CROP PRODUCTION PREDICTION USING DECISION TREE

K. Tejasree | AP23110011434

Project Overview

- This project aims to predict crop production using Machine Learning based on historical agricultural data.
- The goal is to estimate crop output for selected districts in Andhra Pradesh.
- A Decision Tree Regressor is used because it handles categorical and numerical features effectively and provides clear, interpretable predictions.
- Dataset: Crop Production in India (Kaggle)
- Crops Used: Rice, Maize, Onion, Groundnut
- Districts Used: Krishna, Guntur, Kadapa, Kurnool

Introduction

Agriculture plays a major role in India’s economy, and accurate prediction of crop production helps:

✓ Farmers plan resources
✓ Government manage food supply
✓ Improve storage and distribution
✓ Reduce risks in crop failures

- Traditional yield prediction methods are time-consuming and less accurate.
- Machine Learning provides a modern approach by learning patterns from historical data.

Problem Statement:

Crop production varies based on many factors such as:

✓ Area cultivated
✓ Season
✓ Crop type
✓ District/location
✓ Historical output

This makes manual prediction difficult.

Challenges:

1. Wide variation in production values
2. Missing or inconsistent data
3. Need for accurate predictions for planning
4. Categorical fields like Crop, District, Season must be encoded

Problem:

How to develop a reliable ML model that predicts crop production accurately using limited available features?

Proposed Solution

We build a Decision Tree Regression Model using filtered data from Andhra Pradesh.
The model learns relationships between:

- Crop Year
- Area
- Crop Type
- District
- Season

and then predicts Production.
Why Decision Tree?
✓ Handles mixed data types
✓ Captures complex & nonlinear patterns
✓ Easy to interpret
✓ Works well on medium-sized datasets
✓ Fast and efficient

Pipeline Steps

Implementation Details

Tools & Technologies

- Google Colab
- Python 3
- Libraries:Pandas,NumPy, Scikit-learn
- Dataset: Crop Production in India (Kaggle)

Implementation Steps

- Loaded the dataset
- Filtered rows for Andhra Pradesh
- Selected four major districts and crops
- Cleaned missing/invalid rows
- Encoded categorical columns
- Split data into training & testing
- Built a Decision Tree Regressor
- Tuned parameters to reduce overfitting
- Evaluated model performance
- Generated final accuracy metrics

Model Performance

Filtered Data Shape: (547, 7)
Training Accuracy: 96.76%
Testing Accuracy: 90.03%
Difference: 6.73% (Low → good model fit)

Metrics:

- MAE: 37,196.94
- RMSE: 85,977.20
- MAPE: 74.62%
- SMAPE: 45.32%

```
Filtered Data Shape: (547, 7)

===== TRAIN & TEST ACCURACY =====
Train Accuracy (%): 96.76
Test Accuracy (%): 90.03
Difference (%): 6.73

===== MODEL PERFORMANCE =====
MAE: 37196.94
RMSE: 85977.2
MAPE (% Error): 74.62
SMAPE (% Error): 45.32
```

These results show that the model predicts crop production effectively with limited features.

Results & Discussion

Key Findings

✓ The Decision Tree model performed strongly in predicting crop yield

✓ Low overfitting due to parameter tuning

✓ Seasonal and district filters improved accuracy

✓ Model handles nonlinear relationships effectively

Observations

- Agricultural data naturally shows large fluctuations
- Limited features still produced good accuracy
- With more features (rainfall, soil), accuracy could increase

Limitations

- Dataset lacks environmental factors
- Only 4 crops and 4 districts were used

Conclusion

- The project demonstrates that machine learning can be effectively applied to predict crop production.
- Decision Tree Regression gave 90% test accuracy, making it reliable for forecasting.
- Proper filtering, cleaning, and encoding significantly enhanced performance.
- This model can support planning for agriculture, storage, and policymaking.

Future Scope

- Add rainfall, fertilizer, and soil data
- Use Random Forest or XGBoost for higher accuracy
- Deploy as a web dashboard for farmers


THANK YOU