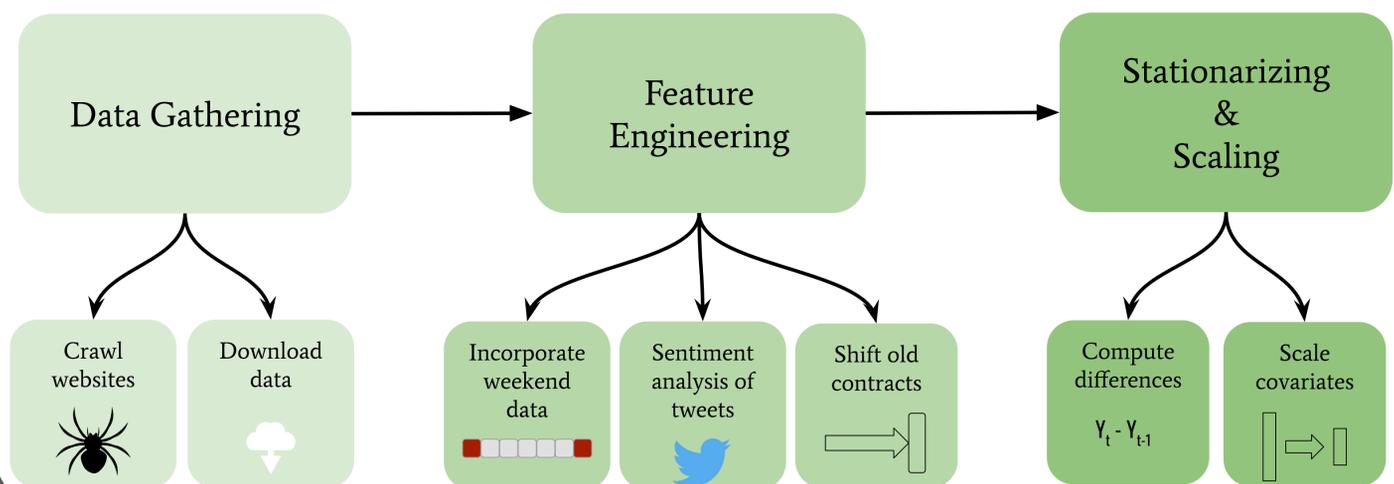# MinneMUDAC Data Science Challenge

Team Name: Women in Math and Stats, G34 | School: University of Minnesota-Twin Cities
Members: Somyi Baek, Cora Brown, Sarah Milstein, Yu Yang | Advisor: Gilad Lerman
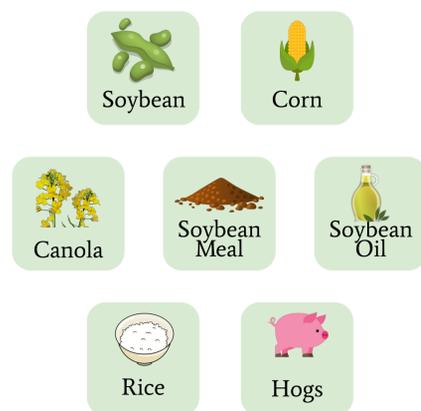
Challenge

- **Objective:** Investigate the factors/characteristics that influence the soybean futures closing prices for 3 different contract months
- **Primary Goal:** Predict soybean closing prices for 5 days: Nov. 4-8, and for 3 contract months: March, May, and July 2020
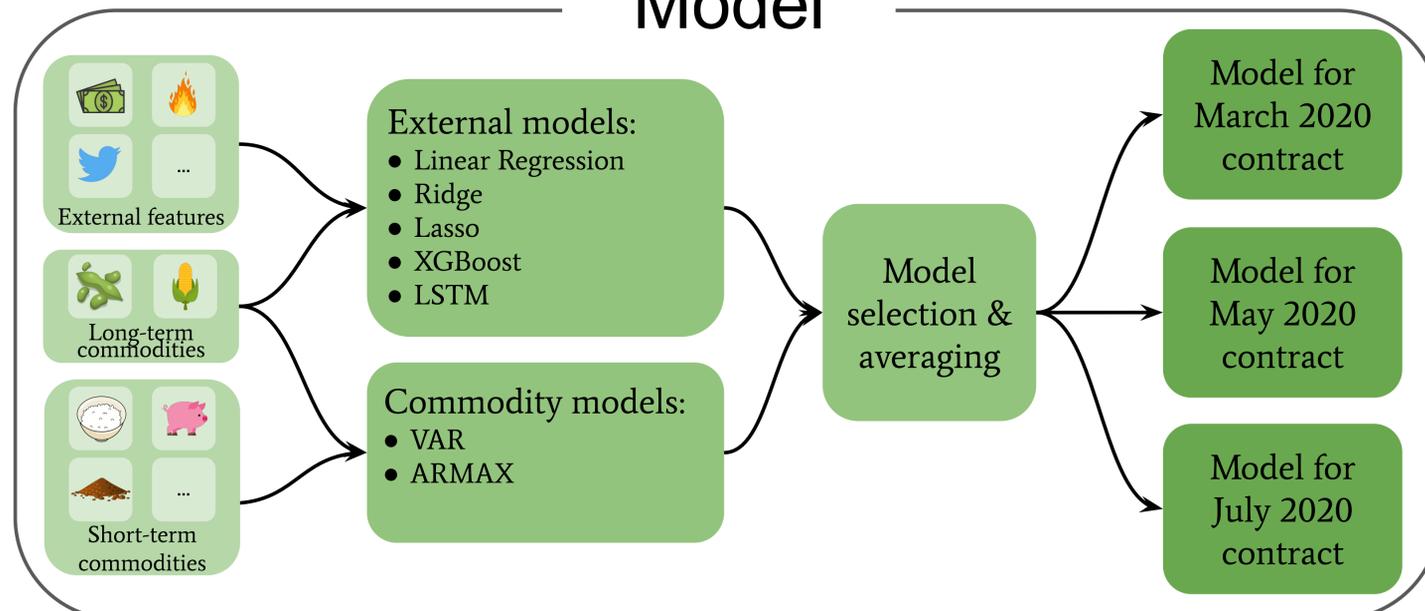
Data Pipeline

Data

Commodities prices

Soybean, Corn, Canola, Soybean Meal, Soybean Oil, Rice, Hogs

External features

Weather, Interest rates, Dow Jones, Tweets, Tariff, Fires, Planting, Disappearance, Stocks

Model

External models:
- Linear Regression
- Ridge
- Lasso
- XGBoost
- LSTM

Commodity models:
- VAR
- ARMAX

Model selection & averaging → Model for March 2020 contract, Model for May 2020 contract, Model for July 2020 contract

Model Interpretation

March 2020: Nov. 4-8 Prediction

	XGBoost	LSTM	VAR
↑ price	Interest Rate ND Weather Dow Jones	Mar 2017 Soybean May 2017 Soybean Argentina Weather	Corn Rice
↓ price	Sunflower Seed Meal Corn Mar 2016 Soybean	MS Weather May 2019 Soybean Jul 2020 Soybean	Soybean Meal Soybean Oil May 2020 Soybean

Results & Conclusions

- Mean absolute error: 4.2167
- Good indicators of soybean prices:
  - Corn
  - Soybean contracts for different months
  - Soybean oil and soybean meal
  - Macroeconomic indicators
  - Weather
- In a 5-day forecast, the predictive power of related commodities outweighs that of random events, such as tweets and tariffs.