



# **BIG DATA ANALYTICS IN AGRICULTURE**

**BY:- NAMRATHA KANCHARLA  
24MBMA15  
MBA (GENERAL)**

# PROBLEM STATEMENT

Agriculture struggles with Inaccurate Forecasting and Inefficient resource use, Leading to financial loss, Environmental Strain and Reactive Management



The use cases answer the question of  
How can Big Data Analytics be strategically applied to integrate diverse agricultural data streams (soil, weather, historical, market) to address the core challenges of modern farming?



# Use Case 1:- Precision Nutrient and Fertilizer Management

Goal: Recommend the most suitable fertilizer based on soil nutrients and crop conditions.

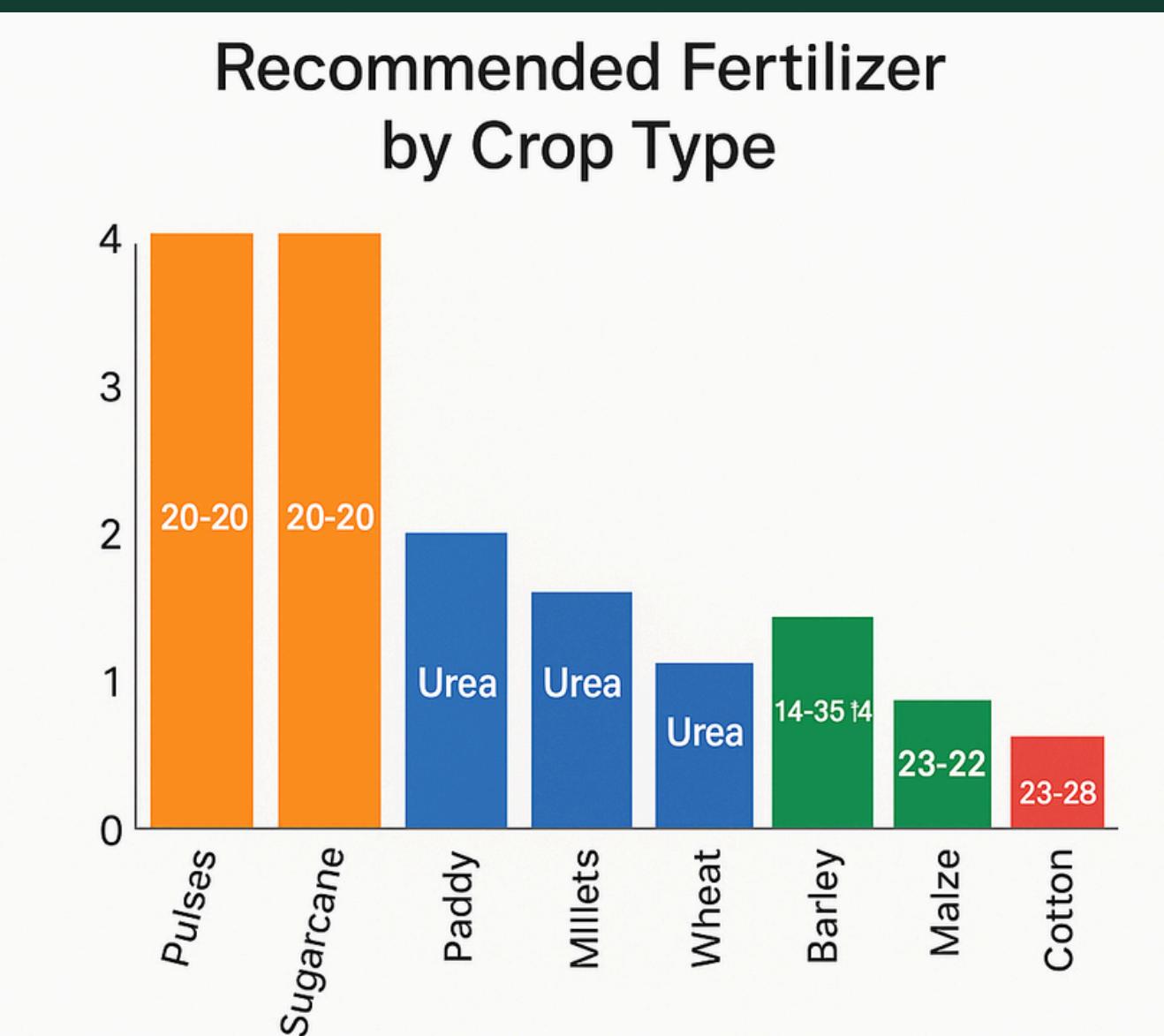
Approach:

- Uploaded fertilizer dataset to Databricks
- Cleaned and transformed data using PySpark / SQL
- Trained a Logistic regression model using MLlib

Big Data Stack: Databricks, Spark MLlib Classification, Vector Assembler, PySpark

Results & Insights:

- The model identifies ideal fertilizers based on N-P-K imbalance
- Helps reduce excessive chemical use and supports sustainable farming



# Use Case 2:- Crop Yield Prediction

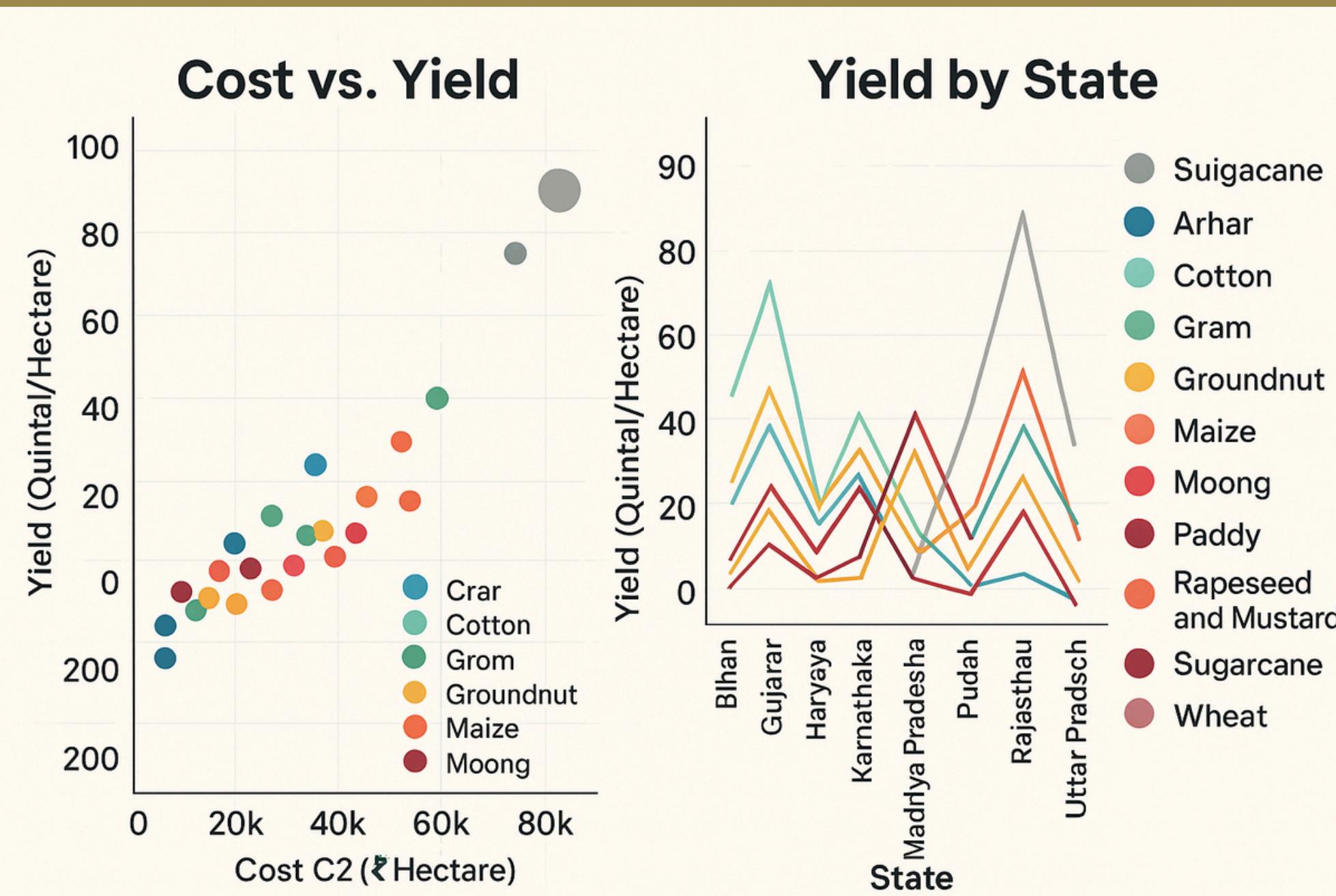
Goal: Predict crop yield based on environmental and soil parameters.

Approach: Data cleaning and feature engineering using Spark

Random Forest Regression model

Output tables generated for predicted yields

Big Data Stack: Spark MLlib Regression, Databricks Notebooks, SQL + PySpark



## Results & Insights:

- Clear patterns found between rainfall, humidity percent, and yield
- Helps farmers plan water usage and crop variety selection
- Supports government planning for production forecasting

## Use Case 3:-Pest Attack Prediction

Goal: Predict pest infestation probability based on environmental factors.

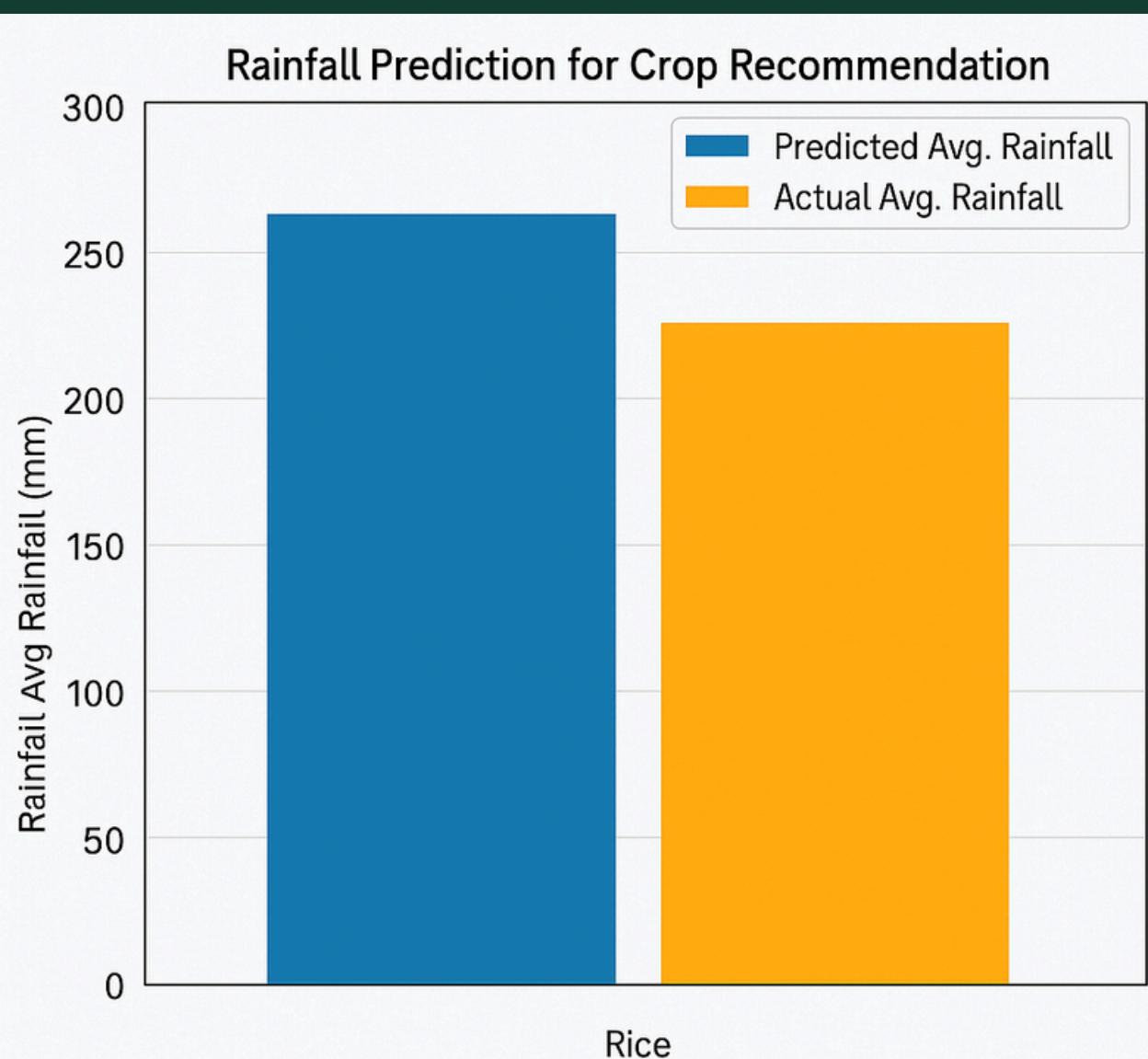
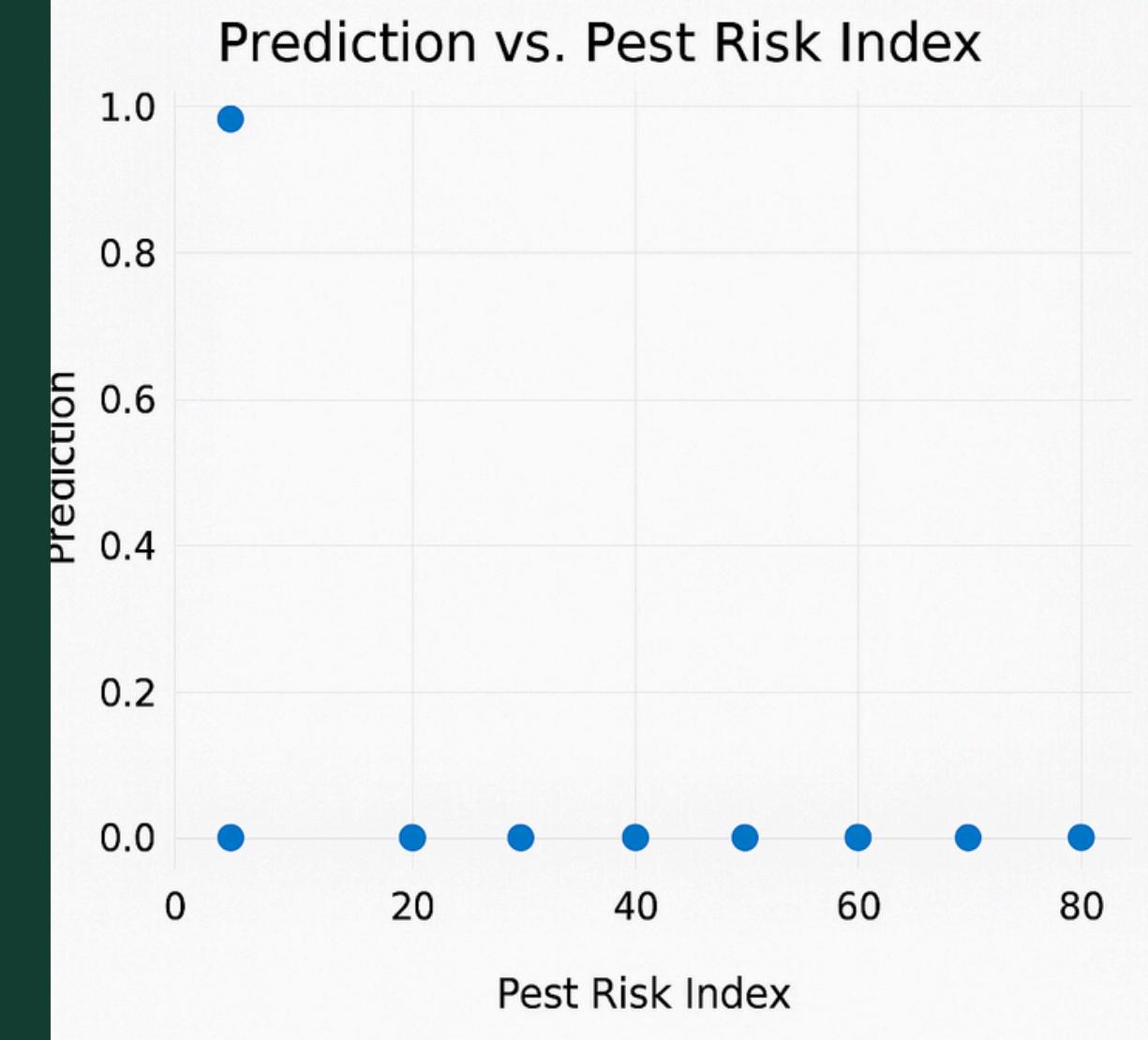
Approach:

- Preprocessed pest dataset using PySpark
- Built a classification model using Spark MLlib

Big Data Stack: Apache Spark, Databricks Delta, MLlib Classification

Results & Insights:

- Model predicts high-risk pest zones
- Allows farmers to plan pesticides only where necessary



## Use Case 4:- Rainfall Prediction for Crop Recommendation

Goal: Predict average rainfall for a crop using environmental, soil, and atmospheric parameters to support crop recommendation decisions.

Approach:

- Data cleaning, renaming, and type-casting in Spark
- Feature engineering (e.g., high-rainfall indicator, categorical encoding)

Big Data Stack: Spark MLlib Regression, Databricks Notebooks, PySpark + SQL, Feature Pipelines, VectorAssembler, DecisionTreeRegressor

Results & Insights:

- Model predicts rainfall with moderate accuracy (avg error ~29 mm)
- Predicted rainfall is slightly higher than actual values for rice

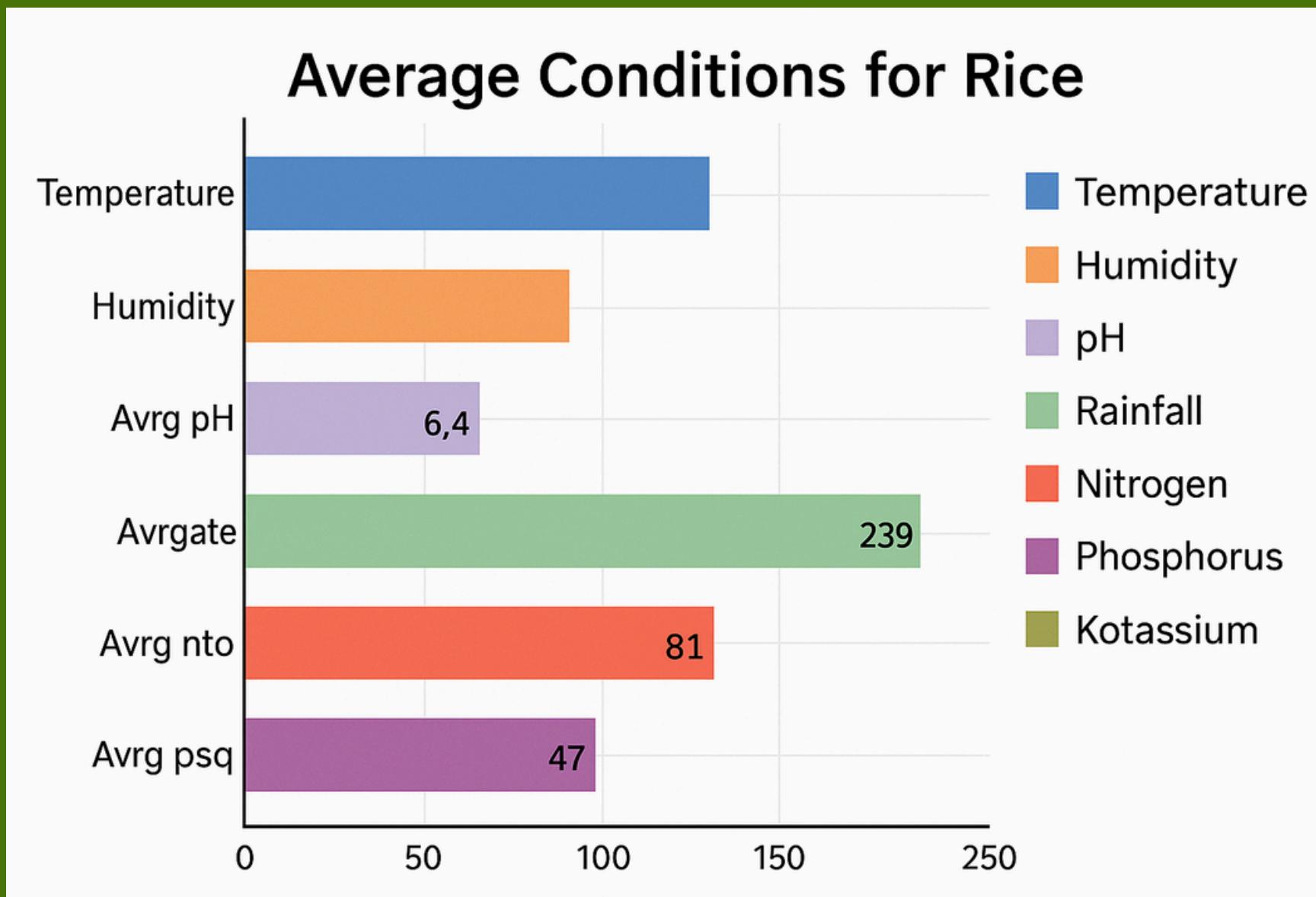
# Use Case 5: Rice Growth Condition Profiling & Prediction Model

Goal: Predict the crop type based on soil nutrients (N, P, K), climate factors (temperature, humidity), soil pH, and rainfall conditions.

Approach:

- Applied feature engineering (VectorAssembler + StandardScaler)
- Trained a Naive Bayes classification model using Spark MLlib

Big Data Stack: Spark MLlib Classification,Databricks Notebooks, SQL + PySpark ,Feature Pipeline (StringIndexer, Assembler, Scaler)



### Results & Insights:

- Dataset contained only one crop (rice), so the model achieved 100% accuracy
  - Visualization highlighted ideal rice-growing conditions:
    - High humidity (~82%)
    - Moderate temperature (~23°C)
    - Slightly acidic soil (pH ~6.4)
    - Very high rainfall (~239 mm)
    - Consistent NPK nutrient levels



A wide-angle photograph of a vast tobacco field at dusk or night. The foreground is filled with the dark green, broad leaves of tobacco plants, arranged in a dense, undulating pattern across the landscape. The sky above is a deep, dark teal or black, suggesting twilight or nighttime.

Thank You!

