Use of Explainable AI using SHAP on Crop Yield Prediction in India

Student Name: Udit Jain
Roll No: 24CSM1R23
Guide Name: Dr. Manjubala Bisi

Scope of the Project

Project Focus

This project focuses on using **SHAP** to make crop yield prediction models not just accurate, but also **understandable**.

The goal is to ensure farmers and policymakers can **trust** and **interpret** the model's decisions—not just see the output.

Expected Outcomes

- A clearer view of which **factors actually drive yield** outcomes.
- Tools that **support smarter, data-backed planning** for those working on the ground.
- Ultimately, this work aims to **bridge the gap** between advanced AI models and real-world agricultural needs.

Importance of Crop Yield Prediction

Economic Impact:

- Accurate predictions can enhance **farmers' income** and **national economic stability**.
- **Food Security:** Predicting yields aids in planning for food supply, addressing hunger issues.
- **Resource Management:** Helps optimize the use of water, fertilizers, and land, promoting sustainable practices.

Crop yield prediction is vital for India's agricultural landscape!

Introduction to Explainable AI

Explainable AI (XAI) refers to methods and techniques that make the output of AI systems understandable to humans.

- **Importance:** Enhances trust and transparency in AI decisions.
- **Relevance in Agriculture:** Critical for decision-making in crop yield prediction, ensuring farmers and stakeholders understand model outputs.

Understanding AI is essential for effective agricultural analytics!

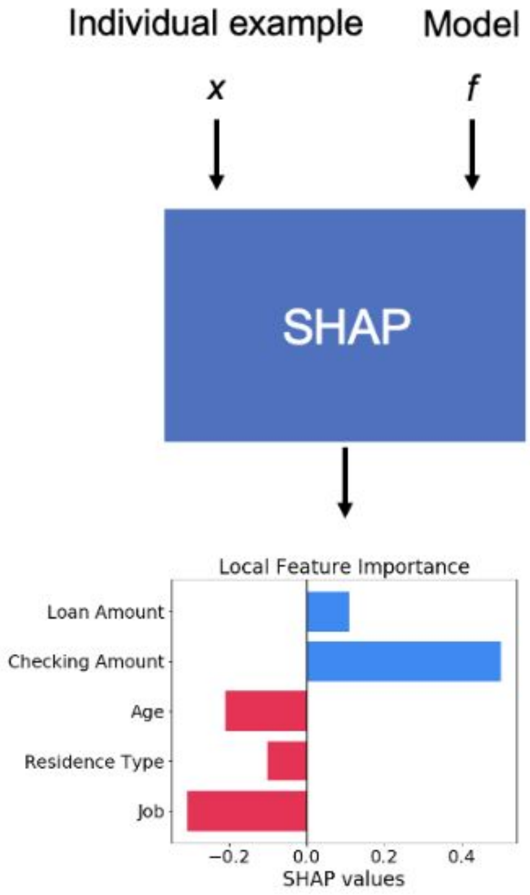
SHAP Methodology Overview

SHAP stands for **SHapley Additive exPlanations**.

Core Principle: Each feature's impact is calculated by comparing what the model predicts with and without that feature.

Working:

1. Start with a **model prediction** for a specific data instance.
2. Identify the **baseline output** (average prediction with no features).
3. Add **each feature** one at a time to observe its impact.
4. Compute **SHAP values** as the average marginal contribution of each feature.
5. Ensure additivity: all **SHAP values sum up** to the final prediction.
6. **Visualize results** using SHAP plots to explain feature influence.



Key Studies and Findings

Significant Research

- Recent studies show how AI is transforming agriculture by improving **forecasting** and **planning**.
- However, researchers emphasize that **without explainability**, the practical value of these models remains limited—especially for non-technical users.

Focus on Explainability

- Research highlights that understanding AI decisions boosts both **performance** and **adoption**.
- Methods like SHAP make models more **transparent**, helping users **trust** and act on predictions.
- These findings form the foundation of our approach, aiming to build models that are both **accurate** and **interpretable**.

Problem Statement

Challenges :

- **Traditional AI methods** often lack transparency in predictions.
- **Difficulty** in understanding how various factors influence yield predictions.
- **Variability** in environmental conditions.
- **Limited access** to reliable data sources.

Understanding these challenges is crucial for effective solutions!

Significance of the Problem

Consequences of Inaccurate Predictions

- Poor yield forecasting can cause **financial setbacks** for farmers due to inefficient planning.
- Leads to imbalanced supply chains, triggering **food shortages** and market instability.
- Results in **wastage of resources** like water, fertilizers, and labor.
- Reduces **farmer confidence in AI tools**, slowing adoption of smart agriculture technologies.

Literature Survey

- Most AI models in agriculture focus on **accuracy**, with limited attention to **explainability**.
- Traditional models do not effectively cover the **data, model, outcome, and end-user dimensions of explainability**

Title of the Paper	Published In	Significance	Limitation
Advanced ML for Regional Potato Yield Prediction	npj Sustainable Agriculture (2025)	Identifies **key environmental drivers** using SHAP;	Focused on **one crop** (potato) and region-specific
Explainable Model for Crop Yield in Indian Conditions	IJEC (2025)	Applies SHAP for **better feature insight** and model trust	Limited dataset and **only local interpretability**
Explainable ML for Corn Yield Using UAV Data	Computers and Electronics in Agriculture (2025)	Uses multispectral UAV data and SHAP for **field-level yield estimation**	**High hardware cost**; domain-limited to UAV-based data

Gaps in Current Research

Identifying Gaps:

- Lack of **interpreted models** in crop yield prediction.
- Need for more research **integrating explainability** into agricultural AI.

Importance of Addressing Gaps:

- Enhancing the **reliability of predictions**.
- Supporting farmers with **actionable insights**. Filling these gaps is crucial for advancing agricultural analytics!

Proposed Solution Overview

Integrating SHAP:

- Utilizing **SHAP** to enhance crop yield prediction models.
- Aiming for improved **interpretability** and **accuracy**.

Expected Benefits:

- **Better insights** into the factors influencing crop yields.
- **Increased trust** in AI-driven predictions among stakeholders. Our solution aims to revolutionize crop yield prediction!

Model Development Process

Stages of Development

- Developed a pipeline for crop yield prediction using **agricultural data**.
- Ensured model interpretability by planning **SHAP integration** early in the pipeline.

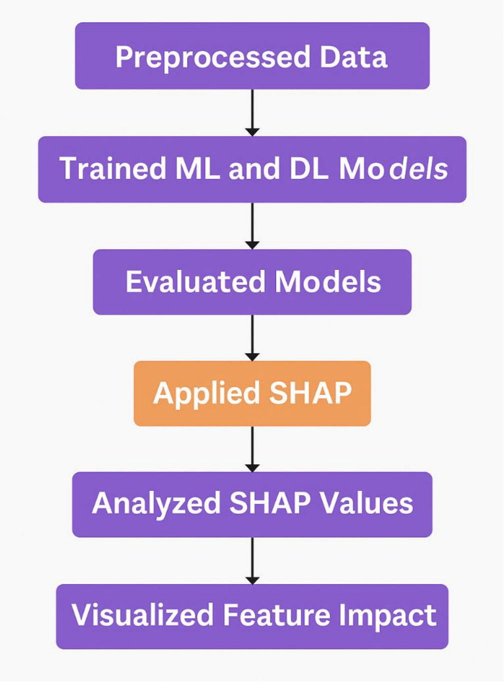
Algorithm Selection

- Explored **ML and DL models**: Decision Tree, Random Forest, XGBoost, RNN, and LSTM.
- Selection was based on **performance** and compatibility with SHAP explainability tools.
- Final model (**Random Forest**) was chosen for its high accuracy (98.96%) and suitability for SHAP-based interpretation.

Implementation of SHAP

Workflow Steps:

1. **Preprocessed** the crop yield dataset (cleaning, encoding, and normalization).
2. **Trained ML and DL models** including Random Forest, Decision Tree, and XGBoost, RNN.
3. **Evaluated models** using metrics: Accuracy, MAE, RMSE, and SD.
4. **Applied SHAP** to the best-performing model (Random Forest) for explanation.
5. **Analyzed SHAP values** to identify key features.
6. **Visualized feature impact** using summary, waterfall, and decision plots for interpretation.



Data Collection and Methodology

Data Sources

Utilization of **agricultural datasets** from government and research institutions, incorporating **environmental**, **economic**, and **historical data**.

The dataset includes data on multiple states and districts across India, capturing **seasonal crop yield information** with parameters such as:

- **State Name** and **District Name**
- **Crop Year** and **Season**
- **Crop Type, Area, Production,** and **Yield**

Methodology

Employing **machine learning and deep learning algorithms** alongside **SHAP** for analysis, ensuring robust validation of models through **cross-validation techniques**.

	state_name	district_name	crop_year	season	crop	area	production	yield
0	Andaman and Nicobar Islands	NICOBARS	0.166667	Kharif	Arecanut	1.461475e-04	1.598977e-06	1.594896
1	Andaman and Nicobar Islands	NICOBARS	0.166667	Kharif	Other Kharif pulses	2.284356e-07	7.994883e-10	0.500000
2	Andaman and Nicobar Islands	NICOBARS	0.166667	Kharif	Rice	1.188331e-05	2.566358e-07	3.147059
3	Andaman and Nicobar Islands	NICOBARS	0.166667	Whole Year	Banana	2.050792e-05	5.124720e-07	3.642045
4	Andaman and Nicobar Islands	NICOBARS	0.166667	Whole Year	Cashewnut	8.391044e-05	1.319156e-07	0.229167
...	...	...	...	...	...	...	...	...
242563	West Bengal	PURULIA	0.944444	Summer	Rice	3.565926e-05	6.403902e-07	2.617647
242564	West Bengal	PURULIA	0.944444	Summer	Sesamum	7.307141e-05	3.701631e-07	0.738437
242565	West Bengal	PURULIA	0.944444	Whole Year	Sugarcane	3.775714e-05	1.299169e-05	50.154321
242566	West Bengal	PURULIA	0.944444	Winter	Rice	3.253470e-02	4.780133e-04	2.141848
242567	West Bengal	PURULIA	0.944444	Winter	Sesamum	2.039137e-05	7.035497e-08	0.502857

242568 rows × 8 columns

Dataset

Performance Metrics

Evaluation Criteria

- **Accuracy**, **MAE**, and **RMSE** were used to evaluate model performance.
- **SHAP values** were used to interpret model predictions and analyze feature importance.
- **SHAP Summary Plot** highlighted key features
- **SHAP Waterfall Plot** visualized the contribution of individual features to a single prediction
- **SHAP Decision Plot** illustrated feature influence across multiple predictions

This approach ensured that models were not only accurate but also explainable, supporting informed decision-making in crop yield prediction.

Results Overview

Key Results Achieved:

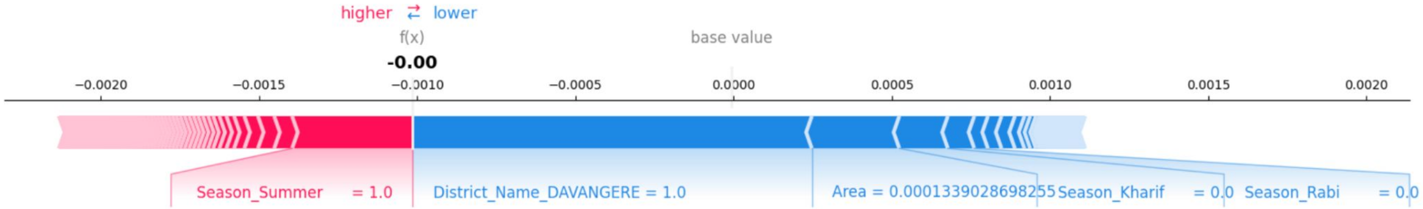
- Achieved 98.96% accuracy using **Random Forest** with MAE: 1.97, RMSE: 2.45.
- **SHAP** identified key features: Crop_Coconut, Area, Season_Kharif, Crop_Rice.
- **Waterfall plot** showed District_TIRUNELVELI and Crop_Urad as strong positive contributors.
- **Decision plot** illustrated consistent positive impact of Area, with variable effect from Crop_Rice.

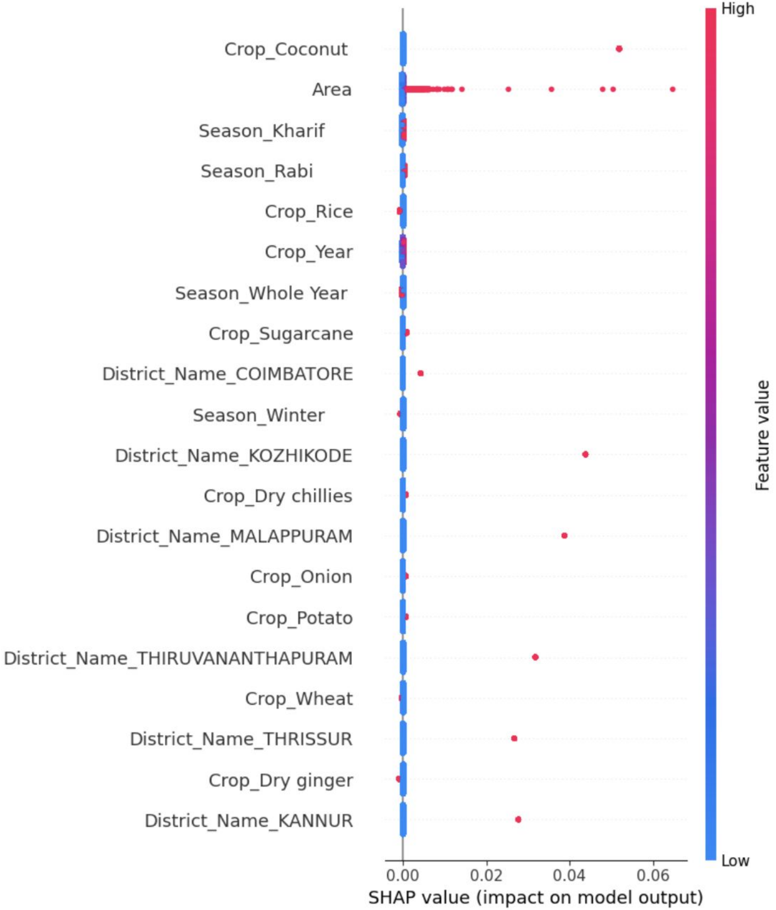
SHAP improved model interpretability and trust, enabling informed agricultural decisions.

TABLE 1. Simulating model performance with area and production as inputs.

Model	Accuracy	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Standard Deviation (SD)
Random Forest	98.96	1.97	2.45	1.23
Decision Tree	89.78	4.58	5.86	2.75
XGBoost	86.46	6.31	7.89	3.54

Force Plot





Summary Plot

Interpretation of Results

In-Depth Analysis

SHAP plots provided clear interpretability by showing:

- Summary plot revealed **Area** had the strongest positive influence on yield.
- Waterfall plot highlighted **District_Name_TIRUNELVELI**, and **Season_Kharif** as key contributors to specific predictions.

Implications of Findings

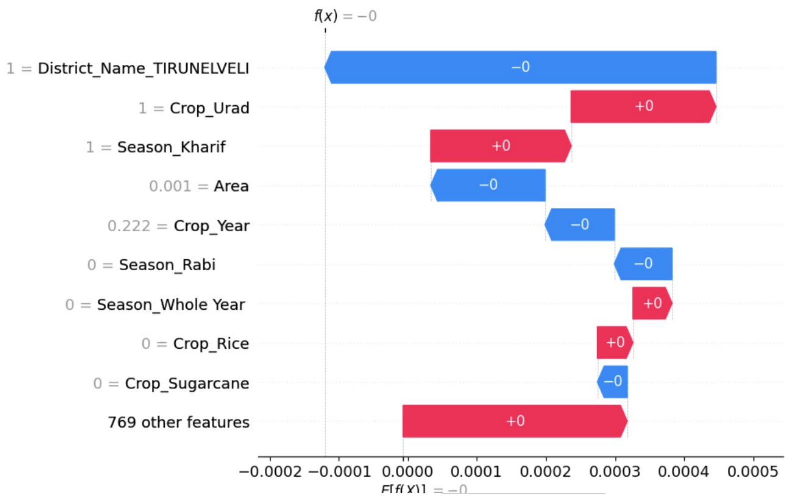
- Supports **data-driven decisions** for optimizing crop planning and resource allocation.
- Insights can guide **region-specific policies** by identifying high-impact crops and conditions.

Visualizing SHAP Values

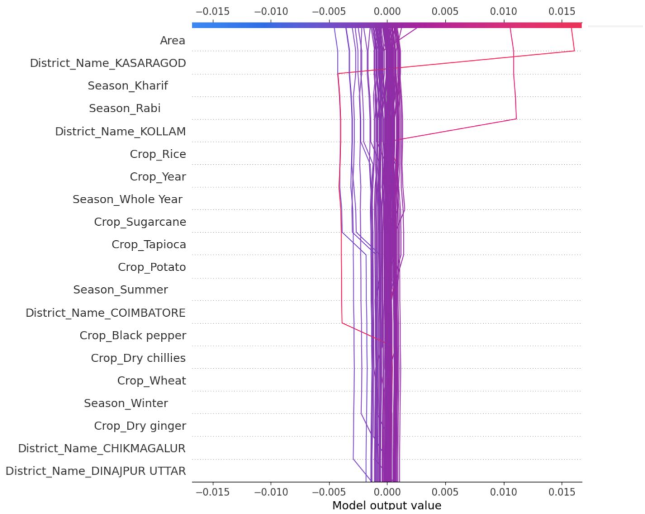
SHAP Value Representations

- **Waterfall plot** breaks down individual predictions feature by feature.
- **District_Name_TIRUNELVELI**, and **Area** had strong positive contributions.
- Features like **Season_Kharif** and **Crop_Urad** showed negative influence.

- **Decision plot** shows cumulative effect of features over multiple samples.
- **Crop_Rice** and **District_Name_KASARAGOD** consistently increased predicted yields.



Waterfall Plot



Decision Plot

Comparison with Traditional Models

Performance Comparison

- Random Forest + SHAP achieved 98.96% accuracy

Outperformed:

- Decision Tree: 89.78%
- XGBoost: 86.46%
- Lowest MAE (1.97) and RMSE (2.45) among all models

Evaluating **SHAP-enhanced models** against traditional predictive models highlights improvements in **interpretability**.

Key Takeaways

- SHAP provide **clear interpretability** through feature contribution visualization
- Enable **greater user trust** and informed decision-making in agriculture

Conclusion and Insights

Project Summary

- Demonstrated how integrating Explainable AI with SHAP enhances both **prediction quality** and **understanding**.
- Enabled stakeholders to not just see predictions, but also understand the **why** behind them.
- Combined high **model accuracy** with **interpretability**, bridging the gap between AI performance and usability in agriculture.

Future Implications

- Opens new directions for **decision-support systems** tailored to farmers and policymakers.
- Lays the groundwork for **scalable AI frameworks** that can adapt across **crops and regions**.
- Reinforces the importance of **trustworthy AI** in critical domains like food security and sustainability.

Thank You