Model Performance Comparison: Rainfall Prediction in Australia

1. Introduction

This report presents a comparative analysis of seven machine learning models trained and evaluated for the task of **Rainfall Prediction in Australia**. The objective is to identify the overall best-performing model and to determine the most suitable models for specific, critical prediction targets, such as minimizing false alarms or maximizing the correct prediction of rain.

The models were evaluated on a binary classification task, where **Class 1** represents the prediction of **"Rain Tomorrow"** and **Class 0** represents **"No Rain Tomorrow"**. The dataset used for evaluation is consistent across all models, with a total of 28,439 samples, including 6,375 instances of Class 1 (Rain) and 22,064 instances of Class 0 (No Rain).

2. Model Performance Summary

The following table summarizes the key performance metrics for all seven models. The metrics are focused on the positive class (Rain Tomorrow) and the overall weighted average.

Model ID	Model Type	Overall Accuracy	F1-Score (Weighted Avg)	Precision (Rain)	Recall (Rain)	F1-Score (Rain)
M6	XGBoost/GBM (Optimized)	**0.851**	**0.84**	**0.75**	0.51	0.60
M1	XGBoost/GBM	0.837	0.83	0.65	0.58	0.61
M2	Random Forest	0.832	0.83	0.63	0.61	0.62

M7	XGBoost/GBM	0.814	0.82	0.57	0.67	0.62
M4	XGBoost/GBM	0.807	0.81	0.56	**0.69**	**0.62**
M5	SVM	0.768	0.78	0.49	0.75	0.59
M3	Logistic Regression	0.698	0.72	0.41	**0.83**	0.55

Note: Precision (Rain) is the ratio of correctly predicted rain events to all events predicted as rain (minimizing False Positives). Recall (Rain) is the ratio of correctly predicted rain events to all actual rain events (minimizing False Negatives).

3. Detailed Analysis and Best Model Selection

The selection of the "best" model depends heavily on the specific business or operational objective. In the context of weather prediction, different metrics correspond to different priorities:

3.1. Best Overall Model (Highest General Performance)

The **Overall Accuracy** and **Weighted Average F1-Score** are the best indicators of general performance across both classes.

> **Best Model: Model 6 (XGBoost/GBM Optimized)**
>
> **Key Metrics:** Accuracy: **0.851**, F1-Score (Weighted Avg): **0.84**
>
> **Analysis:** Model 6 demonstrates the highest overall correctness. However, its lower Recall (0.51) for the "Rain" class suggests it is conservative in predicting rain, leading to a higher number of missed rain events (False Negatives).

3.2. Best Model for Specific Targets

Target 1: Maximizing Correct Rain Prediction (Minimizing Missed Rain Events)

> **Objective:** To ensure that as many actual rain events as possible are correctly predicted. This is crucial for applications where missing a rain event

(False Negative) is costly (e.g., agriculture, outdoor events planning).

Metric to Maximize: Recall (Rain)

Best Model: Model 3 (Logistic Regression)

Key Metric: Recall (Rain): **0.83**

Analysis: Model 3 correctly identified 83% of all actual rain events. However, this high recall comes at the cost of very low Precision (0.41), meaning it also predicted rain when it didn't happen (False Positives) very frequently.

Target 2: Minimizing False Alarms (Reducing "Nuisance" Predictions)

Objective: To minimize the number of times the model predicts rain, but it does not actually rain (False Positives). This is important for applications where a false alarm is disruptive or costly (e.g., unnecessary activation of warning systems).

Metric to Maximize: Precision (Rain)

Best Model: Model 6 (XGBoost/GBM Optimized)

Key Metric: Precision (Rain): **0.75**

Analysis: When Model 6 predicts rain, it is correct 75% of the time. This makes it the most reliable model for minimizing false alarms.

Target 3: Best Balance Between Precision and Recall (General Rain Prediction)

Objective: To find a model that offers a good trade-off between minimizing missed rain events and minimizing false alarms.

Metric to Maximize: F1-Score (Rain)

Best Models: Model 2 (Random Forest) and **Model 4**

(XGBoost/GBM) Key Metric: F1-Score (Rain): **0.62**

Analysis: Both Model 2 and Model 4 achieved the highest F1-Score for the "Rain" class, indicating the best harmonic mean between Precision and Recall.

Model 2: (Precision 0.63, Recall 0.61) - Slightly better balance.

Model 4: (Precision 0.56, Recall 0.69) - Leans slightly more towards catching rain (higher Recall).

Note:

- Although the **Model 1** is not the best mathematical model between **Model 2** and **Model 4** in F1-Score (Rain), but my priority is the highest precision with **good recall** too.
- All of the models their parameters are modified

4. Conclusion

The choice of the final model should align with the operational goal:

Operational Goal	**Best Model**	**Key Metric**	**Value**
Best Overall Performance	**Model 6**	Accuracy	**0.851**
Maximize Rain Detection (Minimize Missed Rain)	**Model 3**	Recall (Rain)	**0.83**
Minimize False Alarms (Reduce Nuisance)	**Model 6**	Precision (Rain)	**0.75**
Best Balanced Rain Prediction	**Model 2 / Model 4**	F1-Score (Rain)	**0.62**

For a general-purpose prediction system where overall correctness is paramount, **Model 6** is the clear winner. However, if the priority is to avoid missing any rain event, even at the expense of some false alarms, **Model 3** is the most effective. The **Random Forest (Model 2)** offers the most robust balance for the “Rain” class.
Model 1 is my own choice for my target.