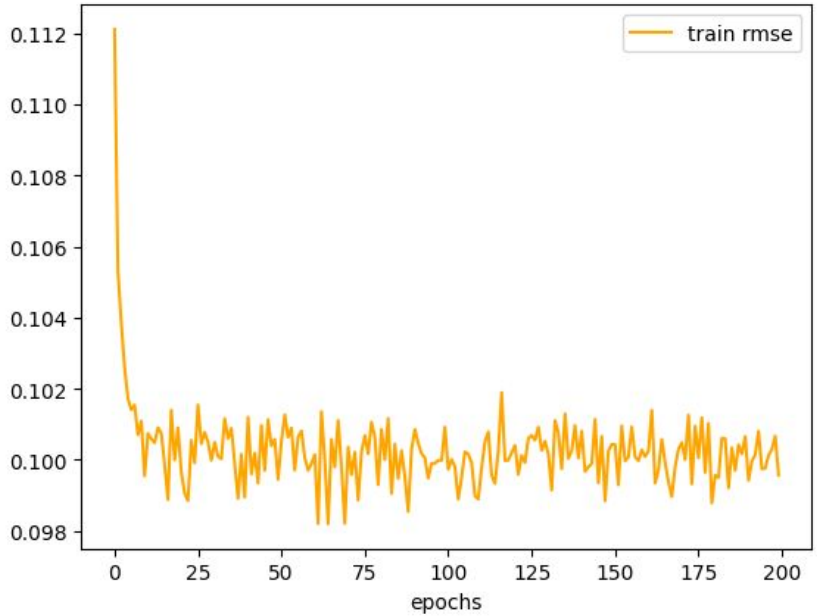






Results of our current model

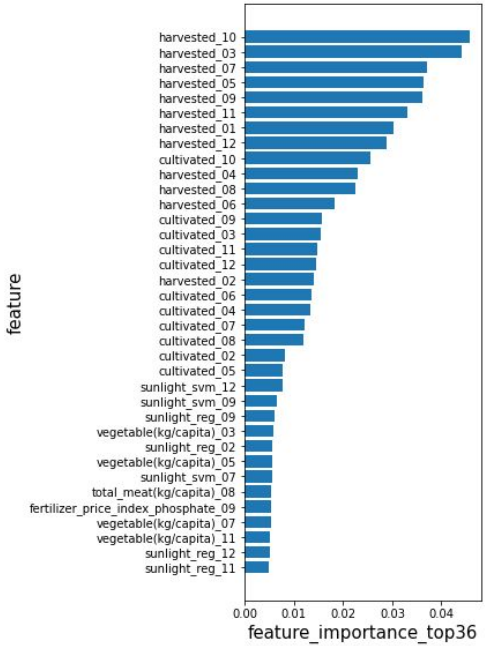


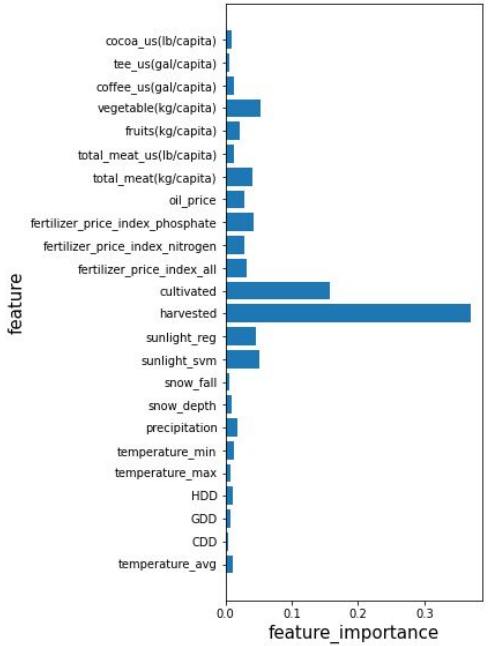


- Overfitting: encountering an interval where increasing the number of epochs no longer reduces the loss.

- Performance: Check the train loss and rmse with validation data

Which features are important?





- Identified features that have a significant impact on results.
- Aside from the obvious correlation between area and yield,
- sunshine, fertilizer prices, and oil price were highly correlated with the yield.
- Interestingly, other vegetable prices were also highly correlated.

Next Plan

1. All-out

- With all the data we've prepared, we'll train a model to predict grain yields.
- With all the types of data we've prepared, we'll also see which data influences the results more.

2. Select And Concentration

- We will retrain the model with data that have a significant impact on the results
- If the important data is climate data, we'll leverage the time series nature of the data to make predictions using an LSTM-based model
- And then use that data to predict future grain production.

Thank you.

Team Progress [Train model with small portion of data] - Team 11
First Semester 2023, Artificial Intelligence Project (SWE3032-41)