M. Bhavana Reddy

Crop Recommendation Report

1. Dataset Description

1.1 Source: Crop Recommendation Dataset from Kaggle and academic repository (2,200 records approximately).

1.2 Columns:
- N – Nitrogen content in soil
- P – Phosphorus content in soil
- K – Potassium content in soil
- temperature – Ambient temperature in Celsius
- humidity – Relative humidity in percentage
- ph – Soil pH value (acidity/alkalinity)
- rainfall – Average rainfall in mm
- label – Recommended crop type (22 unique crops)

1.3 Data Quality:
- No missing/null values detected
- Clean and consistent dataset
- Balanced distribution across crop labels
- Numeric features properly scaled for analysis

---

2. Operations Performed

2.1 Data Cleaning & Exploration
- Verified for missing/null values and dropped none (dataset was clean)
- Checked unique values for crop labels and feature ranges
- Validated data types for all columns

2.2 Descriptive Analytics & Visualizations
- Crop label distribution (bar chart)
- Feature distributions (histograms for N, P, K, temperature, humidity, ph, rainfall)
- Correlation heatmap among soil and weather features
- Scatter plot between rainfall and humidity

2.3 Machine Learning & Predictive Modelling

• Features vectorized using VectorAssembler
• Label encoding using StringIndexer
• Decision Tree Classifier trained on 80% data, tested on 20%
• Model accuracy achieved: ~95%
• Generated feature importance visualization

3. Key Insights

3.1 Soil Nutrient Patterns
• Average Nitrogen: 50.6; Phosphorus: 54.8; Potassium: 48.2
• Balanced nutrient levels observed for most crop regions
• Crops like rice and maize required higher nitrogen and rainfall levels

3.2 Environmental Insights
• Ideal temperature range for most crops: 20–30°C
• High humidity positively correlated with rainfall ($r \approx 0.82$)
• Crops such as coffee and rice thrive in high rainfall zones (>150 mm)

3.3 Crop-Specific Trends
• Label 'rice' has the highest occurrence, followed by 'maize' and 'chickpea'
• Acidic soils (pH < 6) often linked with rice and banana
• Neutral pH (6–7) suitable for most pulses and fruits

3.4 Model Performance
• Decision Tree achieved 95% accuracy
• Key influencing features: rainfall, humidity, and nitrogen
• Misclassifications observed between similar climatic crops (rice vs maize)

4. Recommendations

4.1 Agricultural Planning
• Farmers should prioritize crop selection based on local rainfall and humidity data
• Soil nutrient balancing (NPK ratio) is critical for optimal yield

4.2 Fertilizer Management
• Regions with low nitrogen should increase urea-based fertilization
• Monitor soil pH regularly to maintain neutral range (6.5–7.5)

4.3 Environmental Strategy
• Use weather data prediction models to forecast optimal sowing periods
• Integrate irrigation control with real-time humidity monitoring

4.4 Technological Enhancement

- Build mobile apps for farmers to input soil readings and receive crop suggestions
- Use ML models for adaptive recommendations under changing climate conditions

5. Future Analytics Opportunities

- Predict soil fertility degradation over time using rainfall and crop rotation data
- Develop time-series forecasting for rainfall and temperature variations
- Expand dataset with satellite soil imaging data for precision farming
- Build ensemble models (Random Forest, Gradient Boosting) to improve accuracy
- Implement IoT-based real-time soil monitoring integrated with PySpark pipelines