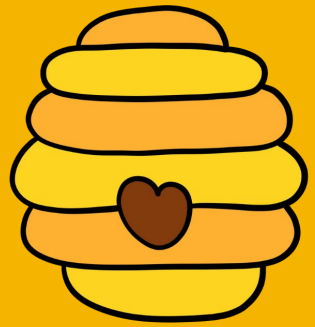
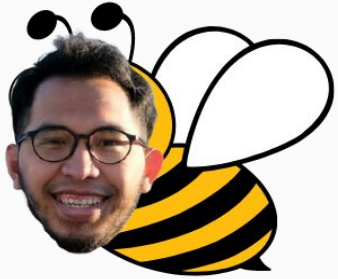


To bee or not to bee:

Hurting honeybee hives in the U.S.

Meet the hive

“Honey” Jack Jacobs



“Bumbly” Irfan Radarma

“Buzzy” Kai Tiede

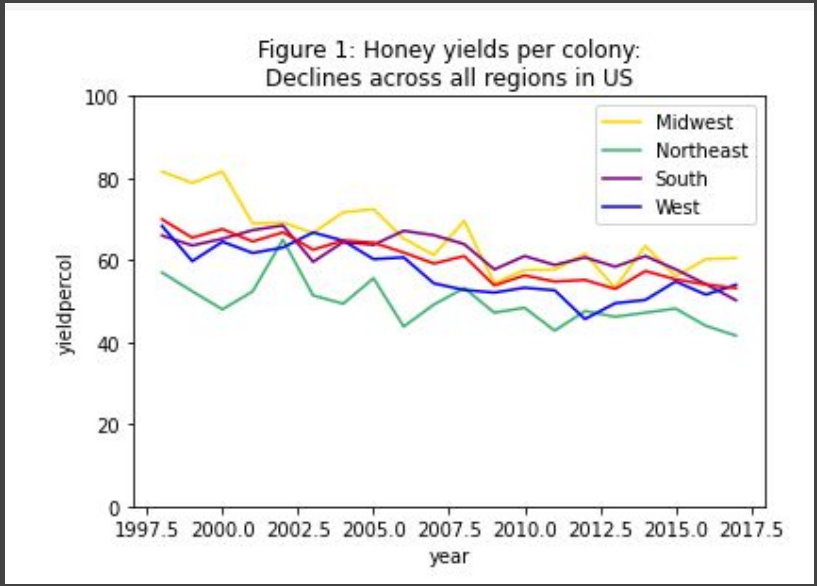
Matt “the drone” Lampl

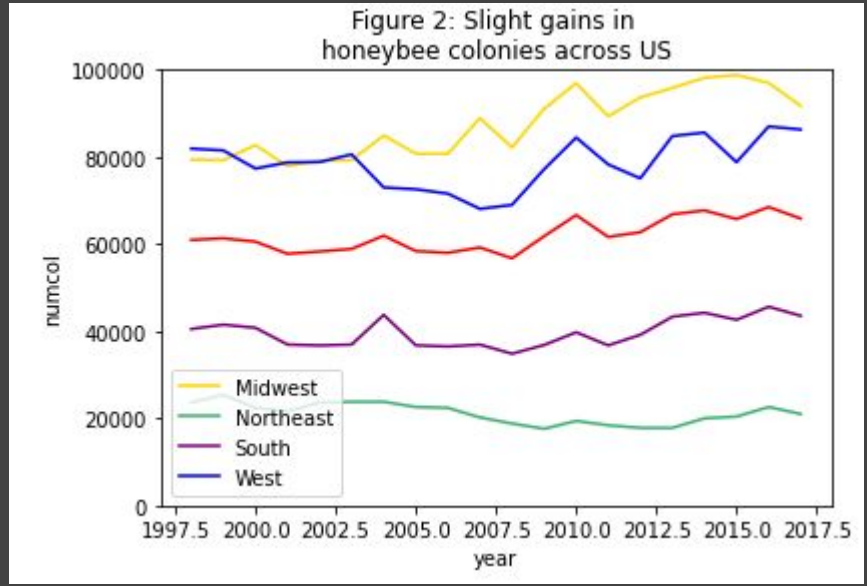
“Stinging” Sara Maillacheruvu

Problem statement:

Hives are hurting

Honeybee colony, production trends





What’s bothering bees?

- Colony collapse
- Climate change
- Pesticides
  - Neonicotinoids
- Parasites
  - Varroa mites
- Poor forage
  - Monoculture cropping

Research method, data summary:

Modeling honeybee decline, understanding data limitations

Research Question

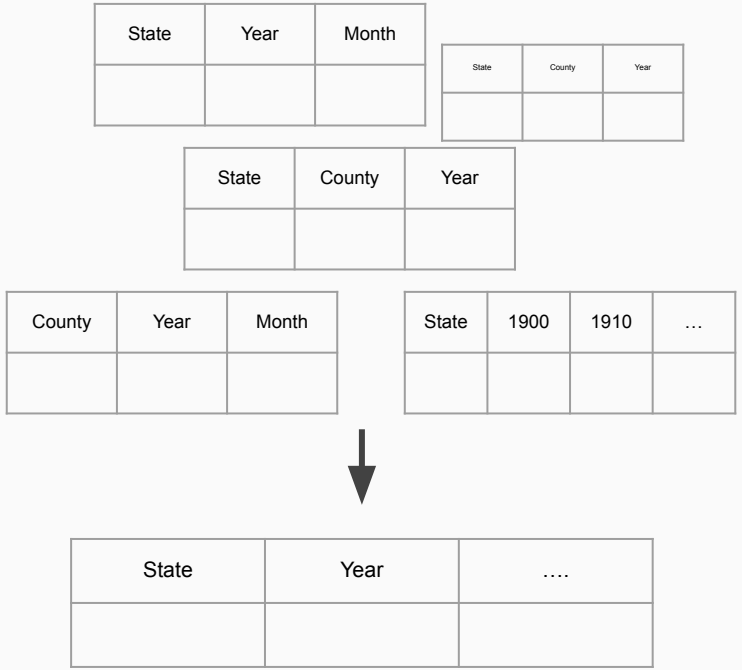
What are the properties/antecedents of honeybee decline in the US?

Dataset

Dataset	Source	Description
APHIS	National Honey Bee Survey	factors that can impact honeybee populations, including levels of varroa mites, Nosema spores, and more
Honey Neonic	USGS's Pesticide National Synthesis Project	number of honeybee colonies and honey production levels as well as various toxins that harm honeybee populations.
Temperature	National Oceanic and Atmospheric Administration	historical temperature at the state level
Urbancity	US Census Bureau	level of urbanicity per state across decennial census years
Air Quality	Environmental Protection Agency	county-level information on air quality
Pesticide	US Department of Agriculture	levels of various pesticides, such as Metribuzin and Trifloxystrobin.

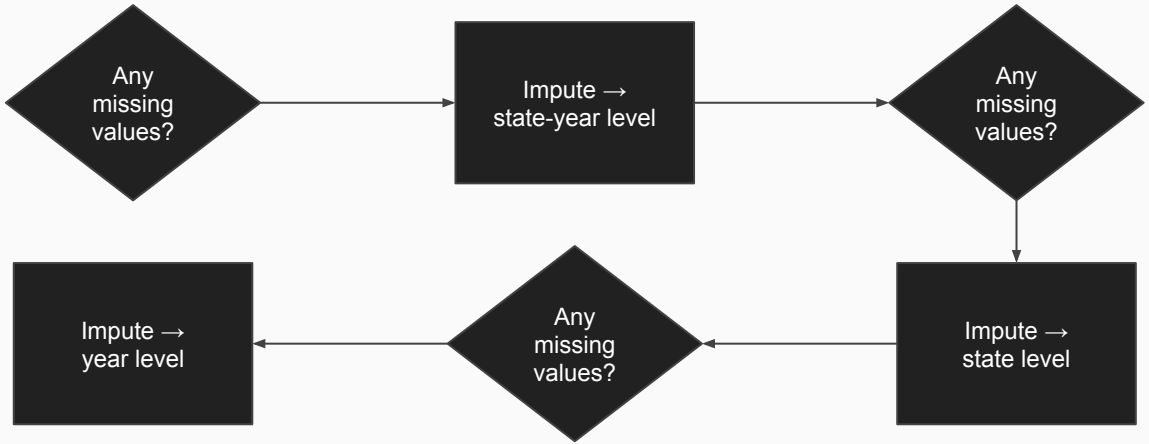
Data limitations

- **Granularity**: State-year honeybee yield per colony levels
- Inherently **local** phenomenon
- Target variable at the state level
  - Nuance lost



Missingness

Missing values were handled using **mean (continues)** and **mode (binary/categorical)** on the following procedure:

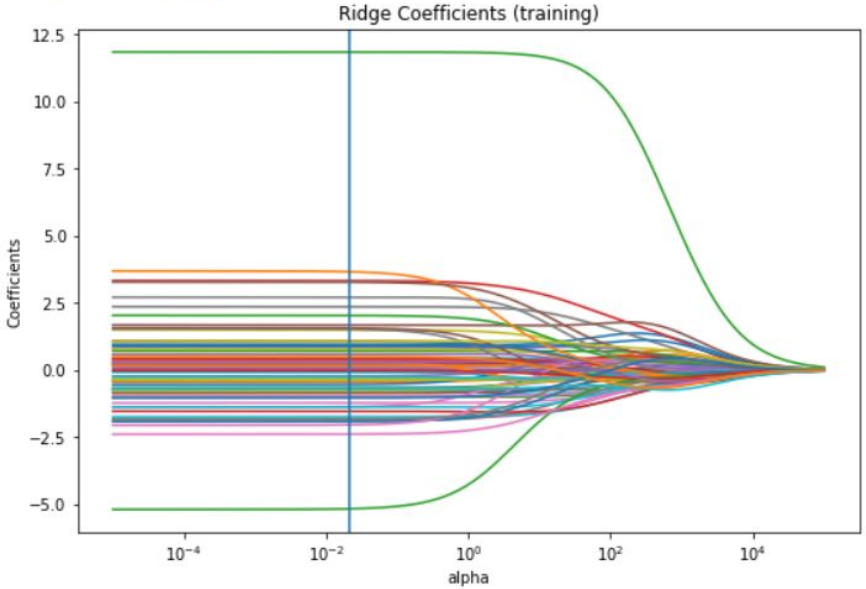


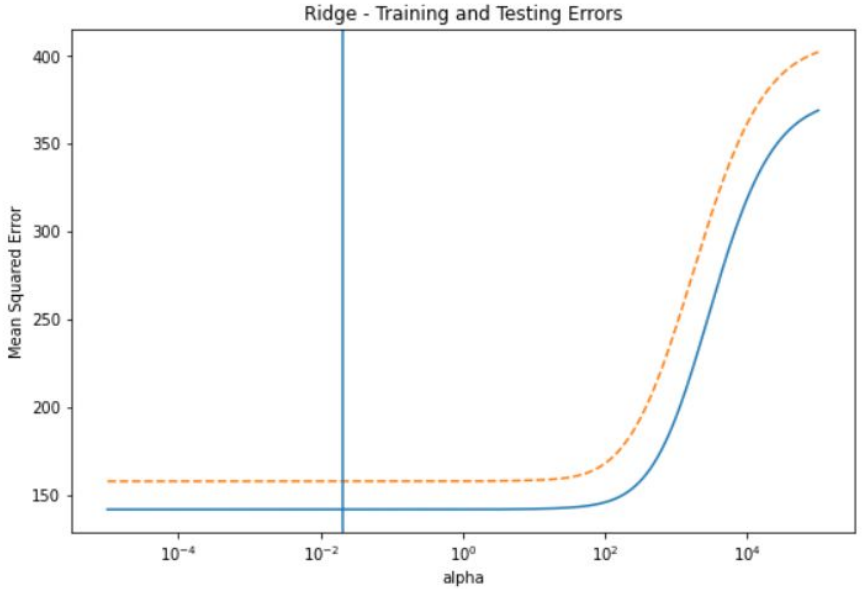
Prediction:

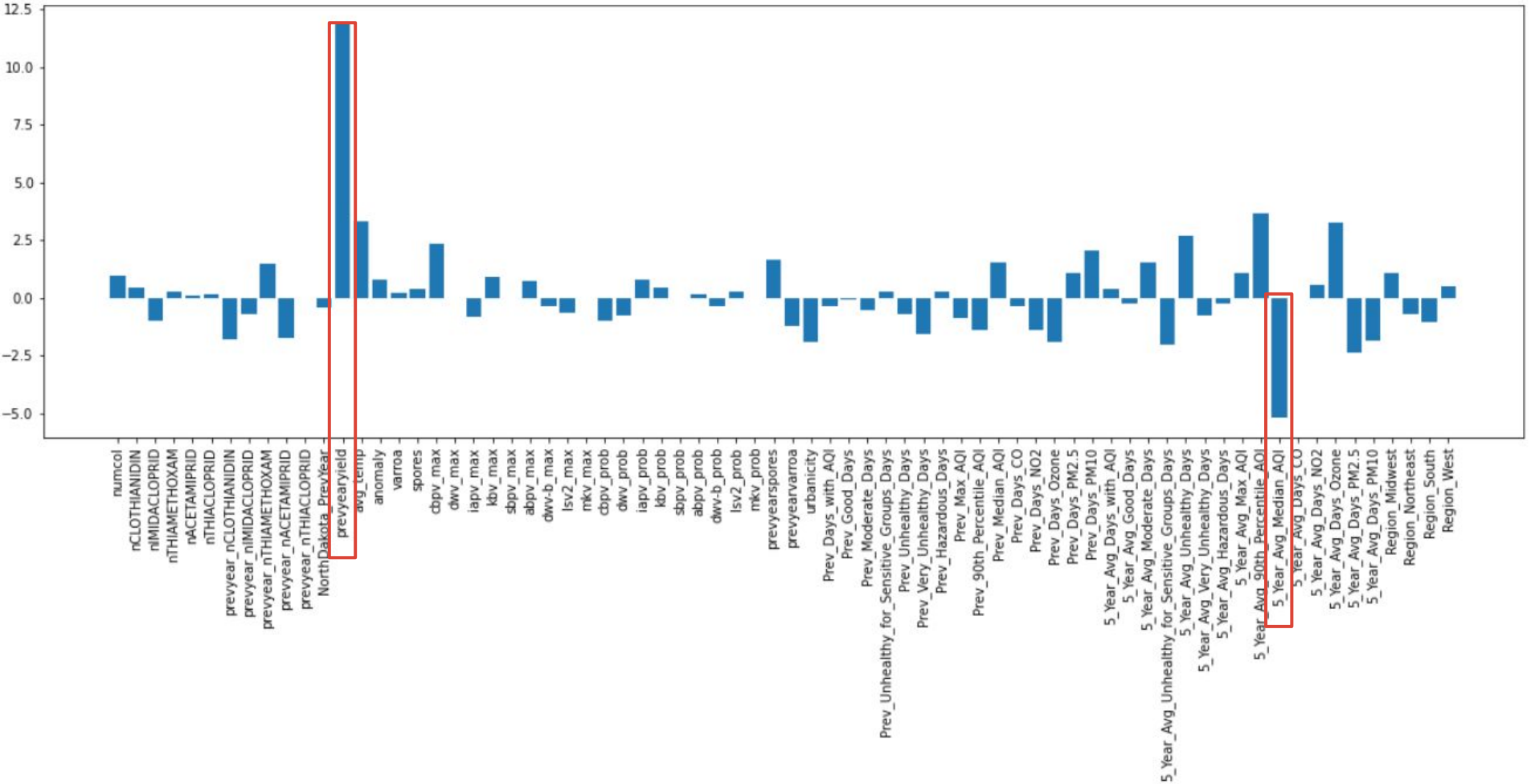
Modeling honey production

Ridge

Alpha: 0.02, Train MSE: 141.92, Test MSE: 157.94, Train RMSE: 11.91, Test RMSE: 12.57

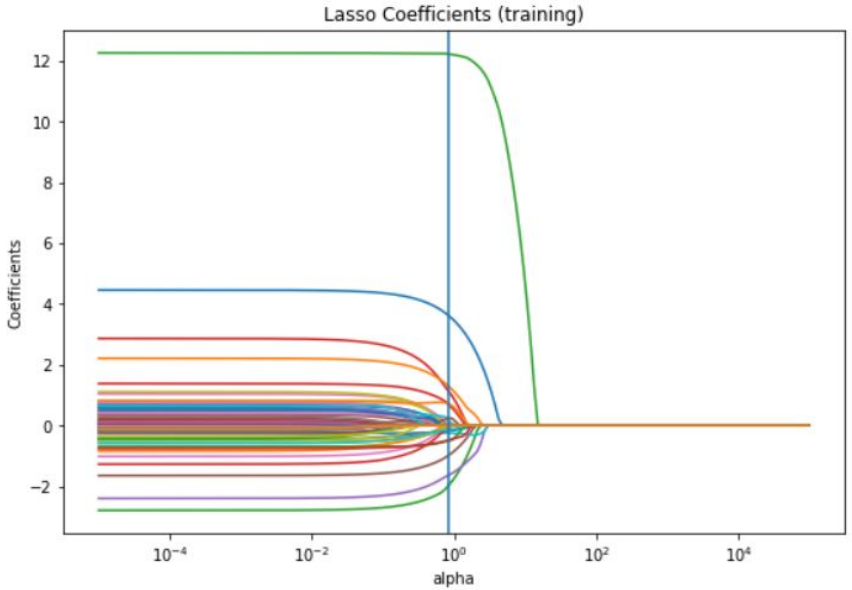


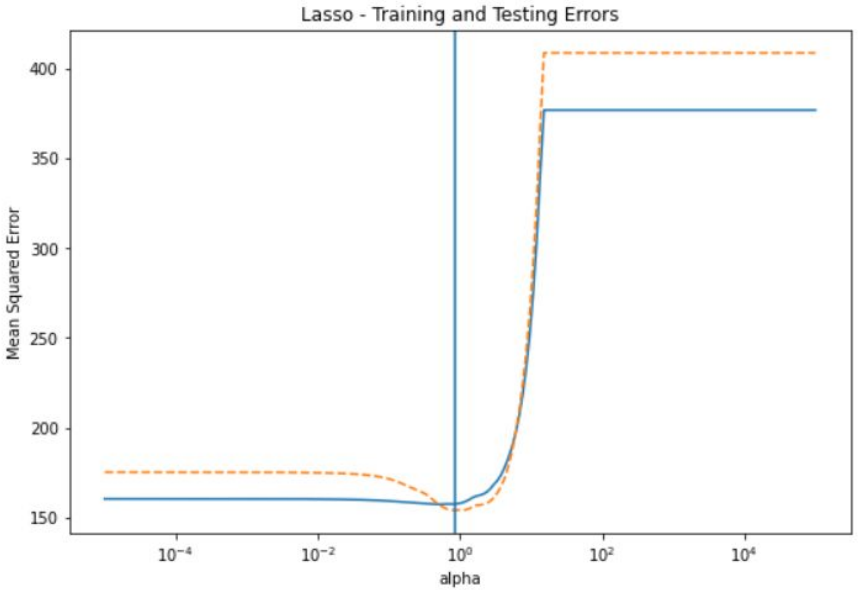


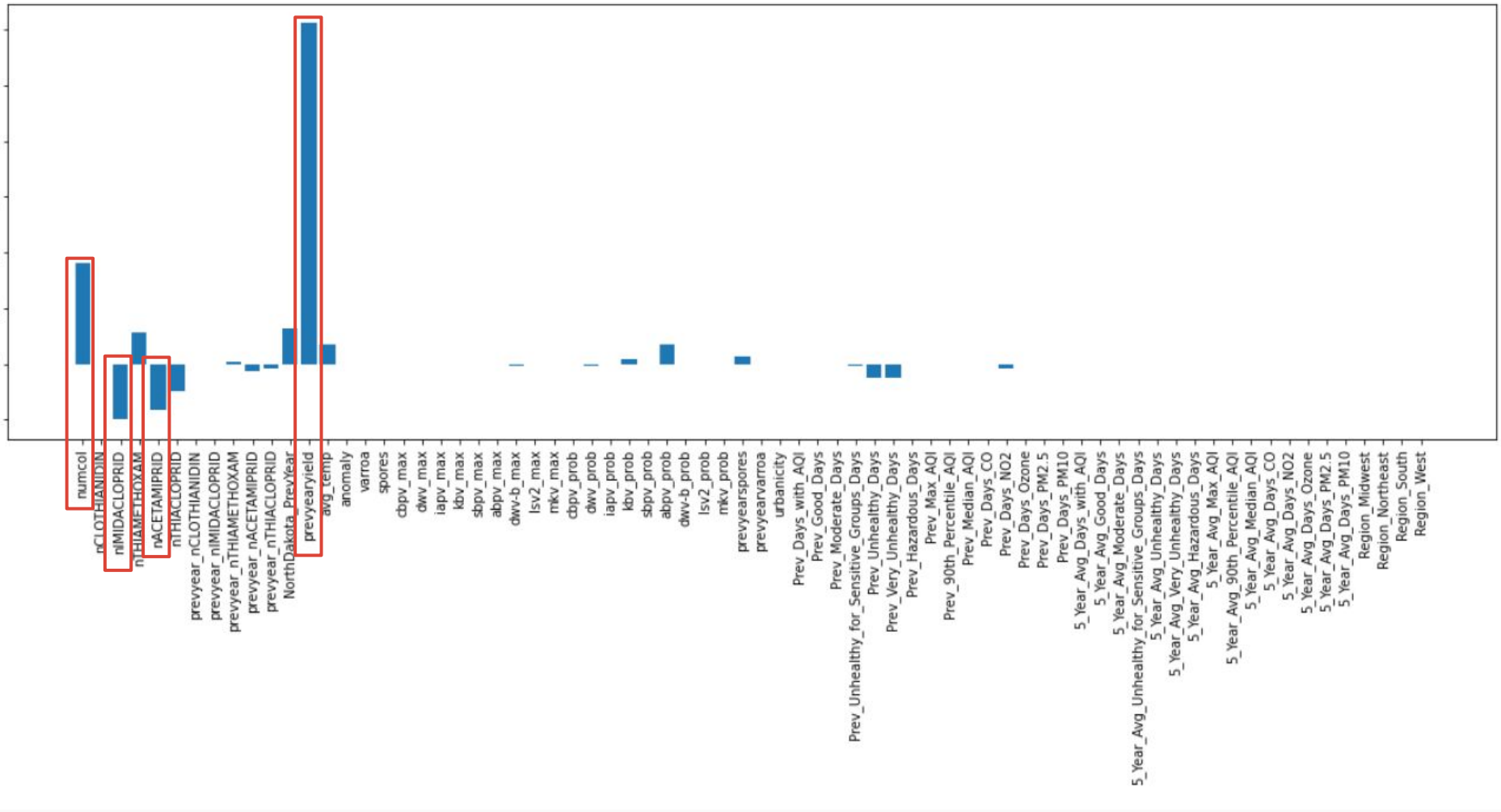

12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
numcol
nCLOTHIANIDIN
nIMIDACLOPRID
nTHIAMETHOXAM
nACETAMIPRID
nTHIACLOPRID
prevyear_nCLOTHIANIDIN
prevyear_nIMIDACLOPRID
prevyear_nTHIAMETHOXAM
prevyear_nACETAMIPRID
prevyear_nTHIACLOPRID
NorthDakota_PrevYear
prevyearyield
avg_temp
anomaly
varroa
spores
cbpv_max
dwv_max
iapv_max
kbv_max
sbpv_max
abpv_max
dwv-b_max
lsv2_max
mkv_max
cbpv_prob
dwv_prob
iapv_prob
kbv_prob
sbpv_prob
abpv_prob
dwv-b_prob
lsv2_prob
mkv_prob
prevyearspores
prevyearvarroa
urbanicity
Prev_Days_with_AQI
Prev_Good_Days
Prev_Moderate_Days
Prev_Unhealthy_for_Sensitive_Groups_Days
Prev_Unhealthy_Days
Prev_Very_Unhealthy_Days
Prev_Hazardous_Days
Prev_Max_AQI
Prev_90th_Percentile_AQI
Prev_Median_AQI
Prev_Days_CO
Prev_Days_NO2
Prev_Days_Ozone
Prev_Days_PM2.5
Prev_Days_PM10
5_Year_Avg_Days_with_AQI
5_Year_Avg_Good_Days
5_Year_Avg_Moderate_Days
5_Year_Avg_Unhealthy_for_Sensitive_Groups_Days
5_Year_Avg_Unhealthy_Days
5_Year_Avg_Very_Unhealthy_Days
5_Year_Avg_Hazardous_Days
5_Year_Avg_Max_AQI
5_Year_Avg_90th_Percentile_AQI
5_Year_Avg_Median_AQI
5_Year_Avg_Days_CO
5_Year_Avg_Days_NO2
5_Year_Avg_Days_Ozone
5_Year_Avg_Days_PM2.5
5_Year_Avg_Days_PM10
Region_Midwest
Region_Northeast
Region_South
Region_West

Lasso

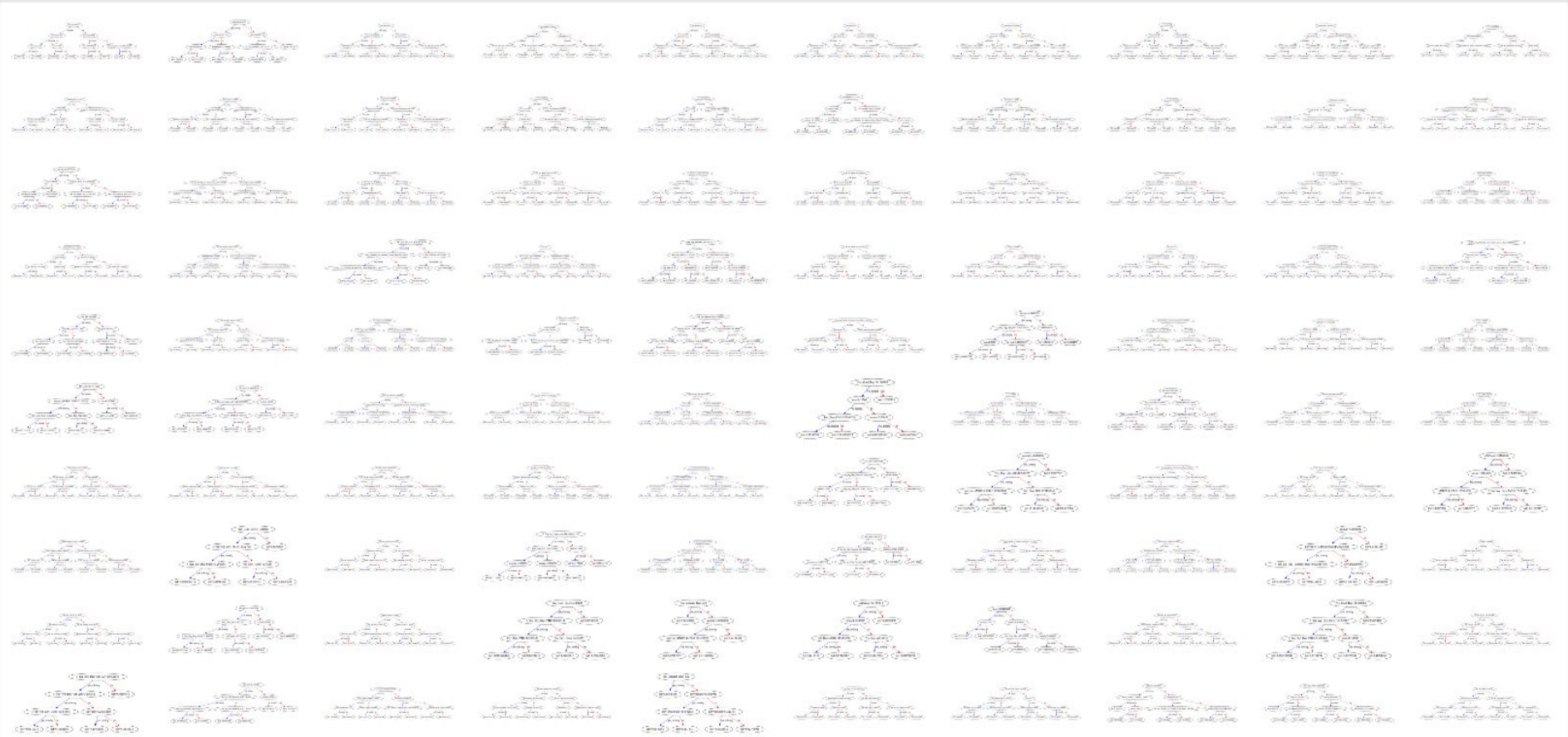
Alpha: 0.84, Train MSE: 157.52, Test MSE: 153.93, Train RMSE: 12.55, Test RMSE: 12.41





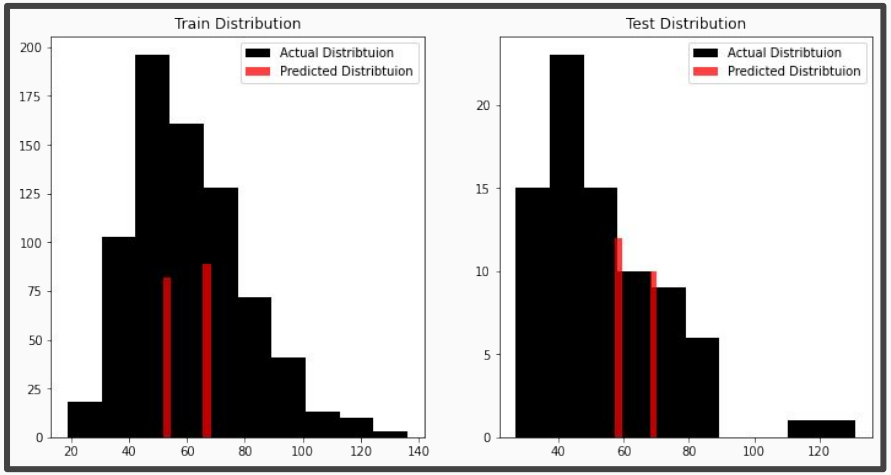

12
10
8
6
4
2
0
−2
numcol
nCLOTHIANIDIN
nIMIDACLOPRID
nTHIAMETHOXAM
nACETAMIPRID
nTHIACLOPRID
prevyear_nCLOTHIANIDIN
prevyear_nIMIDACLOPRID
prevyear_nTHIAMETHOXAM
prevyear_nACETAMIPRID
prevyear_nTHIACLOPRID
NorthDakota_PrevYear
prevyearyield
avg_temp
anomaly
varroa
spores
cbpv_max
dwv_max
iapv_max
kbv_max
sbpv_max
abpv_max
dwv-b_max
lsv2_max
mkv_max
cbpv_prob
dwv_prob
iapv_prob
kbv_prob
sbpv_prob
abpv_prob
dwv-b_prob
lsv2_prob
mkv_prob
prevyearspores
prevyearvarroa
urbanicity
Prev_Days_with_AQI
Prev_Good_Days
Prev_Moderate_Days
Prev_Unhealthy_for_Sensitive_Groups_Days
Prev_Unhealthy_Days
Prev_Very_Unhealthy_Days
Prev_Hazardous_Days
Prev_Max_AQI
Prev_90th_Percentile_AQI
Prev_Median_AQI
Prev_Days_CO
Prev_Days_NO2
Prev_Days_Ozone
Prev_Days_PM2.5
Prev_Days_PM10
5_Year_Avg_Days_with_AQI
5_Year_Avg_Good_Days
5_Year_Avg_Moderate_Days
5_Year_Avg_Unhealthy_for_Sensitive_Groups_Days
5_Year_Avg_Unhealthy_Days
5_Year_Avg_Very_Unhealthy_Days
5_Year_Avg_Hazardous_Days
5_Year_Avg_Max_AQI
5_Year_Avg_90th_Percentile_AQI
5_Year_Avg_Median_AQI
5_Year_Avg_Days_CO
5_Year_Avg_Days_NO2
5_Year_Avg_Days_Ozone
5_Year_Avg_Days_PM2.5
5_Year_Avg_Days_PM10
Region_Midwest
Region_Northeast
Region_South
Region_West

Gradient Boosted Trees



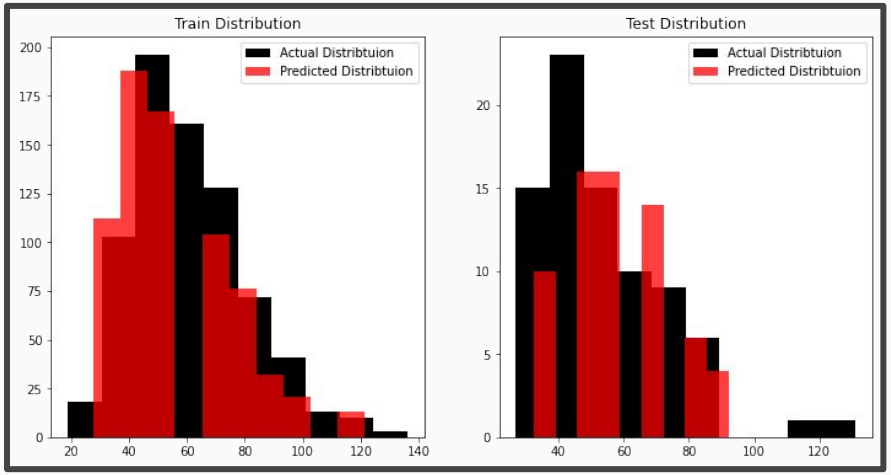
Gradient Boosted Trees Performance

Random Forest with 5000 Trees



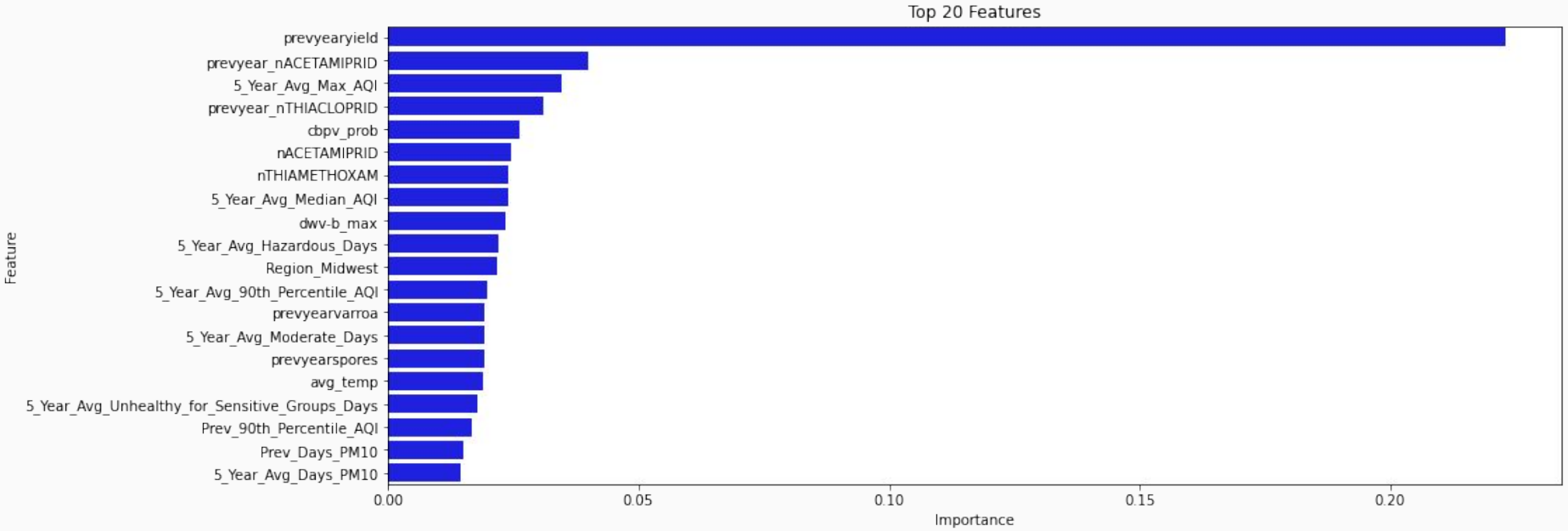
MSE_test: 290.31
MSE_train: 212.45
RMSE_test: 17.04
RMSE_train: 14.58

Gradient Boosted Forest with 100 Trees



MSE_test: 108.96
MSE_train: 54.62
RMSE_test: 10.44
RMSE_train: 7.39

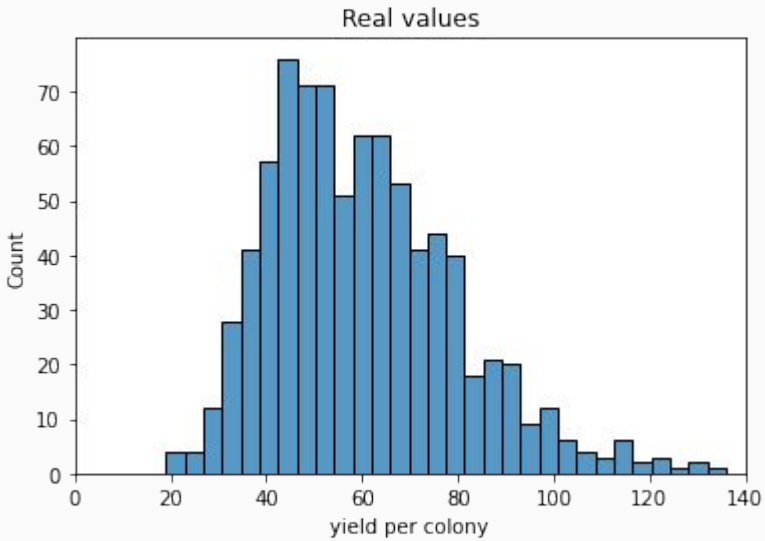
Gradient Boosted Trees Feature Importances

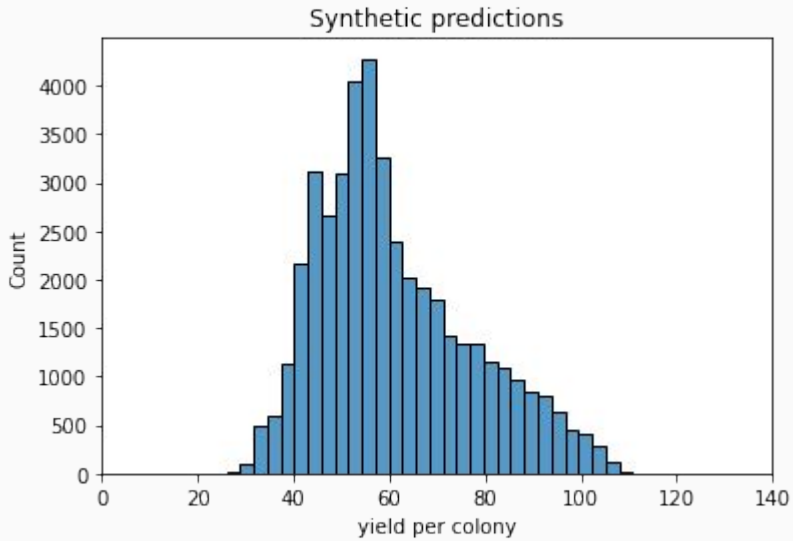


Simulation:

Looking deeper into yield variability

Similar distributions, even with uniform sampling





Insights from our best trained model

- Location, location, location!
  - Hawaii and Louisiana are overrepresented in the best years for colony yield, and the Northeast region is underrepresented
  - Insecticides don’t compensate for this: their use is negatively correlated with high-yield years
- Agricultural colonies appear to be more resilient to pests
  - Previous-year pests positively associated with current-year yield

Analysis and discussion:

Potential policy interventions

Takeaways

- Previous year’s yield is important
- Other important features:
  - 5 Year average AQI
  - Usage of certain neonicotinoid insecticides
- Simulation helps us understand our models’ insights
  - E.g., the primacy of location in per-colony yield

Next Steps

- More granular data
  - More local levels (counties, metropolitan areas)
  - Monthly or weekly observations instead of annual
- Include features measuring extreme weather

Anything abuzz?

Feel free to share thoughts or questions.