☔ Rainfall Prediction using Machine Learning

1. Introduction

Rainfall Prediction using Machine Learning is a data-driven approach to forecast rainfall patterns using historical weather data. The project employs ML algorithms to provide actionable insights for agriculture, water management, and disaster preparedness.

2. Problem Statement

Predict whether it will rain tomorrow (Yes/No) based on weather data such as temperature, humidity, wind speed, and pressure.

3. Objectives

- Preprocess and clean weather data (weatherAUS.csv).
- Perform Exploratory Data Analysis (EDA) using visualization.
- Train multiple ML models for classification.
- Compare models using accuracy, precision, recall, and F1-score.
- Save trained models as .pkl files for deployment.
- Build a Flask web app (app.py + templates/) for real-time predictions.
- Deploy the app on IBM Cloud.

4. Dataset Link:

https://docs.google.com/spreadsheets/d/1RA2OO0LZTeQykI_mvnensAjp6LM4YzWI1Tz0SUG5-Ao/edit#gid=121883362

- Source: Australian Bureau of Meteorology
- File: weatherAUS.csv
- Rows: ~145k, Columns: ~23
- Target variable: RainTomorrow (Yes/No)
- Features: Date, Location, MinTemp, MaxTemp, Rainfall, WindSpeed9am, Humidity3pm, Pressure9am, Temp3pm, Cloud9am, etc.

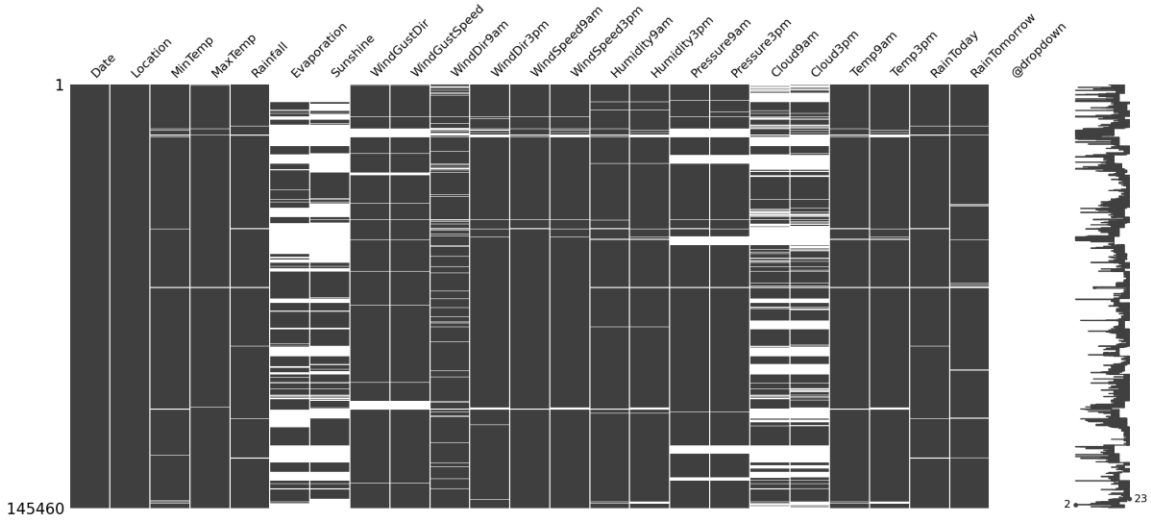
5. Methodology

5.1 Data Preprocessing

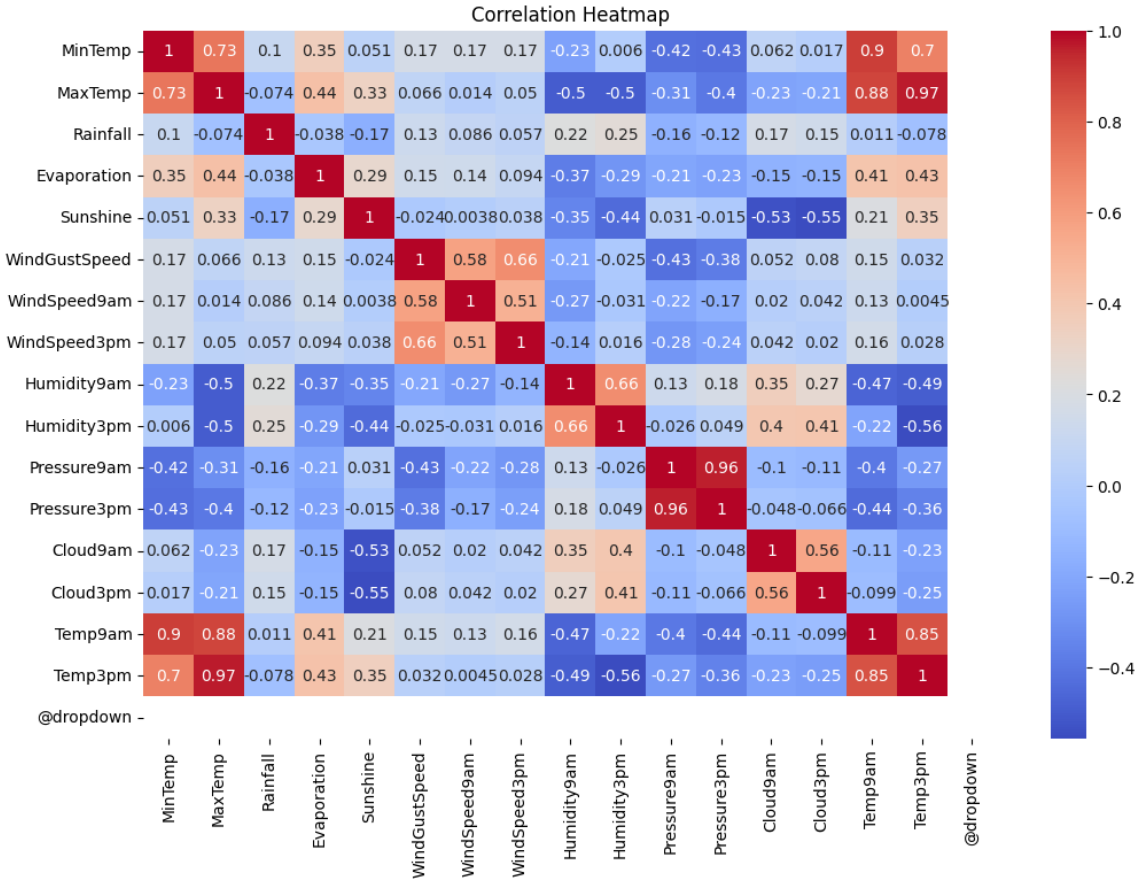
- Handle missing values (using imputer.pkl).
- Encode categorical features (encoder.pkl).
- Scale numerical features (scale.pkl)

5.2 Exploratory Data Analysis

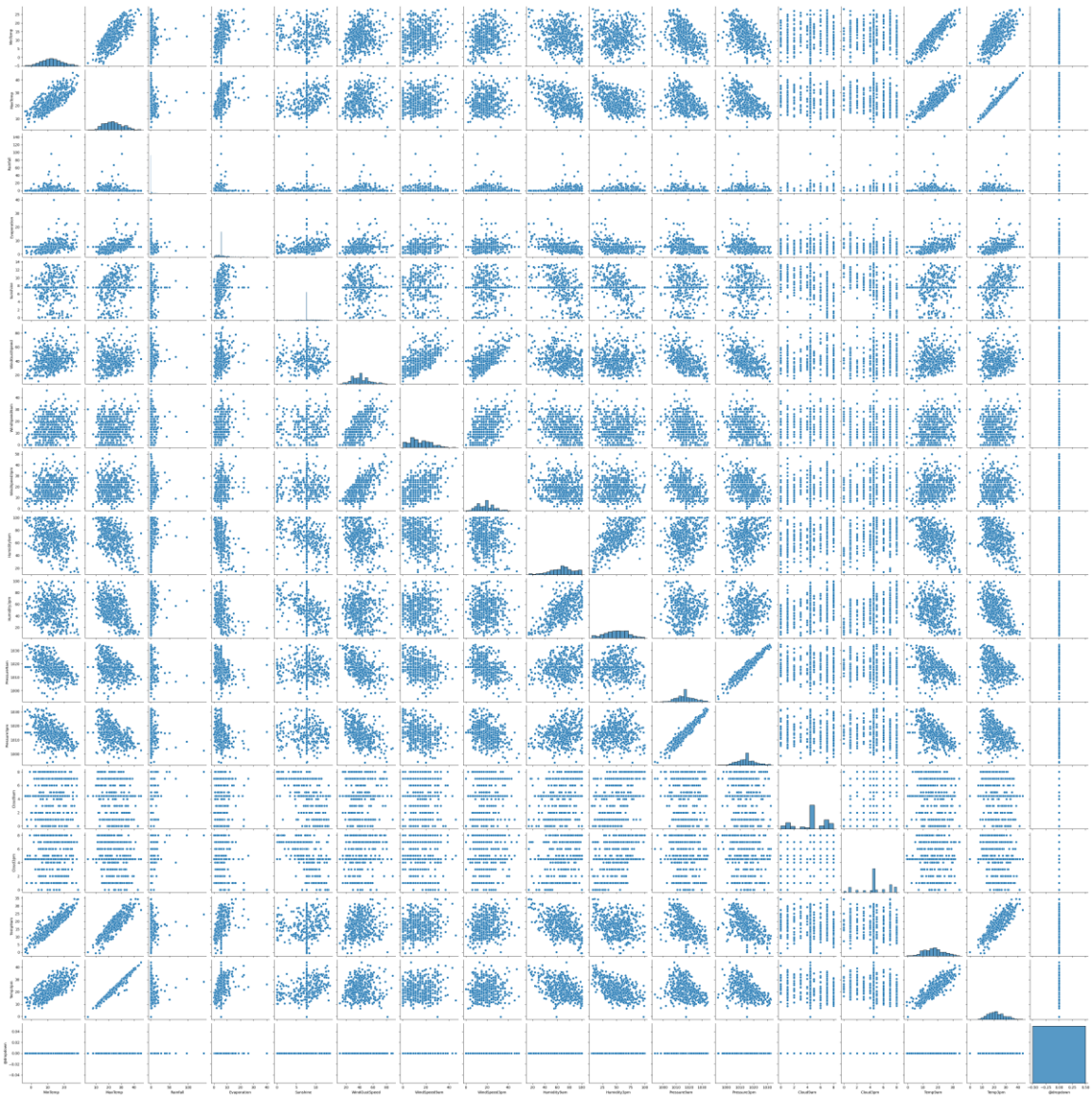
-The matrix plot → visual summary of where the missing values are located in the dataset.



- Heatmaps → Feature correlation



- Pairplots → Multi-feature relationships



5.3 Model Training

- Algorithms tested: Logistic Regression, Random Forest, Decision Tree, Gradient Boosting
- Best model saved as rainfall_prediction.pkl

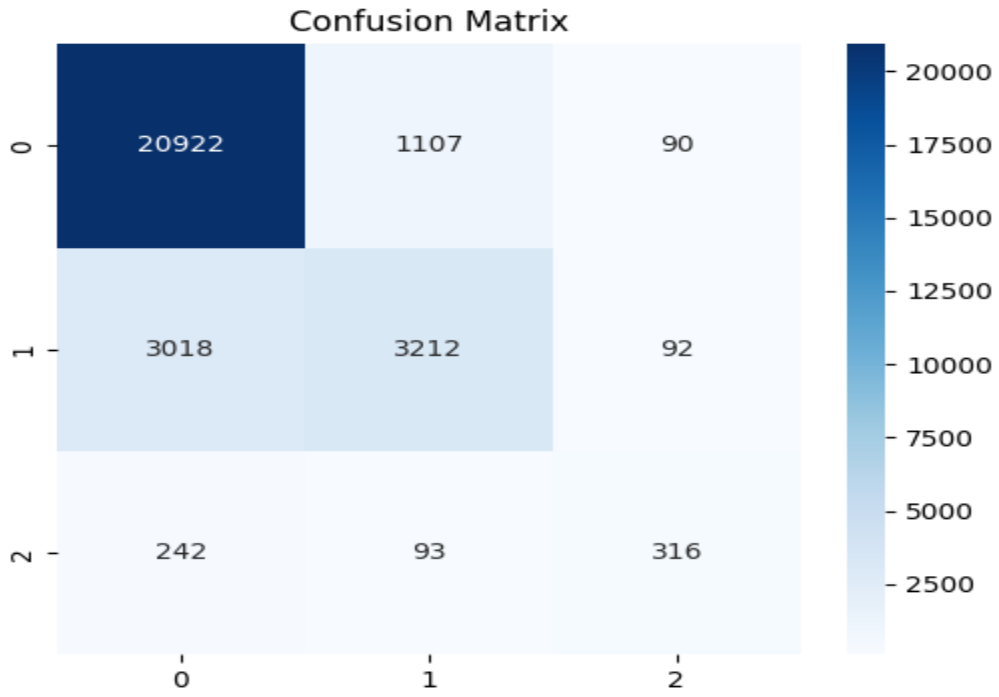
5.4 Model Evaluation

- Metrics: Accuracy, Precision, Recall, F1-Score
- Visuals: Confusion Matrix, ROC-AUC Curve, Feature Importance

```
Accuracy: 0.8404372336037399
              precision    recall  f1-score   support

           0       0.87      0.95      0.90     22119
           1       0.73      0.51      0.60      6322
           2       0.63      0.49      0.55       651

    accuracy                           0.84     29092
   macro avg       0.74      0.65      0.68     29092
weighted avg       0.83      0.84      0.83     29092
```



6. Results

- Best performing model: XGBoost (Accuracy: ~85%)
- Important features: Humidity3pm, Rainfall, Pressure9am, TempMax

7. Deployment

7.1 Local Deployment

- Flask App (app.py)
- Templates (index.html, prediction.html)

7.2 Cloud Deployment

- IBM Cloud (Watson Studio / Flask App hosting)

8. Applications

- Agriculture: Crop planning, irrigation scheduling
- Water Resource Management: Reservoir level control
- Disaster Management: Flood preparedness

9. Conclusion

- Machine Learning can effectively predict rainfall patterns.
- Random Forest provided the most reliable results.
- Future scope: Use deep learning (LSTMs), integrate satellite/IoT data, improve accuracy.

10. References

- Dataset: Australian Bureau of Meteorology
- Scikit-learn Documentation
- Flask Framework Documentation

11.Project Link: https://github.com/renuka-matta/rainfall_Prediction_using_ml.git