Rice Cultivation Daily Dataset Documentation

Overview

The **Rice Cultivation Daily Dataset** is a synthetic dataset designed to facilitate analysis of daily weather conditions for rice cultivation suitability. Each record represents environmental conditions for a single day, labeled as "suitable" or "not suitable" for rice growth based on predefined criteria derived from agricultural guidelines. The dataset is intended for researchers, data scientists, and agricultural analysts to study the impact of daily weather parameters on rice cultivation or to develop predictive models for suitability.

Dataset Description

- **Size**: 10,000 records (days)
- **Format**: CSV
- **File Name**: rice_cultivation_daily_dataset.csv
- **Location**: Saved in the data/ directory
- **Purpose**: To provide a realistic dataset for evaluating whether daily weather conditions are conducive to rice cultivation across different seasons.

Columns

The dataset contains six columns, each representing a feature or label:

Column Name	Data Type	Description
season	String	The agricultural season: pre-kharif, kharif, or rabi.
average_day_temp	Float	Average daytime temperature (°C) for the day, rounded to 1 decimal place.
average_night_temp	Float	Average nighttime temperature (°C) for the day, rounded to 1 decimal place.

daily_sunshine_hours	Float	Number of sunshine hours in the day, ranging from 0 to 12 hours.
daily_rainfall	Float	Rainfall amount (mm) for the day, non-negative.
suitability	String	Label indicating suitability for rice cultivation: suitable or not suitable.

Seasons

- **Pre-kharif**: Early monsoon transition period (warmer, moderate rainfall).
- **Kharif**: Main monsoon season (high rainfall, slightly cooler).
- **Rabi**: Post-monsoon season (cooler, lower rainfall).

Data Generation Process

The dataset is generated using Python with the pandas and numpy libraries, ensuring realistic variability and seasonal differences. Key aspects of the generation process include:

1. Season Assignment

- Seasons are assigned randomly to each record with probabilities:
    - Pre-kharif: 30%
    - Kharif: 40%
    - Rabi: 30%

2. Parameter Generation

Each parameter is generated with season-specific characteristics, incorporating randomness and correlations to mimic real-world conditions.

Season-Specific Parameters

Season	Day Temp (°C)	Night Temp (°C)	Sunshine Hours	Rainfall (mm)
Pre-kharif	Mean: 30, SD: 3	Mean: 22, SD: 1.5	Mean: 8, SD: 2	Prob: 0.4, Mean: 10, SD: 5

Kharif	Mean: 28, SD: 2	Mean: 21.5, SD: 1	Mean: 7, SD: 1.5	Prob: 0.6, Mean: 15, SD: 7
Rabi	Mean: 25, SD: 2.5	Mean: 20, SD: 1.2	Mean: 9, SD: 2	Prob: 0.3, Mean: 5, SD: 3

- **Day and Night Temperatures**:
  - Generated using a normal distribution with season-specific means and standard deviations.
  - Measurement noise (±0.5°C) is added to simulate real-world sensor inaccuracies.
- **Daily Sunshine Hours**:
  - Generated using a normal distribution, correlated with day temperature (warmer days tend to have more sunshine).
  - Clipped to 0–12 hours to reflect maximum daylight.
- **Daily Rainfall**:
  - Modeled as a two-step process:
    - A Bernoulli distribution determines if it rains (based on rainfall_prob).
    - On rainy days, rainfall amount is drawn from a normal distribution.
    - Non-rainy days have 0 mm rainfall.
    - Clipped to ensure non-negative values.

3. Extreme Events

- 5% of samples (500 records) are randomly selected to experience extreme events:
  - **Extreme Temperature**: Day temperature is shifted by ±5°C, night temperature by ±3°C (50% of extreme events).
  - **Extreme Rainfall**: Rainfall is multiplied by 0.5 or 2 (50% of extreme events).
- These events simulate real-world anomalies like heatwaves or heavy rain.

4. Data Constraints

- Temperatures are rounded to one decimal place for consistency.
- Sunshine hours are constrained to 0–12 hours.
- Rainfall is non-negative, with realistic upper bounds based on extreme events.

Suitability Criteria

A day is labeled as suitable for rice cultivation if **all** of the following conditions are met:

Parameter	Condition
Average Day Temperature	20°C ≤ average_day_temp ≤ 36°C
Average Night Temperature	20°C ≤ average_night_temp ≤ 23°C
Daily Sunshine Hours	daily_sunshine_hours ≥ 6 hours
Daily Rainfall	2 mm ≤ daily_rainfall ≤ 30 mm

- If any condition is not met, the day is labeled not suitable.
- **Note**: The daily rainfall condition assumes that irrigation is not available to compensate for low rainfall days. In practice, irrigation could relax the lower bound (2 mm), but this dataset uses a strict threshold for simplicity.

Rationale for Daily Conditions

- **Temperatures**: The ranges (20–36°C for day, 20–23°C for night) are derived from rice cultivation guidelines, applicable to daily averages as they reflect optimal growth conditions.
- **Sunshine Hours**: A threshold of ≥6 hours ensures sufficient sunlight for photosynthesis, typical for tropical growing seasons.
- **Rainfall**: The 2–30 mm range reflects adequate moisture for rice (a semi-aquatic crop) without excessive flooding, based on typical daily rainfall in rice-growing regions.

Usage Notes

- **Analysis**: The dataset can be used to:
    - Study the impact of daily weather parameters on rice cultivation suitability.
    - Train machine learning models to predict suitability based on weather conditions.
    - Analyze seasonal differences in suitability rates.
- **Limitations**:
    - The dataset is synthetic and does not account for irrigation, soil type, or stage-specific requirements (e.g., flowering, grain formation).

    - Daily rainfall suitability assumes no irrigation, which may not reflect real-world practices.
    - Extreme events are simplified and may not capture all types of weather anomalies.
- **Extending the Dataset**:
    - Additional parameters (e.g., humidity, wind speed) can be included to enhance realism.
    - Suitability criteria can be adjusted to reflect specific rice varieties or regional practices.