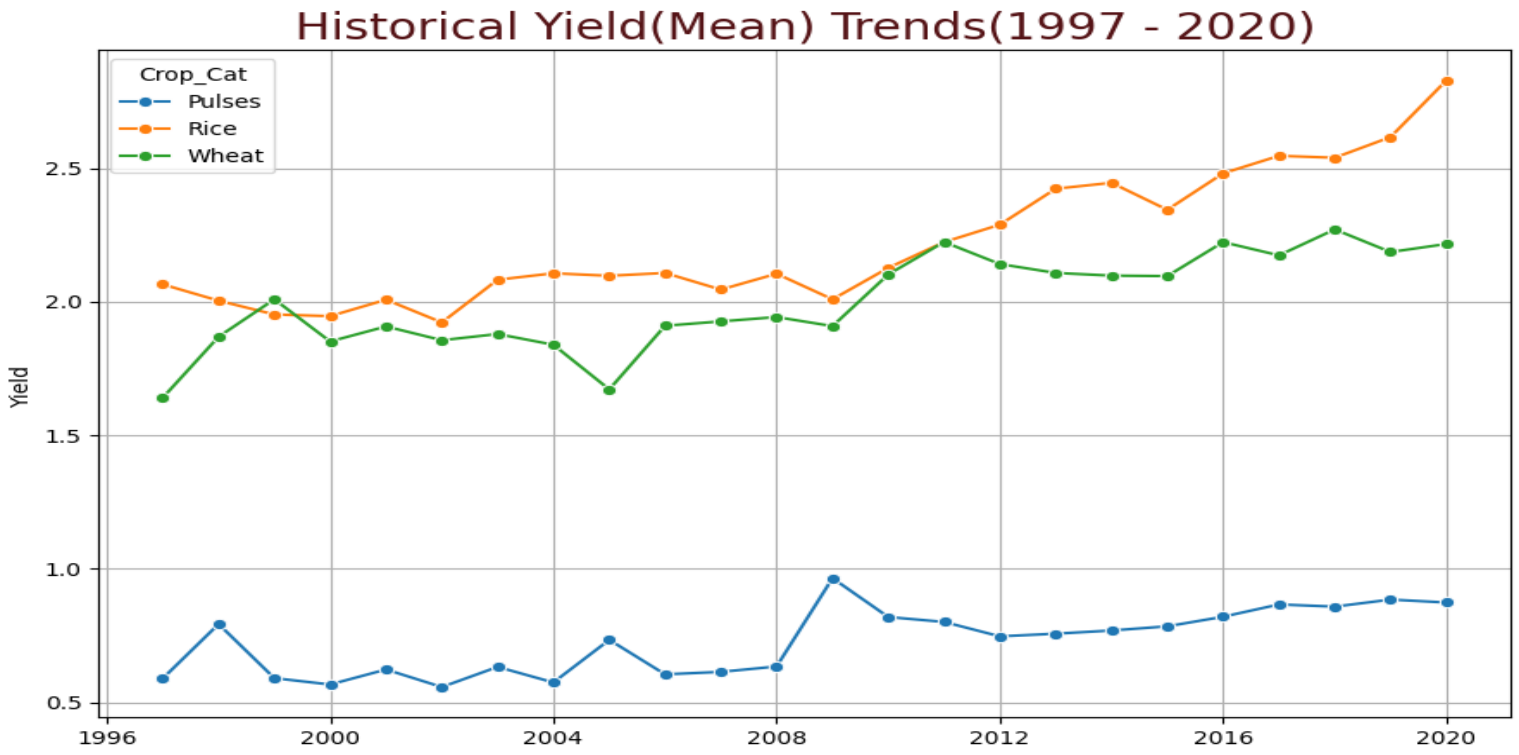
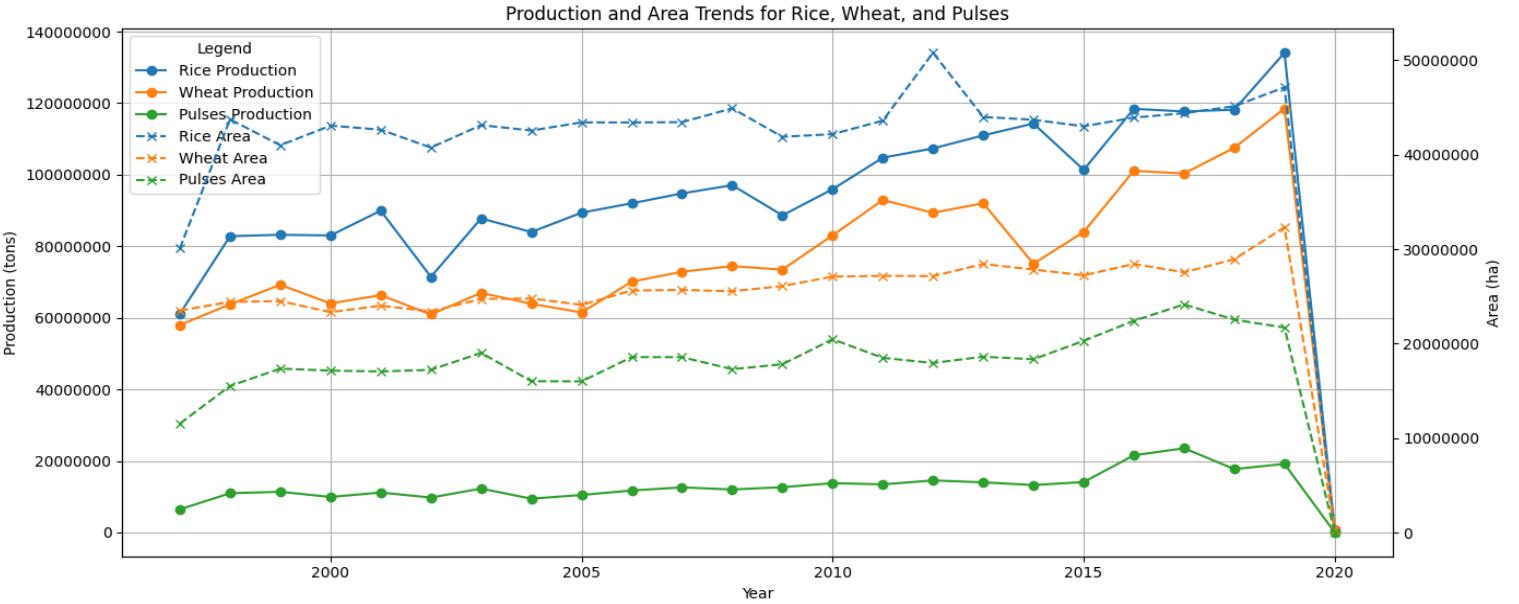
Crop Yield Prediction

Objective & Data

The project aimed to predict next-season agricultural yields for three major crop categories — Rice, Wheat, and Pulses — across Indian states and seasons.

- Historical Data: Crop yield records from 1997 to 2020 by Year, Season, and State.
- Additional Features: Annual rainfall, fertilizer use, pesticide use, cultivated area, and production volumes were included as regressors.

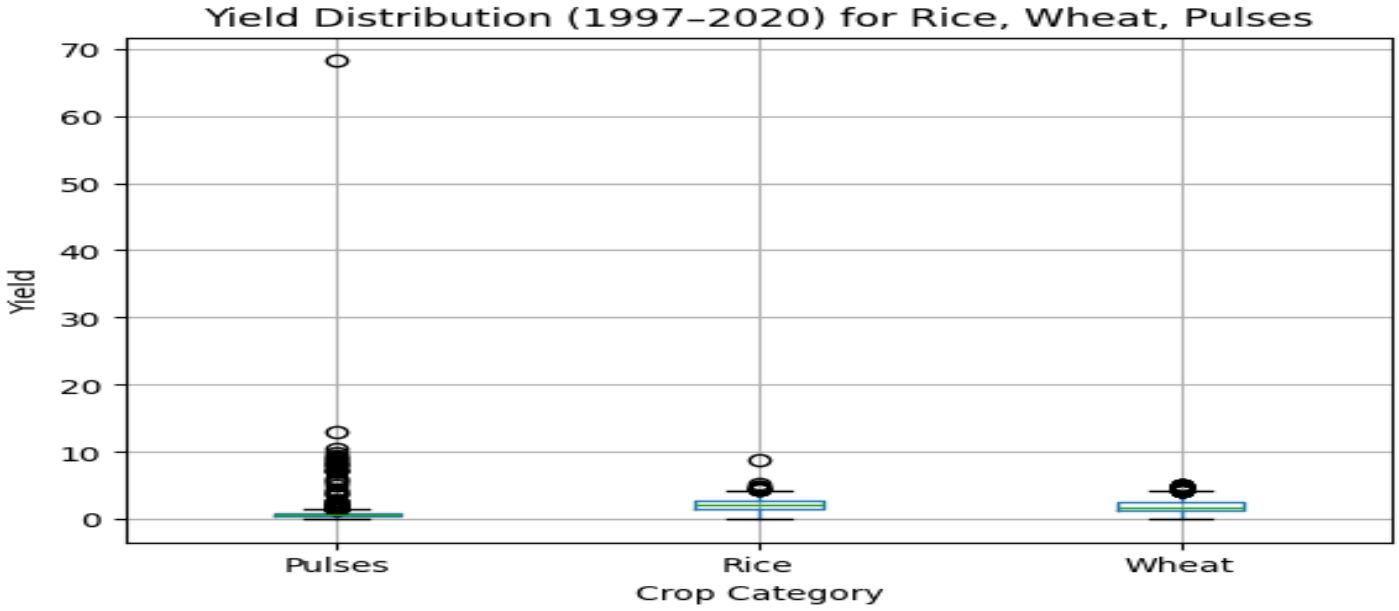




- Seasons Considered: Whole Year, Kharif, Rabi, Autumn, Summer, and Winter.

Methodology

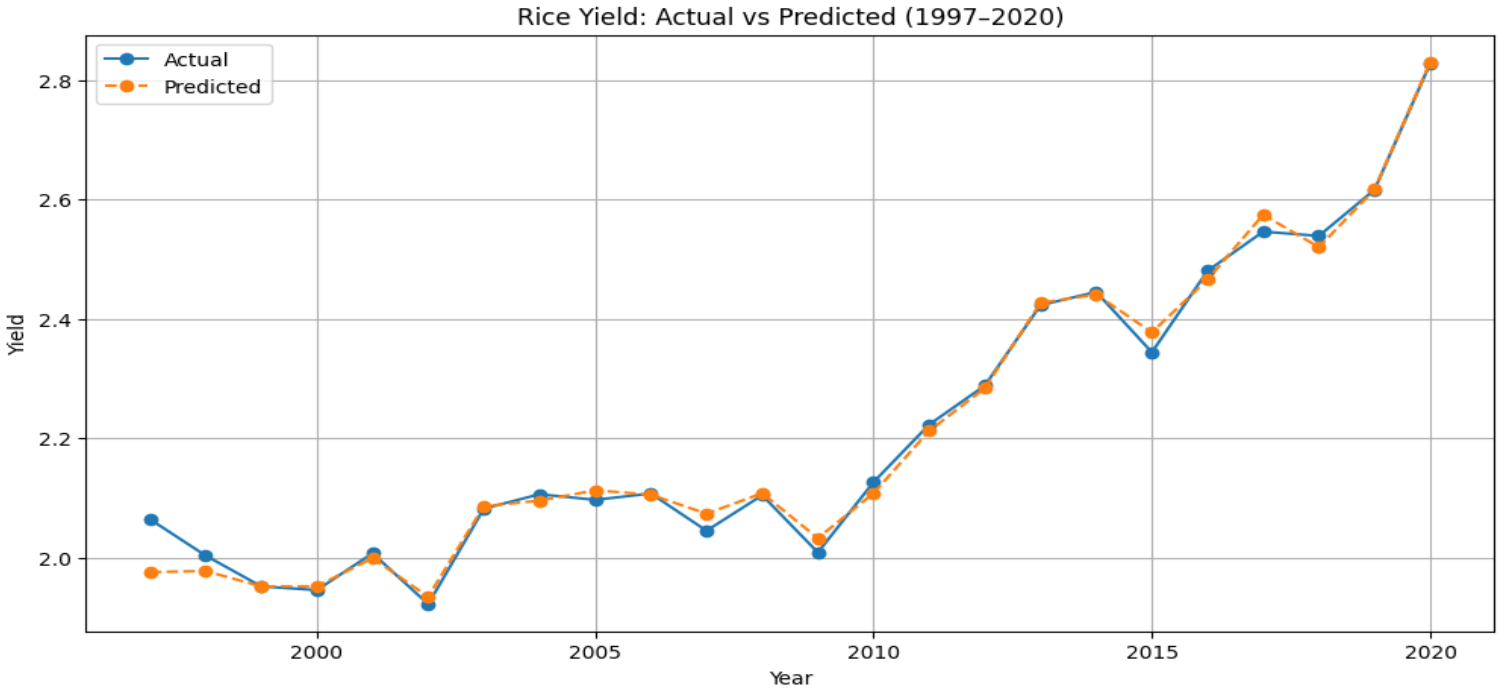
- Data Cleaning & Transformation:
    - Removed entries with zero or missing values in key columns (Area, Production, Yield).
    - Encoded categorical columns like State and Season.
    - Lag features (Yield_Lag1, Yield_Lag2), rolling averages, and sudden yield spike indicators were engineered.

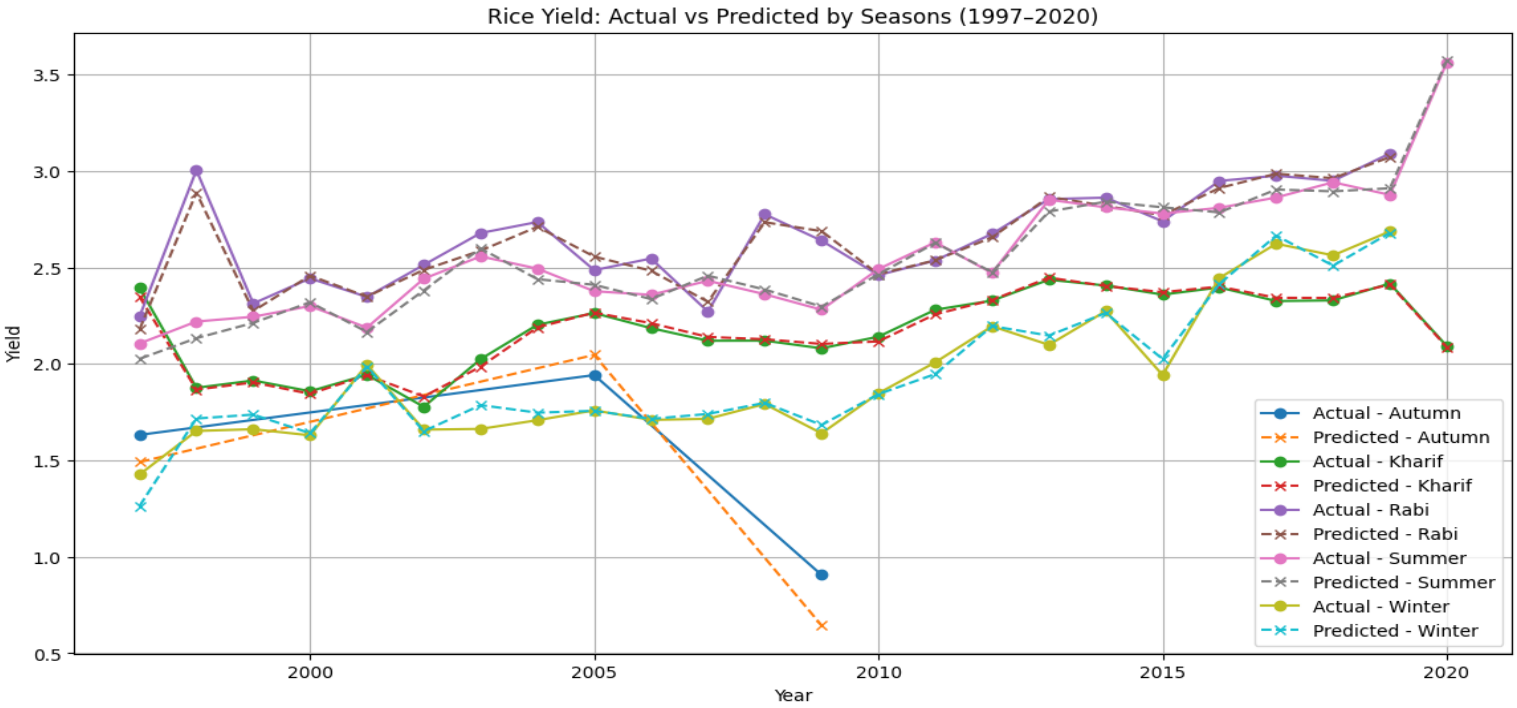


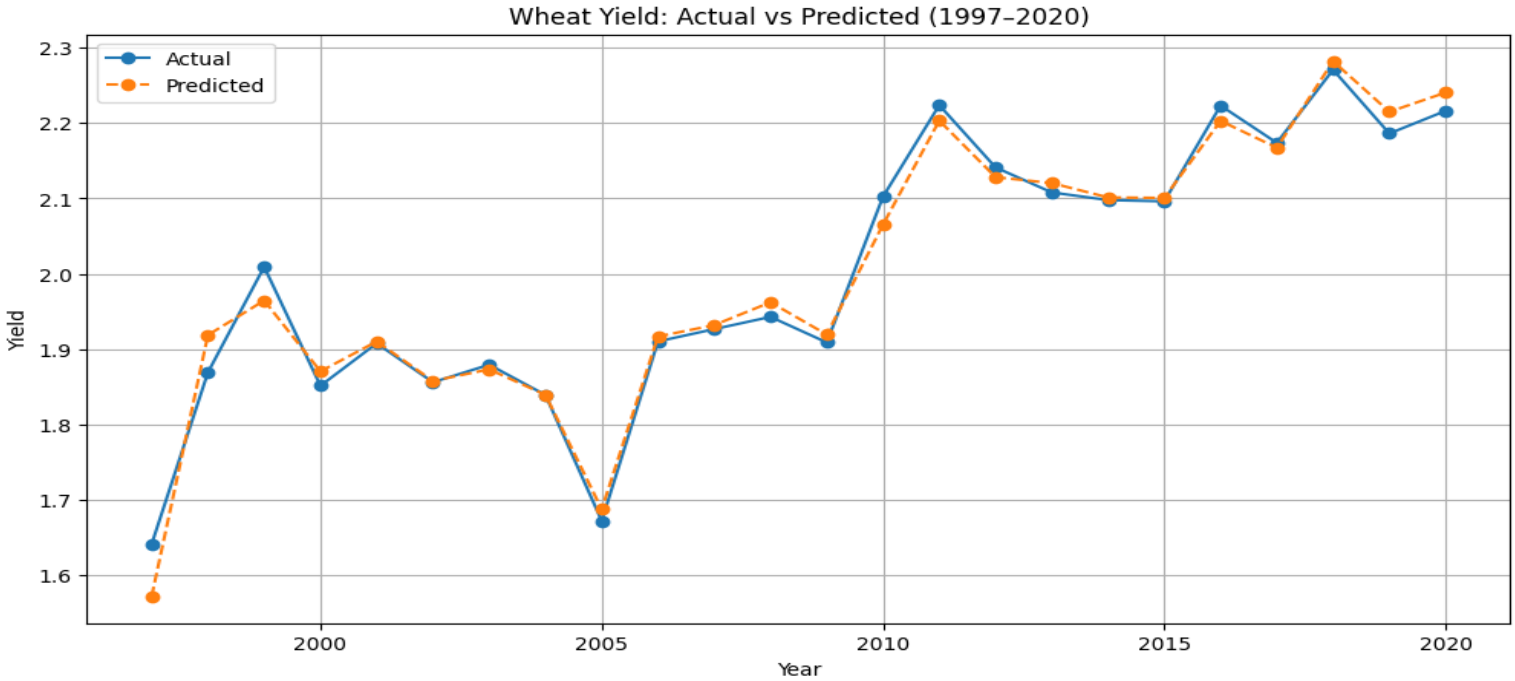
Spike counts per crop:

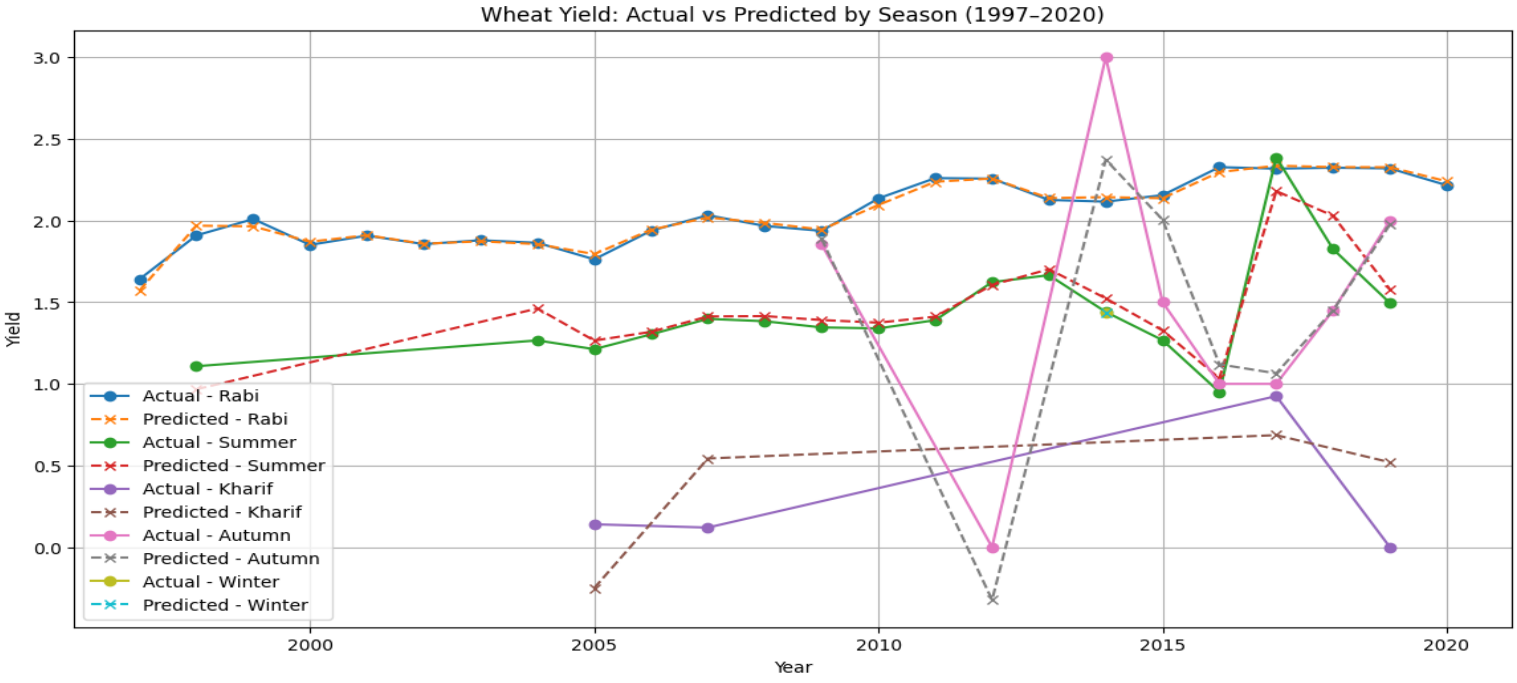
Crop Category	
Pulses	**119**
Rice	**9**
Wheat	**40**

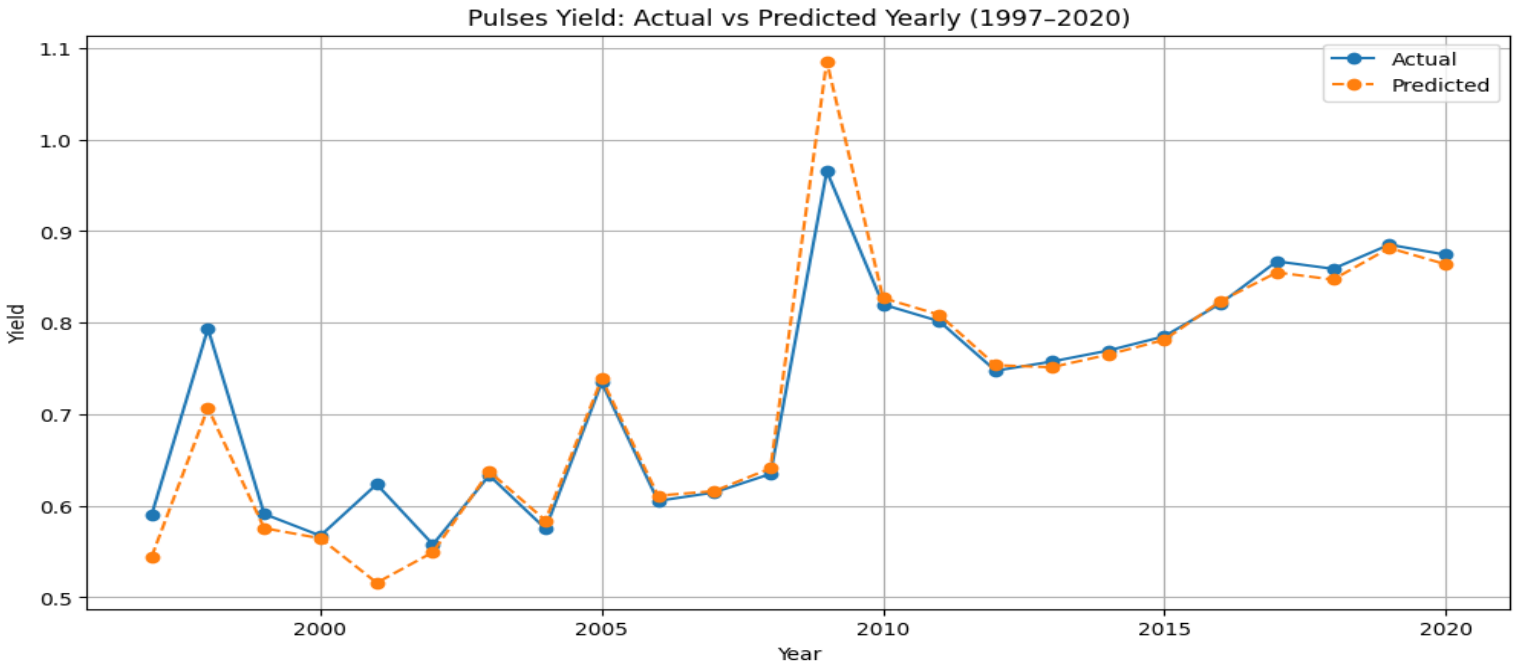
- Modeling:
    - Facebook Prophet was used for each crop type, with climate and input variables as external regressors.
    - Predictions extended from 2021 to 2023.
    - Cross-validation was done to validate accuracy.

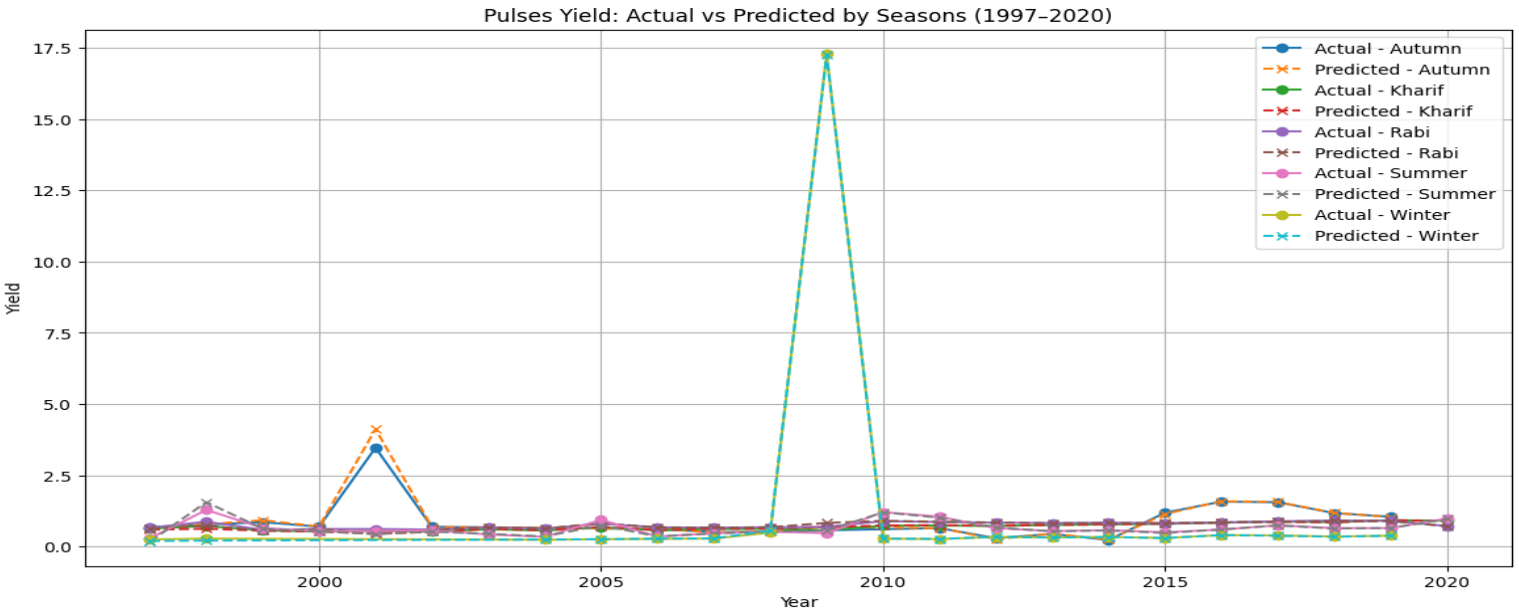



Rice Yield: Actual vs Predicted by Seasons (1997–2020)
Yield
Year
Actual - Autumn
Predicted - Autumn
Actual - Kharif
Predicted - Kharif
Actual - Rabi
Predicted - Rabi
Actual - Summer
Predicted - Summer
Actual - Winter
Predicted - Winter

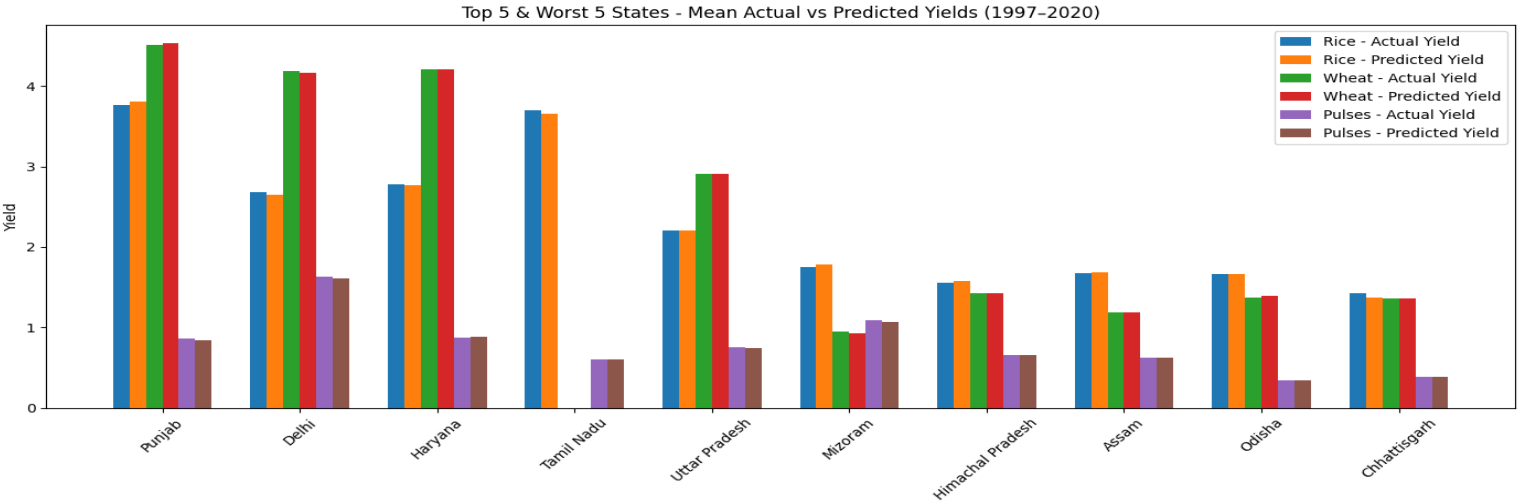

Wheat Yield: Actual vs Predicted (1997–2020)
Yield
Year
Actual
Predicted


Wheat Yield: Actual vs Predicted by Season (1997–2020)
Yield
Year
Actual - Rabi
Predicted - Rabi
Actual - Summer
Predicted - Summer
Actual - Kharif
Predicted - Kharif
Actual - Autumn
Predicted - Autumn
Actual - Winter
Predicted - Winter





- Analysis:
  - Top 5 and bottom 5 states were ranked based on predicted yields.
  - Correlation heatmaps identified the most influential factors for yield changes.



Top & Bottom 5 States by Predicted Yield (Mean values)

Top 5 States by Predicted Yield (Mean values)

State	Actual_Yield	Predicted_Yield
Punjab	3.045287	3.056572
Delhi	2.832879	2.809993
Haryana	2.620682	2.621628
Tamil Nadu	2.151110	2.127485
Uttar Pradesh	1.954193	1.951510

Bottom 5 States by Predicted Yield (Mean values)

State	Actual_Yield	Predicted_Yield
Mizoram	1.262866	1.260254
Himachal Pradesh	1.213576	1.219190
Assam	1.159276	1.163430
Odisha	1.124258	1.130991
Chhattisgarh	1.056414	1.040789

Key Findings

Predictive Accuracy

- The model demonstrated **high predictive accuracy** for historical years, with actual vs predicted yields closely aligned (**very low RMSE** and **very Low MAPE[avg 4% for all 3 crops]**).
- Given **Year–Season–State** level data for future years, yield forecasts can be **reasonably accurate**.

High & Low Yield Regions

- **High-yield next season**: States like **Punjab, Tamil Nadu, and Andhra Pradesh** (Rice), **Haryana & Punjab** (Wheat), and select southern states for Pulses.
- **Low-yield next season**: States like **Jharkhand, Chhattisgarh, Madhya Pradesh** for Rice and some northeastern states for Pulses.

Correlation Insights

- **Fertilizer, Pesticide,** show a **strong positive correlation** with yield for Wheat nut almost negligible with Rice and Pulses.
- **Seasons** do correlate with all e crops in positive manner though States do correlate positive with some crops or negative with other crops.
- **Annual Rainfall** shows mixed correlation — beneficial in some cases but negatively correlated in over-irrigated or flood-prone states.
- Lag-based, Rolling mean yield features (Yield_Lag1, Yield_Roll3, etc.) also strongly correlate with future yields, confirming yield persistence patterns.

Recommendations

Resource Allocation

- Prioritize **fertilizer and pesticide supply** in low yield states, ensuring optimized usage.
- **Targeted irrigation infrastructure** in drought prone regions.
- Provide **storage and processing facilities** to reduce post-harvest losses in high-production states.

Policy Strategies

- Launch **precision agriculture programs** in low erforming states.
- Expand **seasonal weather forecasting** for farmers to optimize planting and input use.
- Encourage **crop diversification** in low yield zones to reduce dependency on a single crop.

Conclusion

With the implemented methodology, the system can **accurately predict crop yields at a seasonal and state level** when provided with updated climate, input, and seasonal data for future years.

The top & bottom state yield rankings can guide **government interventions, subsidy allocation, and agricultural planning** for the upcoming seasons.