

# Australian Weather Analysis and Rain Prediction

Analyzing Australian weather data to predict rainfall using machine learning models. Our study examines weather patterns and builds classification models to forecast rain events.

 作者: sen jay





# Project Overview



## Data Collection

Weather data from Australia with daily observations.



## Data Cleaning

Imputation, renaming, and feature engineering.



## Analysis

EDA focusing on target variable balance.



## Modeling

Random Forest and Logistic Regression with GridSearchCV.



Made with Gamma

# Data Wrangling Methods

1

## Initial Dataset

3271 entries with 22 columns of weather observations.

2

## Column Renaming

RainToday → RainYesterday, RainTomorrow → RainToday for clarity.

3

## Missing Values

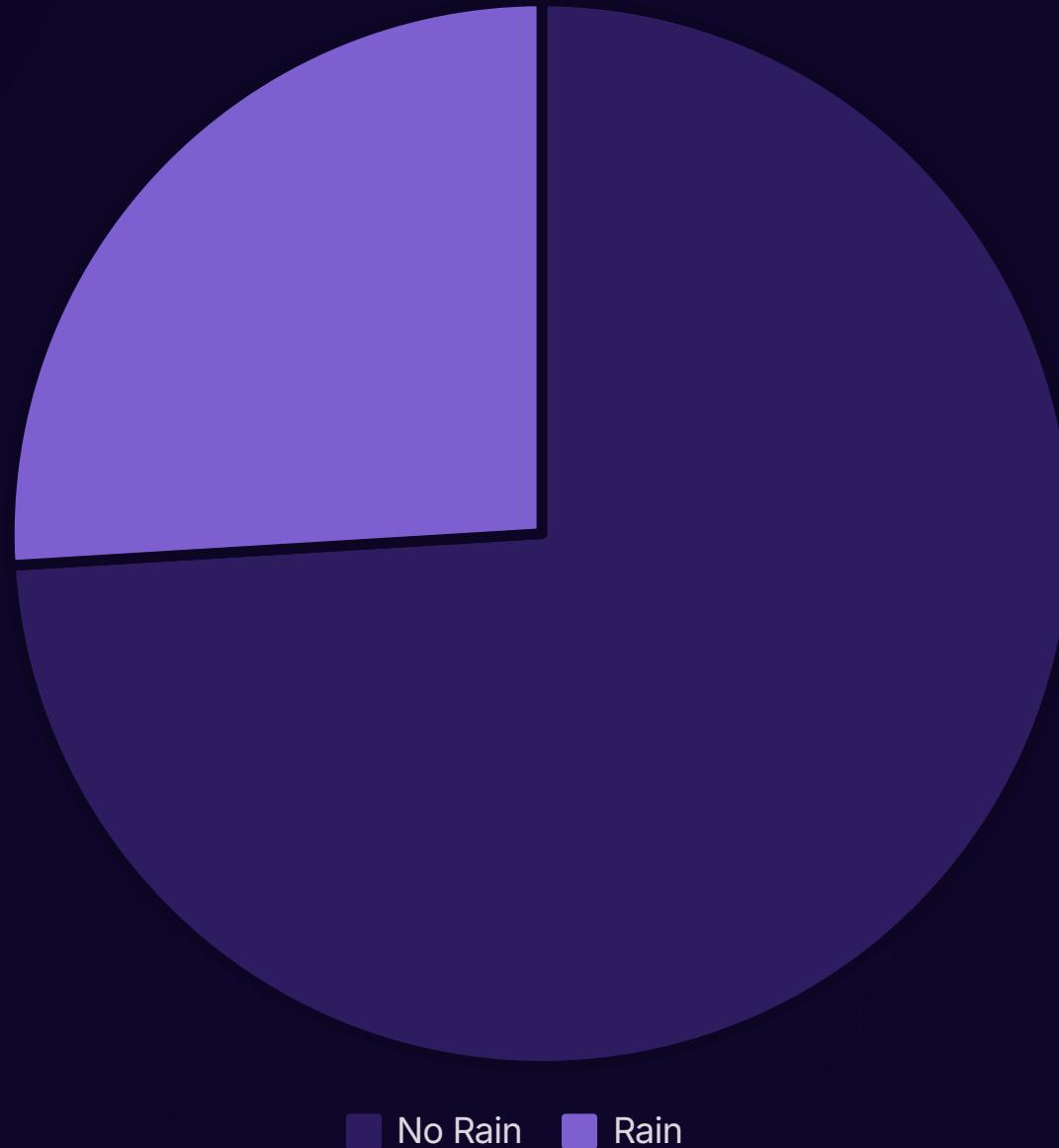
Used SimpleImputer within sklearn pipeline.

4

## Feature Engineering

Created 'Season' column based on 'Date'.

# Target Variable Distribution



The dataset shows significant imbalance: 74% no-rain days versus 26% rain days. This imbalance affects model training and evaluation strategies.

# Preprocessing Pipeline



## Feature Detection

Automatic identification of numeric and categorical features.



## Numeric Processing

Imputation with mean values followed by StandardScaler.



## Categorical Processing

Imputation with most frequent values followed by OneHotEncoder.



Made with Gamma

# Random Forest Results

## Best Parameters

- max\_depth: 10
- min\_samples\_split: 5
- n\_estimators: 50

## Performance

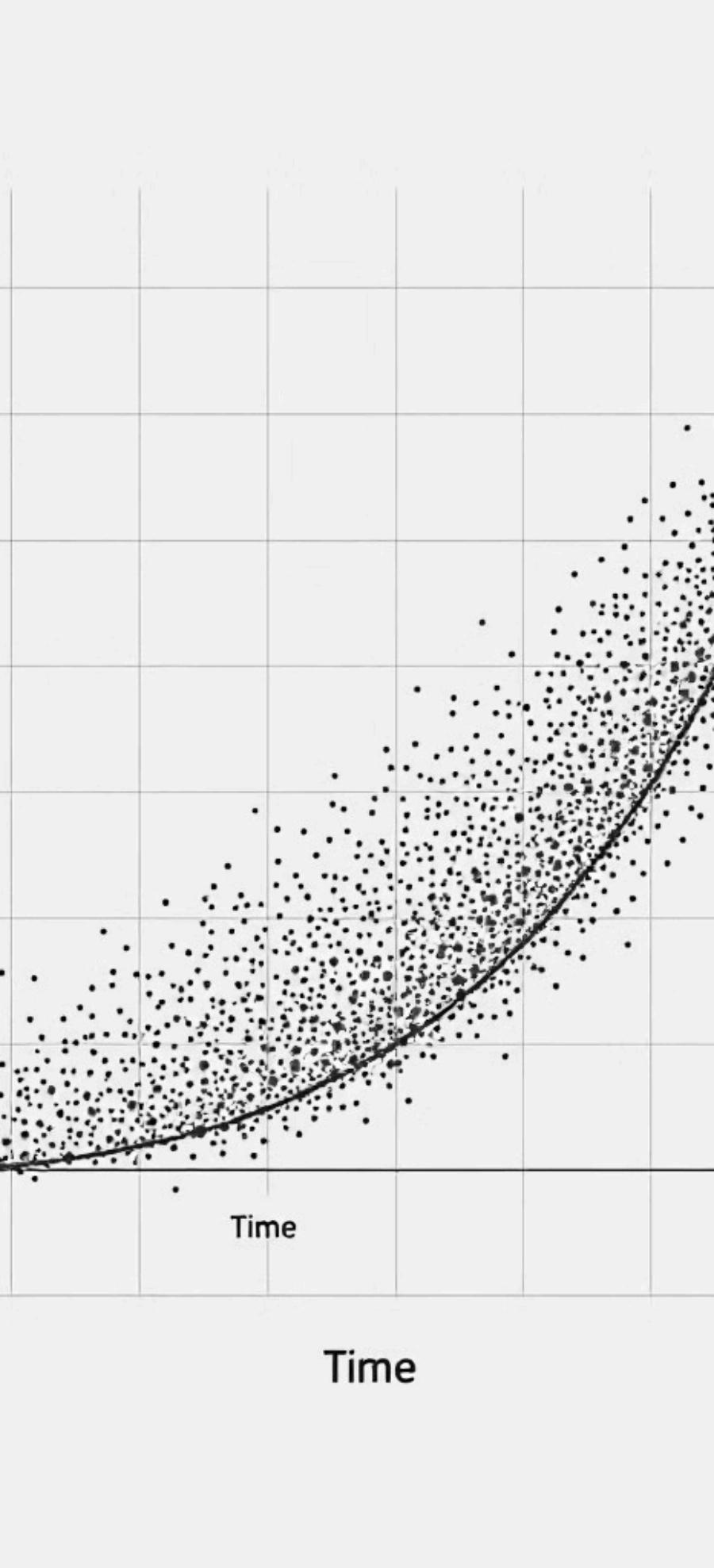
- CV Accuracy: 84.33%
- Test Accuracy: 81.68%

## Rain Prediction

- Precision: 0.73
- Recall: 0.47
- F1-Score: 0.57



# Logistic Regression Results



**82.29%**

**Test Accuracy**

Slightly better than Random Forest

**0.69**

**Precision**

For rain prediction

**0.58**

**Recall**

Better at identifying actual rain days

**0.63**

**F1-Score**

For rain prediction

# Key Findings & Next Steps

## Key Findings

- Humidity3pm is the most important predictor
- Both models achieve ~82% accuracy
- Logistic Regression better identifies rain days
- Class imbalance affects model performance

## Next Steps

- Address class imbalance with SMOTE
- Try additional algorithms
- Conduct deeper feature analysis
- Evaluate with Precision-Recall AUC