FORECASTING AGRO-FOOD CO_2 EMISSIONS UNDER POLICY SCENARIOS

Abby Peck, Abhiroop Kumar, Grace Lin,
Shruthi Chembu, Vyshnavi Maringanti

TABLE OF CONTENTS

01

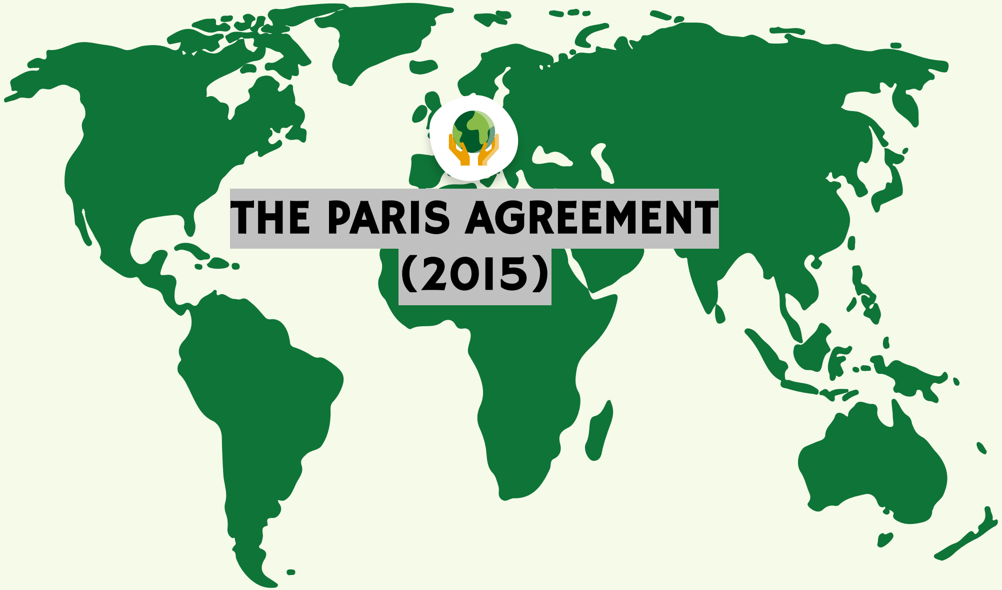
THE PROBLEM

AGRICULTURE'S CLIMATE CHALLENGE

- Agriculture emits **CO_2, CH_4, N_2O** through fires, fertilizers, manure, and land-use change
- Compared to energy or transport, agriculture is **harder to modernize** due to ecological and biological constraints
- This motivates the need for **predictive, scenario-based modeling** to explore future emission trajectories
- What we are attempting:

Agricultural practices → Emissions → Long-term temperature impact

GLOBAL TEMPERATURE BENCHMARKS



BY YEAR 2100

- **< 2°C** warming above pre-industrial levels
- Aim for **1.5°C**

AGRICULTURE IS BOTH:

- A major emitter, and
- A potential carbon sink

AGRI-FOOD CO_2 EMISSION DATASET

~7,000 rows × ~31 columns of country-year agricultural emission records.

INCLUDES:

- *Burning emissions:* Savanna fires, Forest fires
- *Crop processes:* Rice cultivation, Crop residues
- *Land-use:* Net forest conversion, Forestland
- *Industrial agriculture:* Fertilizer manufacturing, Pesticide production

VECTORS:

- Covers both emission flows and production activity indicators.
- Source: **Kaggle dataset by Alessandro Lobello**

TARGET: *total_emission*

"*What-If*" AGRICULTURAL FUTURES

BAU

BASELINE

Business-as-usual, assuming current trajectories continue

MODERATE

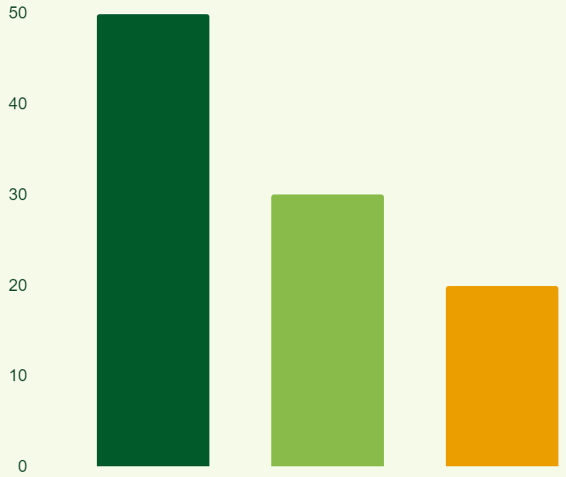
SCENARIO A

- Mild reductions to key emission-related features
- Represents achievable, policy-friendly agricultural reform

AMBITIOUS

SCENARIO B

- Large reductions across the same features
- Represents high-ambition Paris-aligned transformation



Objective: Understand how agricultural changes ripple through

Emissions → Scenario-based reductions → Climate modeling

02

THE SETUP

CLEANING, QUALITY AND PREPARATION

Data Integrity Checks:
Standardized dtypes,

- Area → categorical
- Year → integer

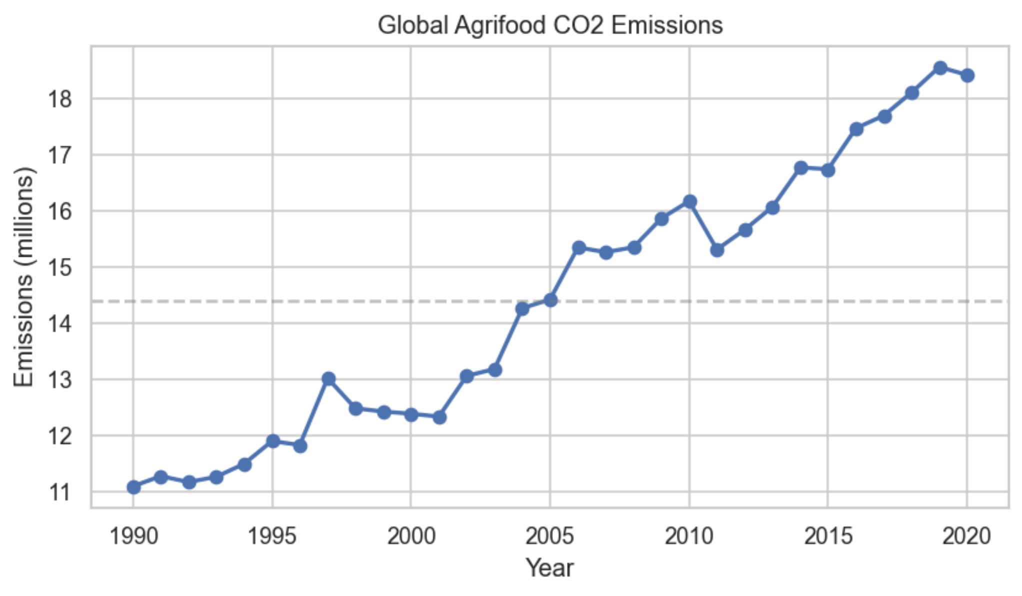
Several emission variables have close to 20–29% missingness:
3-stage imputation pipeline (country & time-aware)

1. Within each country:
Interpolate across years

2. Country-level fill:
Remaining gaps → country means

3. Fallback:
Global median fill

WHAT DOES THE DATA ACTUALLY LOOK LIKE?



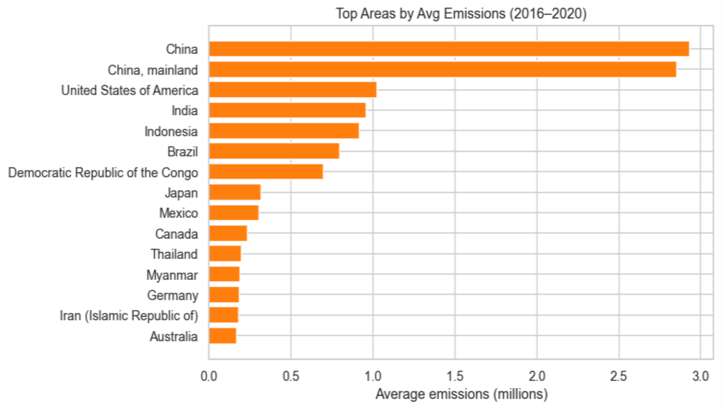
Key Insights

- **236 areas** with annual records from 1990–2020
- Global agrifood CO_2 emissions **increased ~66%** since 1990
- **Clear upward trend**, with minor downward patterns in-between

WHERE DO EMISSIONS COME FROM?

GEOGRAPHIC PATTERNS

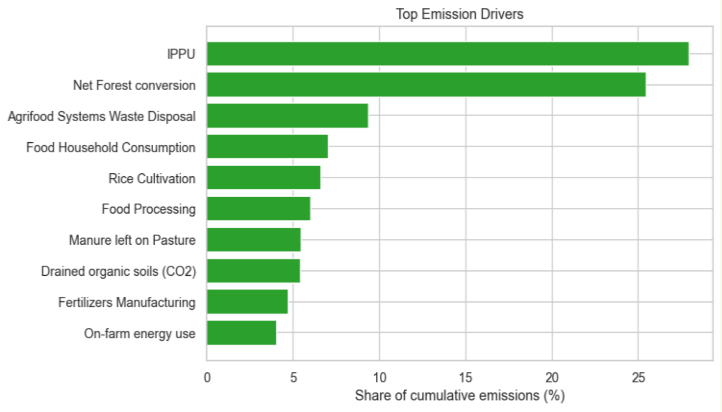
- Top 5 emitting countries together account for **only ~9.6%** of global emissions
- Rest of the emissions are widely spread across countries, not dominated by a few



COMPOSITION PATTERNS

- Industrial & land-use sources dominate:

IPPU + *Net Forest conversion* contribute **>50%** of total emissions



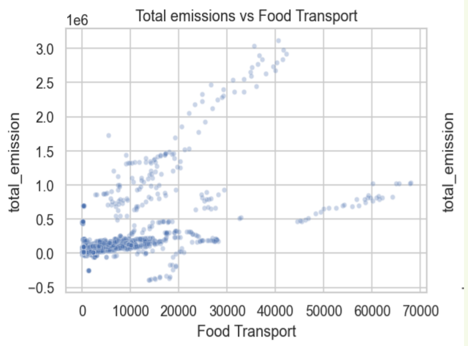
WHAT CORRELATES WITH EMISSIONS?

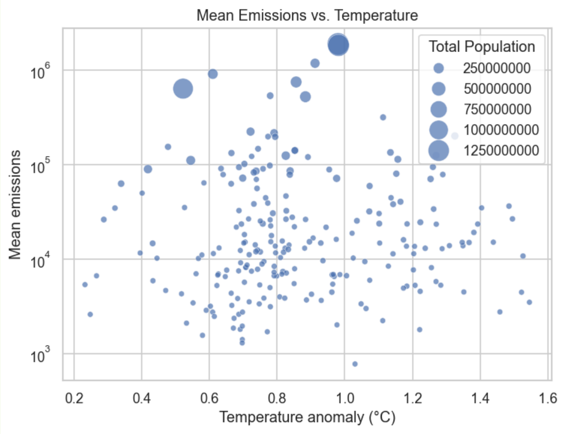
Correlations:

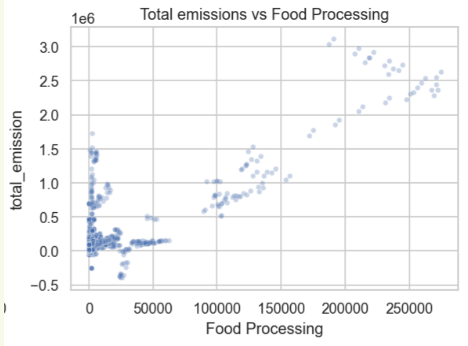
- **Emissions vs Population:** strong (**r ≈ 0.86**)
- **Emissions vs Temperature:** almost zero (**r ≈ -0.01**)

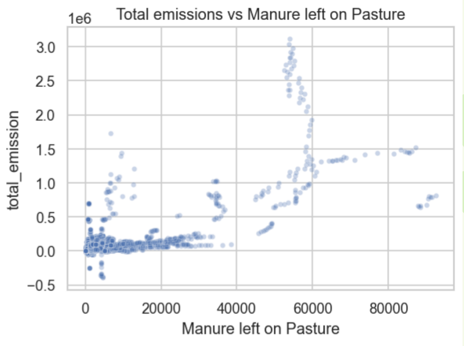
Driver Correlations:

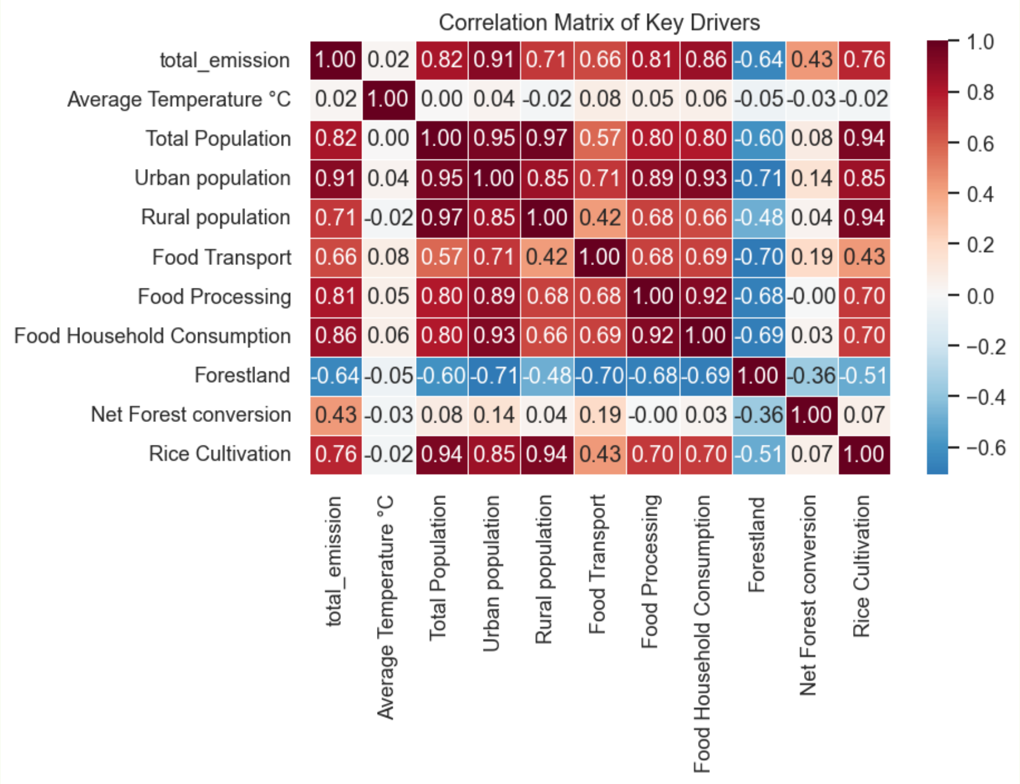
- **Food Transport** → strong positive association
- **Food Processing** → strong positive association
- **Manure left on Pasture** → moderate positive association











ARE THE DRIVERS CORRELATED?

Transport, processing, and household consumption show strong positive ties to total emissions,
while *Forestland* is inversely related to *Net Forest conversion*, capturing the reforestation vs. deforestation tug-of-war that can inform scenario modeling.

DISTRIBUTIONS, SKEW & OUTLIERS

DISTRIBUTION FINDINGS

- Heavy **right-skew** in major drivers (*total_emission, IPPU, Net Forest conversion*, etc.)
- **Log transformations** to stabilize variance

OUTLIER ANALYSIS

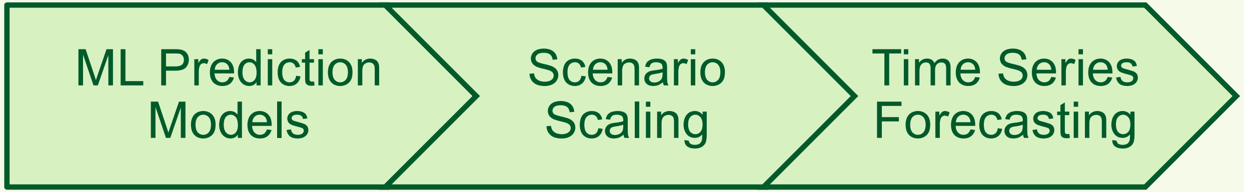
- **Extreme values in land-use variables**, but are legitimate (real events)
- No implausible population or temperature values

03

THE WORK

THE PIPE OF RESEARCH

1. **ML Modeling:** Select the best prediction model by trying several methods
2. **Scenario Linear Scaling:** Scale Agrofood Variables into three scenarios
3. **Time Series:** Forecast future CO_2 by the three scenarios (trained and tested by past data)



END GOAL

Forecast future CO_2 emissions using Agrofood Variables

THE MODELS

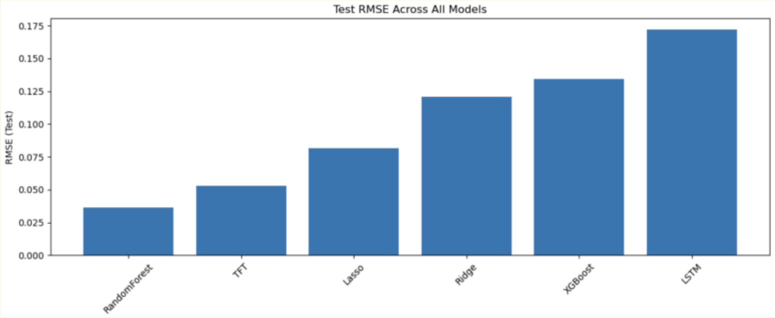
ESTABLISHING THE BASELINE (Ridge, Lasso, RF and XGBoost)

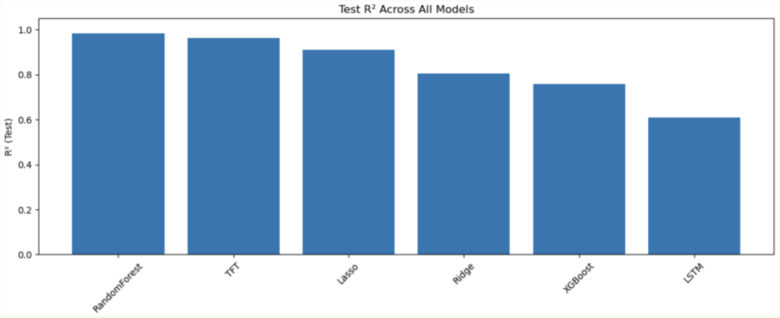
- **Handling Multicollinearity:** Ridge/Lasso penalize correlated features
- **Capturing Non-Linearity:** Tree ensembles (XGBoost) identify saturation points in agricultural inputs
- **Feature Importance:** Identification of high-impact drivers (Forestland, Forest fires)

CAPTURING TEMPORAL DYNAMICS (LSTM & TFT)

- **Sequential Dependencies:** LSTMs capture long-term lags
- **Temporal Fusion Transformer (TFT):**
    - Explicitly separates static metadata from time varying inputs
    - Automatically selects relevant features while suppressing noise via gating
    - Uses attention mechanisms to identify significant historical time steps
    - Interpretable weights reveal which past events drive current predictions

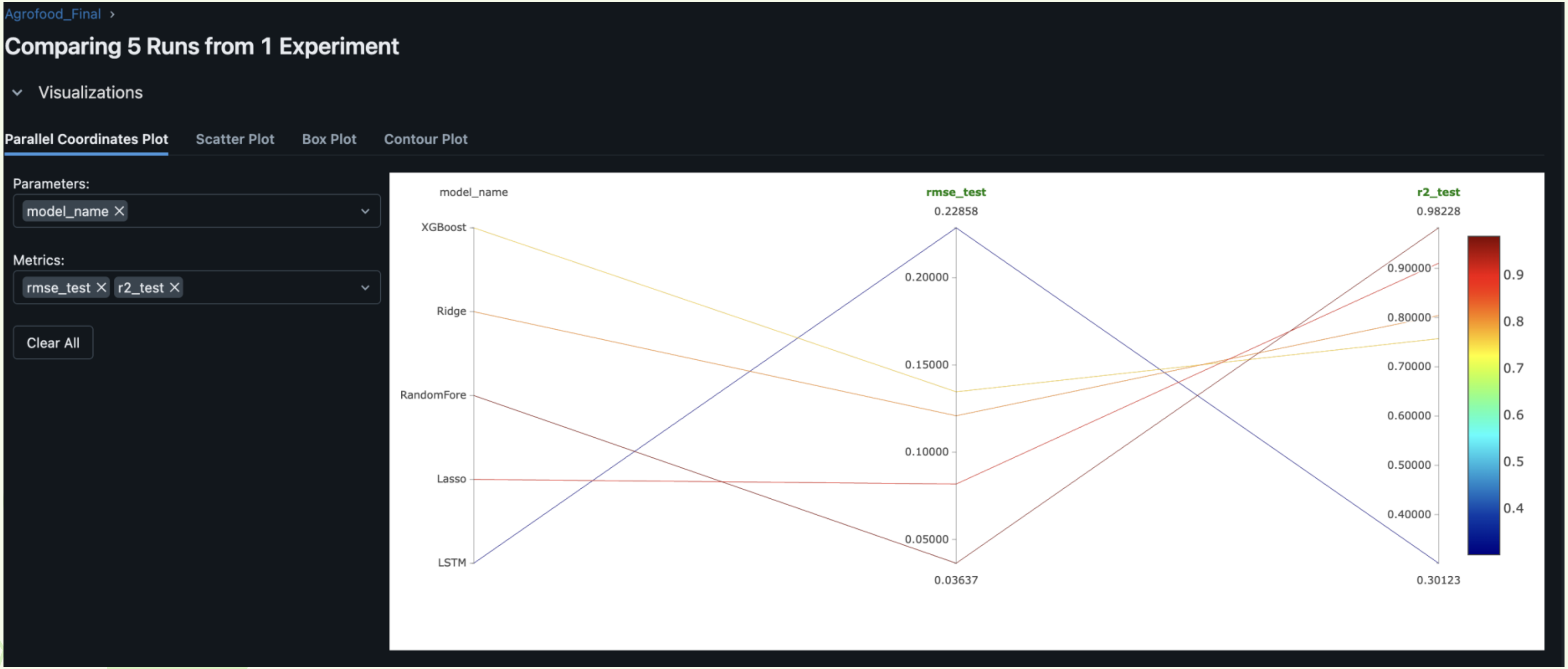
HOW DID THE MODELS PERFORM?





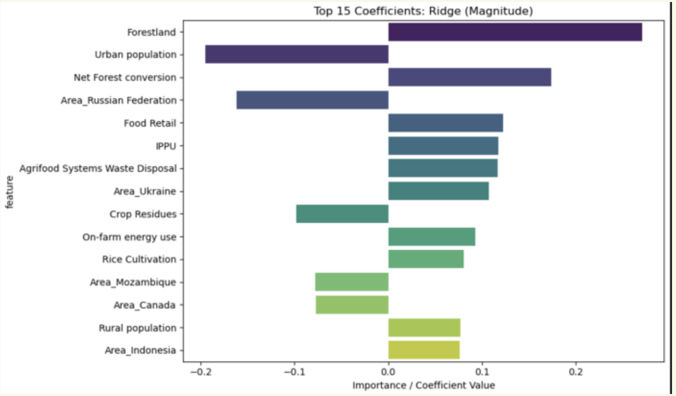
- **Random Forest performed the best** of all the models, with TFT as a close second
- All the models were all tuned to find the **best hyper parameters**
- The models were trained on **data from 1990-2014** and **tested on data from 2014-2020**

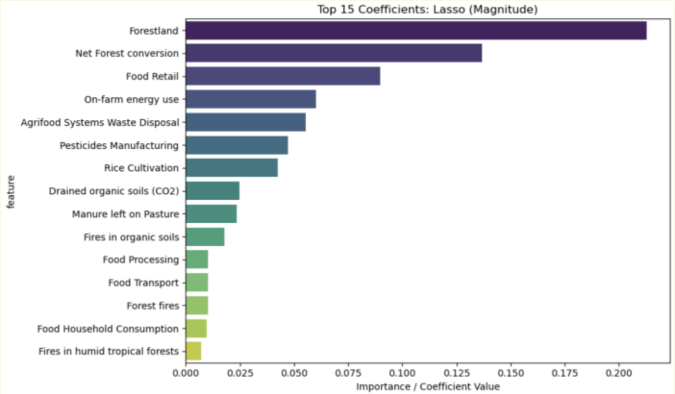
MLFLOW SS

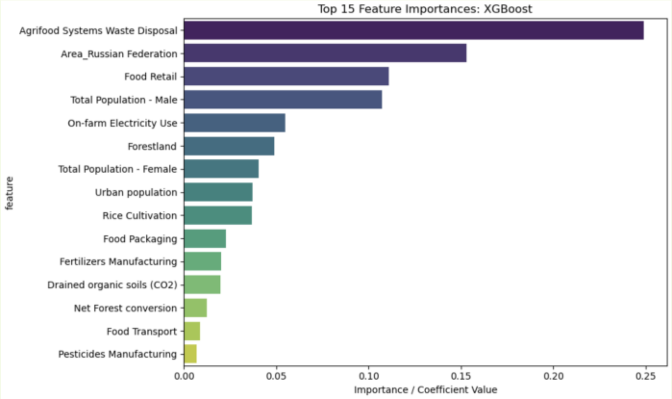


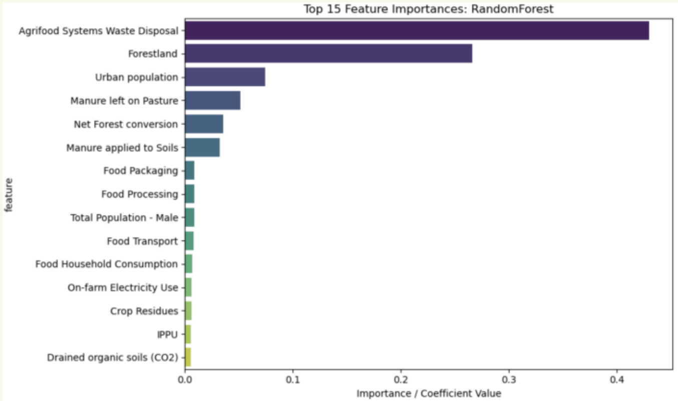
Agrofood_Final ›
Comparing 5 Runs from 1 Experiment
Visualizations
Parallel Coordinates Plot
Scatter Plot
Box Plot
Contour Plot
Parameters:
model_name
Metrics:
rmse_test
r2_test
Clear All
model_name
rmse_test
r2_test
XGBoost
Ridge
RandomFore
Lasso
LSTM
0.22858
0.20000
0.15000
0.10000
0.05000
0.03637
0.98228
0.90000
0.80000
0.70000
0.60000
0.50000
0.40000
0.30123
0.9
0.8
0.7
0.6
0.5
0.4

LOOKING AT IMPORTANT FEATURES





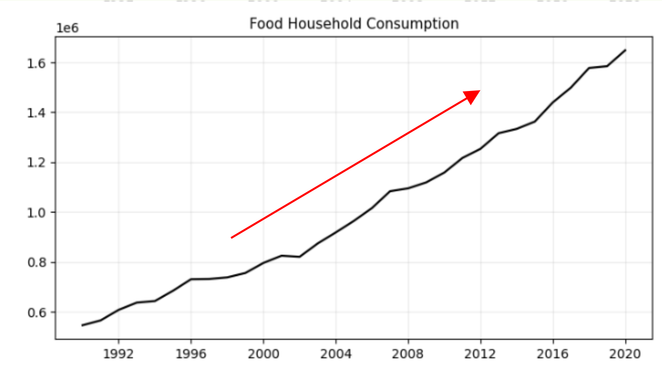




FORECASTING DATA ~2100

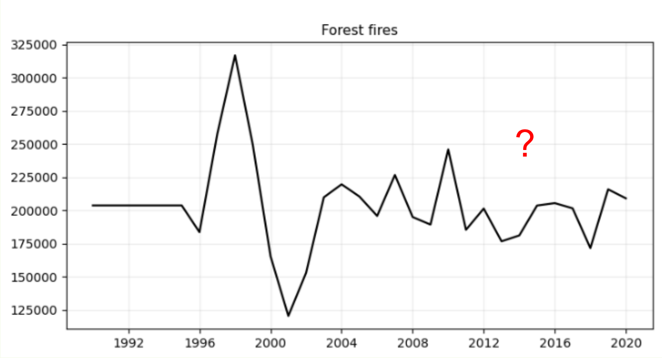
SMOOTH VARIABLES

- Recognizable trend (mostly upward)
- Damped Exponential Smoothing
- Variables like: *On-Farm Electricity, Food Retail, Food Household Consumption, Population*



NOISY VARIABLES

- No significant trends (mostly related to natural disasters)
- ARIMA(1, 1, 1)
- Variables like: *Forest Fires, Savanna Fires, Fires in Organic Soils*



CREATING SCENARIOS

TWO POLICY SCENARIOS GROUNDED IN RANGES USED BY:

- IPCC AR6 Mitigation Pathways
- FAO Agrifood Systems Mitigation Analyses (2021–2023)
- Integrated Assessment Models (IAMs) such as SSP1 / SSP2
- Literature on methane & N_2O mitigation in agriculture

THESE FRAMEWORKS COMMONLY ASSUME MULTIPLIERS:

- **10–25%** reductions for moderate mitigation (Scenario A)
- **30–70%** reductions for high-ambition pathways (Scenario B)

SCENARIO-WISE FEATURE **REDUCTION**

Land-use and Fire

Column	Scenario A	Scenario B
`Savanna fires`	0.80	0.40
`Forest fires`	0.80	0.40
`Fires in organic soils`	0.80	0.40
`Fires in humid tropical forests`	0.80	0.40
`Drained organic soils (CO2)`	0.80	0.40
`Net Forest conversion`	0.80	0.30
`Forestland` *(sink, often negative)*	**1.20**	**1.50**

Fertilizer, Manure, Rice, and Residues

Column	Scenario A	Scenario B
`Rice Cultivation`	0.80	0.50
`Fertilizers Manufacturing`	0.80	0.50
`Manure Management`	0.80	0.50
`Manure applied to Soils`	0.80	0.50
`Manure left on Pasture`	0.85	0.60
`Crop Residues`	0.90	0.60
`IPPU`	0.90	0.70

Energy, Supply Chain, and Consumption

Column	Scenario A	Scenario B
`On-farm Electricity Use`	0.85	0.50
`On-farm energy use`	0.85	0.50
`Pesticides Manufacturing`	0.85	0.60
`Food Processing`	0.90	0.60
`Food Packaging`	0.85	0.60
`Food Transport`	0.85	0.50
`Food Retail`	0.90	0.70
`Food Household Consumption`	0.90	0.70
`Agrifood Systems Waste Disposal`	0.80	0.30

BREAKING DOWN SCENARIO CREATION

$t = year\ (t \in [2021, 2100])$
$S(t) = linear\ scaling\ factor\ applied\ to\ each\ feature\ in\ year\ t$
$m = target\ multiplier\ assigned\ for\ that\ feature$
$p(t) = progress\ ratio\ from\ 0 \rightarrow 1\ across\ 2021 \rightarrow 2100$

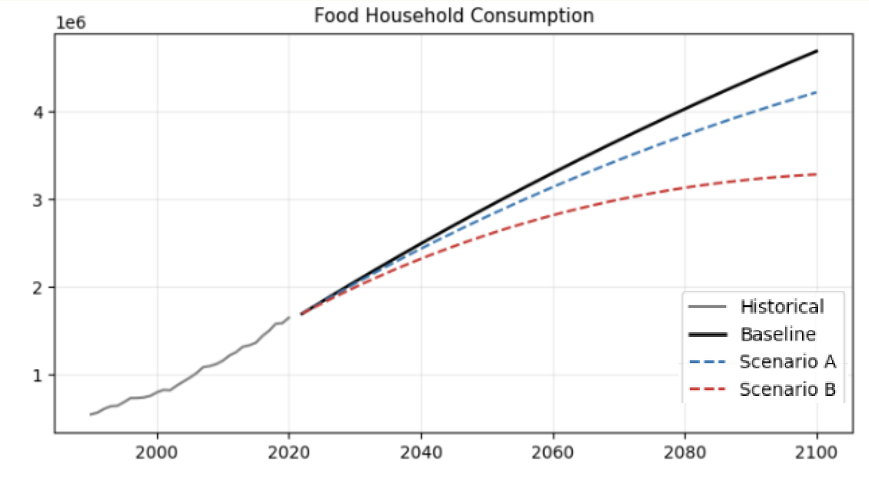
$$x\_scenario(t) = x\_baseline(t) \cdot S(t)$$

$$S(t) = 1 - p(t) \cdot (1 - m)$$

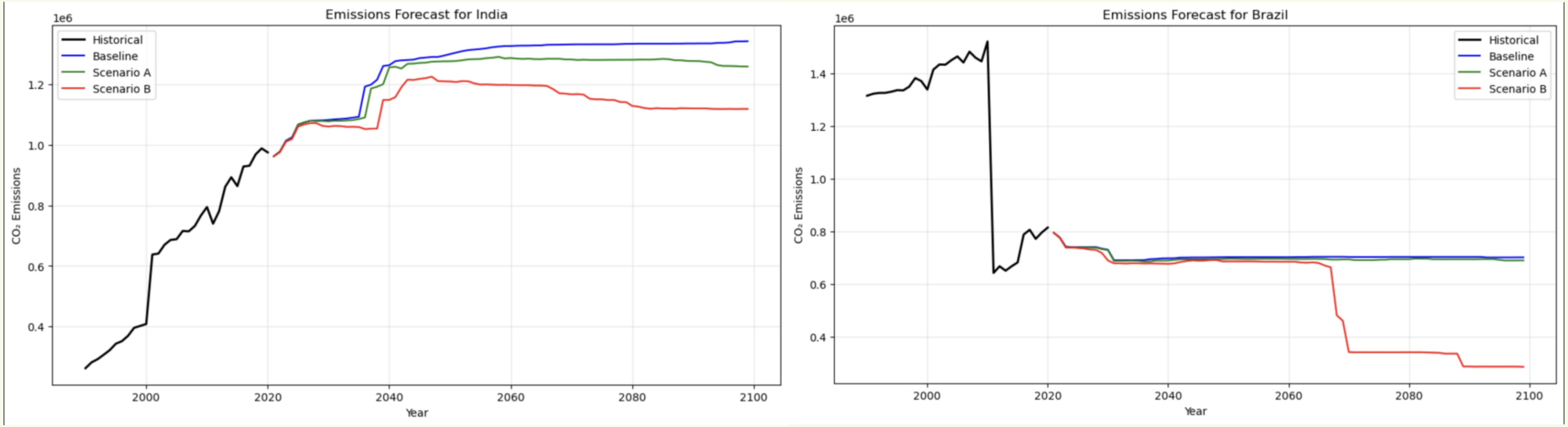
$$p(t) = (t - 2021) / (2100 - 2021)$$

$$S(2021) = 1$$
$$S(2100) = m$$

$$Scenario(t) = Baseline(t) \times [1 - p(t) \cdot (1 - m)]$$



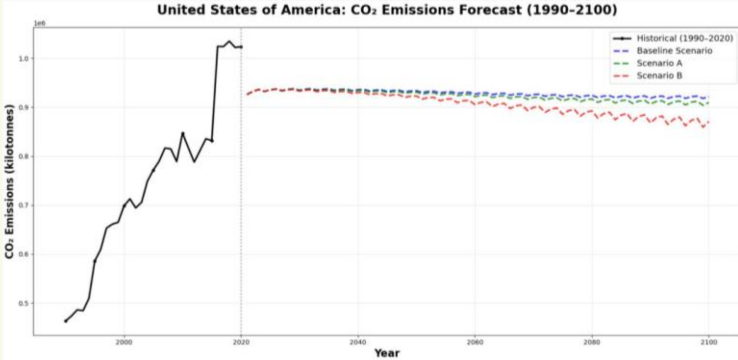
FORECASTING THE RANDOM FOREST

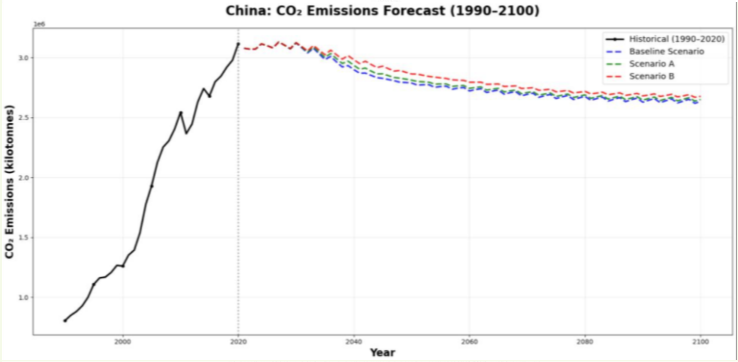


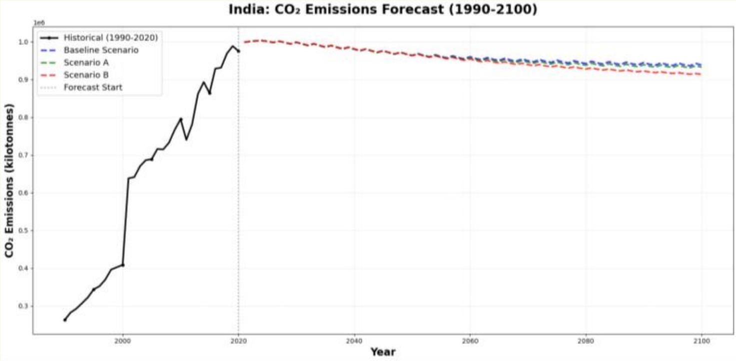
- Baseline projects **continued emissions growth** following historical momentum
- Moderate reduction scenario **slows the rate of increase** but *fails to reverse the trend*
- Heavy reduction scenario **successfully breaks the upward trajectory and stabilizes emissions** after 2040 for India and **triggers a massive structural drop** around 2065 for Brazil

THE TFT REPORT

- **Long-range emission forecasts** for major emitters using the trained TFT model
- All three countries show **stabilizing or slightly declining** baseline emissions once the historical trend peaks
- Scenario A produces **moderate reductions**, while Scenario B consistently achieves the **largest and earliest declines** across all countries
- The spread between Baseline, A, and B illustrates how **policy intensity directly shapes long-term emission pathways**







CONCLUDING OUR ANALYSIS

SUMMARY:

- Models achieved **high accuracy** (R^2: RF - 0.96, TFT - 0.98)
- We **successfully extrapolated emissions** through 2100 with stable long-term behavior
- Scenario A (10-25%) and Scenario B (30-70%) **tuned key agricultural emission drivers**, reducing the key emission drivers
- Both scenarios show **lower long-term emissions** for most countries

This demonstrates that our *policy-based parameter adjustments* can effectively model and guide potential emission-reduction strategies.

04

THE FINALE

Just ___ more slides to go!

OPTIONAL IMPROVEMENTS

RICHER FEATURES

Add interactions and temporal trends

BROADER MODELS

Include SGD and other Boost variations

BETTER SCALING

Apply Box-Cox or robust transformations

SCENARIO TESTING

Vary policy multiplier intensities

TIME HORIZONS

Explore alternative forecast window lengths

FaIR INTEGRATION

Introduce climate-model temperature conversion

FUTURE CONSIDERATIONS

SCOPE EXPANSION
Incorporate other major climate drivers *(energy, automobiles, factories)* for holistic forecasts

RICHER SCENARIOS
Move from fixed multipliers to policy-driven interventions *(carbon pricing, fertilizer caps, reforestation)*

ADVANCED MODELING
Explore nonlinear ML, uncertainty modeling, and richer FaIR inputs *(CH_4, N_2O, aerosols)*

BETTER DATA
Add sub-national detail, crop-level breakdowns, and forward-looking projections

FEEDBACK LOOPS
Integrate climate-agriculture interactions *(temperature → yields → emissions)*


THANK YOU
QUESTIONS?