Variety-wise Daily Market Prices Data of Commodity Analysis

Source : https://data.gov.in/
Github :
LinkedIn :

Dataset Overview

- The dataset contains 20447 rows and 11 columns.
- Columns include: State, District, Market, Commodity, Variety, Grade, Arrival_Date, Min_Price, Max_Price, Modal_Price...
- This dataset focuses on predicting or analyzing cotton yield based on factors such as rainfall, temperature, humidity, and pesticide use.

Data Cleaning & Preparation

- - Handled missing and non-numeric values (e.g., 'NA', 'NaN').
- - Casted numeric columns to proper data types.
- - Created engineered features like Rain-Temp Ratio and Humid-Temp Diff.
- - Ensured dataset consistency for machine learning.

Exploratory Data Analysis (EDA)

- - Checked summary statistics and correlations.
- - Visualized relationships between rainfall, temperature, and yield.
- - Observed how environmental factors affect cotton yield.
- - Identified trends for prediction model preparation.

Machine Learning Model

- - Used Linear Regression model with PySpark MLlib.
- - Selected numeric features like rainfall, temperature, humidity, etc.
- - Target variable: cotton yield.
- - Model provided feature importance, R^2 score, and RMSE.
- - Demonstrated rainfall and temperature as major influencing factors.

Visual Insights

- - Scatter plot between rainfall and yield shows a positive trend.
- - Heatmap shows correlation between weather and yield factors.
- - Helps in identifying key contributors for cotton productivity.

Key Findings

- - Rainfall and temperature significantly impact cotton yield.
- - Humidity and pesticide use also show strong correlation.
- - Proper climatic balance improves cotton production.
- - Model accuracy demonstrates potential for predictive agriculture.

Conclusion

The cotton prediction analysis demonstrates how data-driven techniques can optimize agricultural productivity. Using PySpark, data cleaning, and machine learning models, we can forecast yield trends and assist farmers and policymakers in better decision-making.