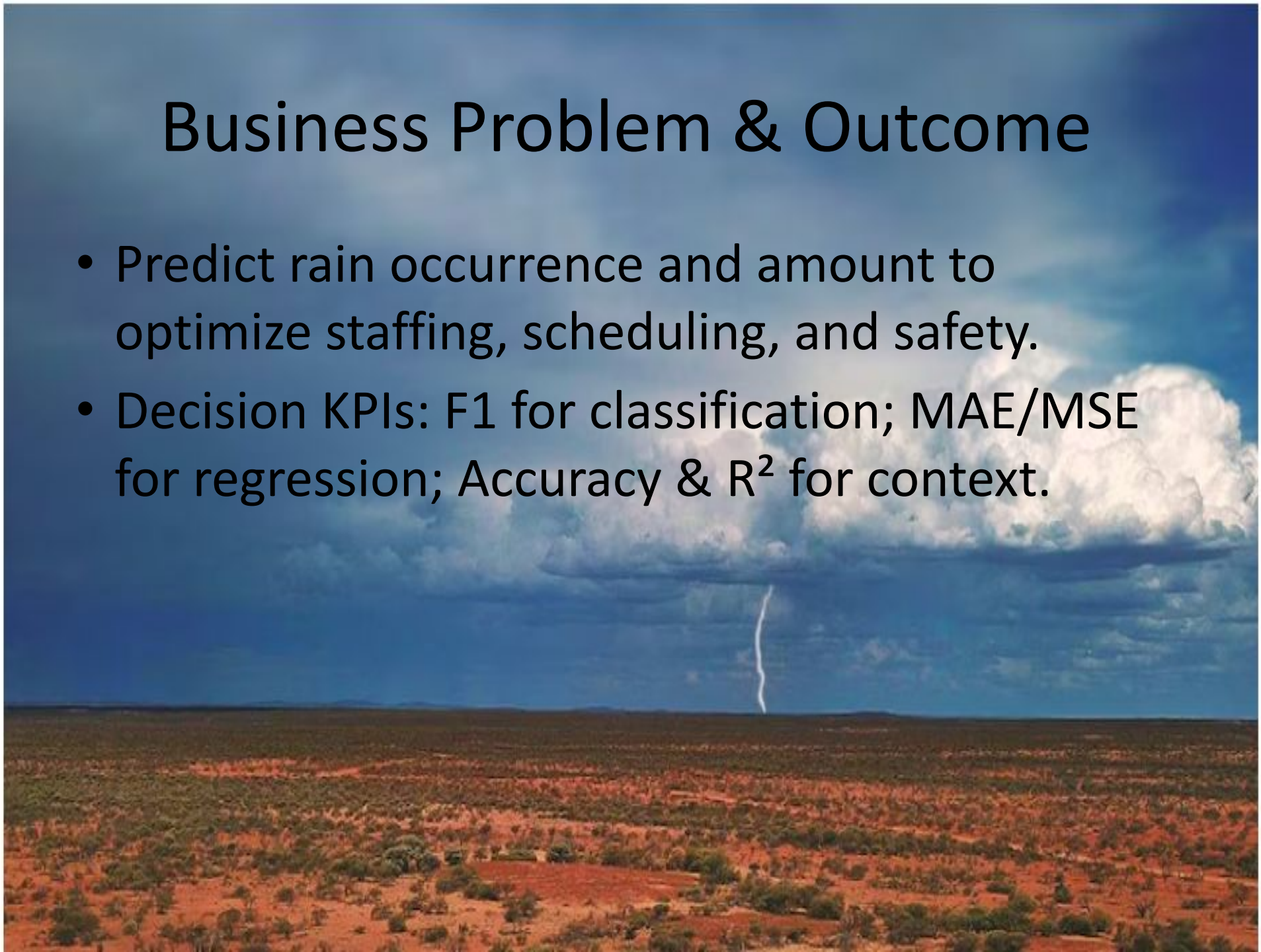
Australian Rain — Technical & Executive Summary

- Audience: Technical staff + Executives
- Deliverables: EDA framing, preprocessing approach, evaluation matrices, clear recommendations

Business Problem & Outcome

- Predict rain occurrence and amount to optimize staffing, scheduling, and safety.
- Decision KPIs: F1 for classification; MAE/MSE for regression; Accuracy & R^2 for context.

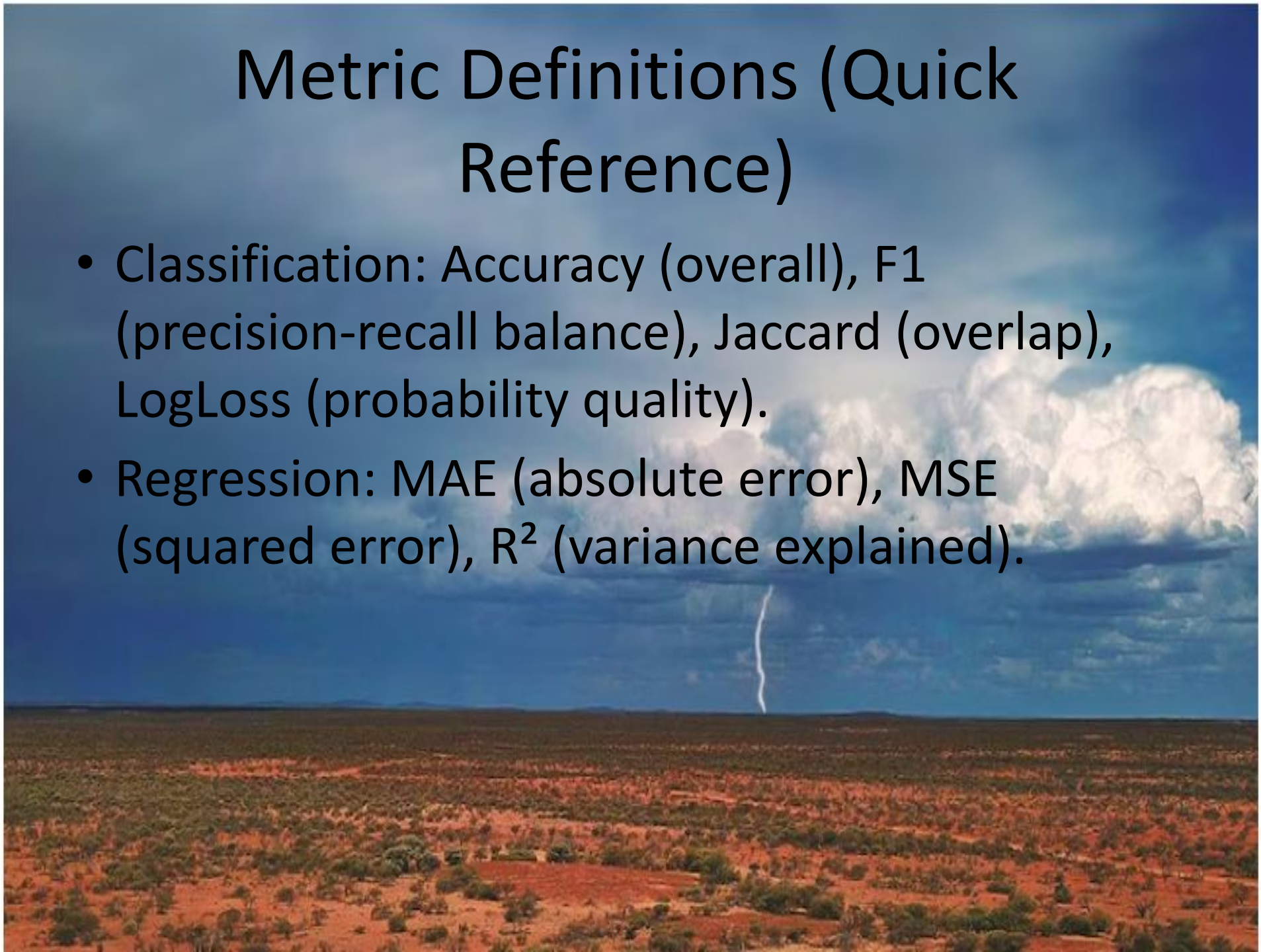


Data & Features (from project)

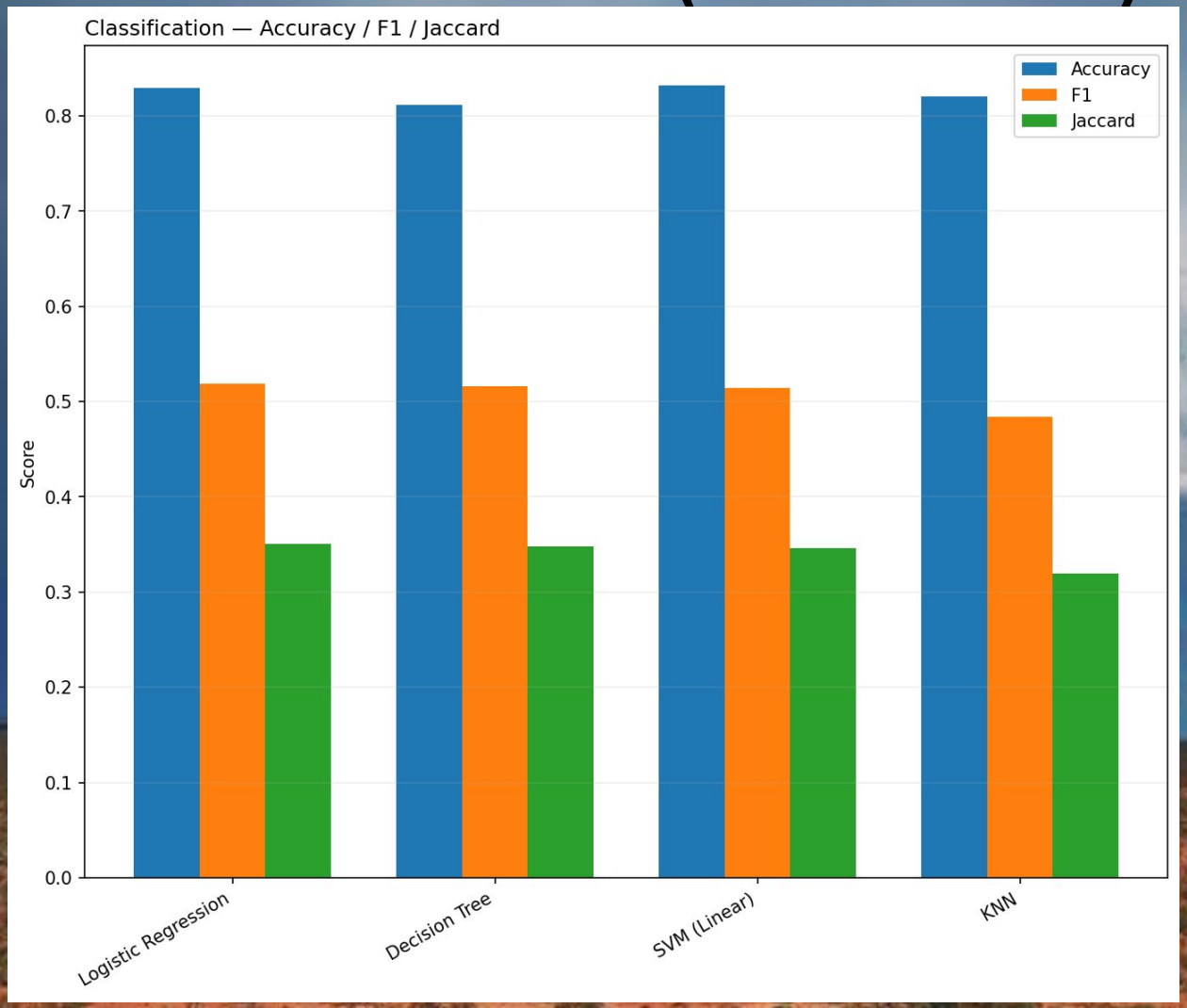
- Targets: RainTomorrow (0/1) and Rainfall (mm).
- Core signals: temperature, humidity, pressure, wind, cloud, RainToday.
- Preprocessing: impute missing values; standardize for KNN/SVM; train/test split.

Metric Definitions (Quick Reference)

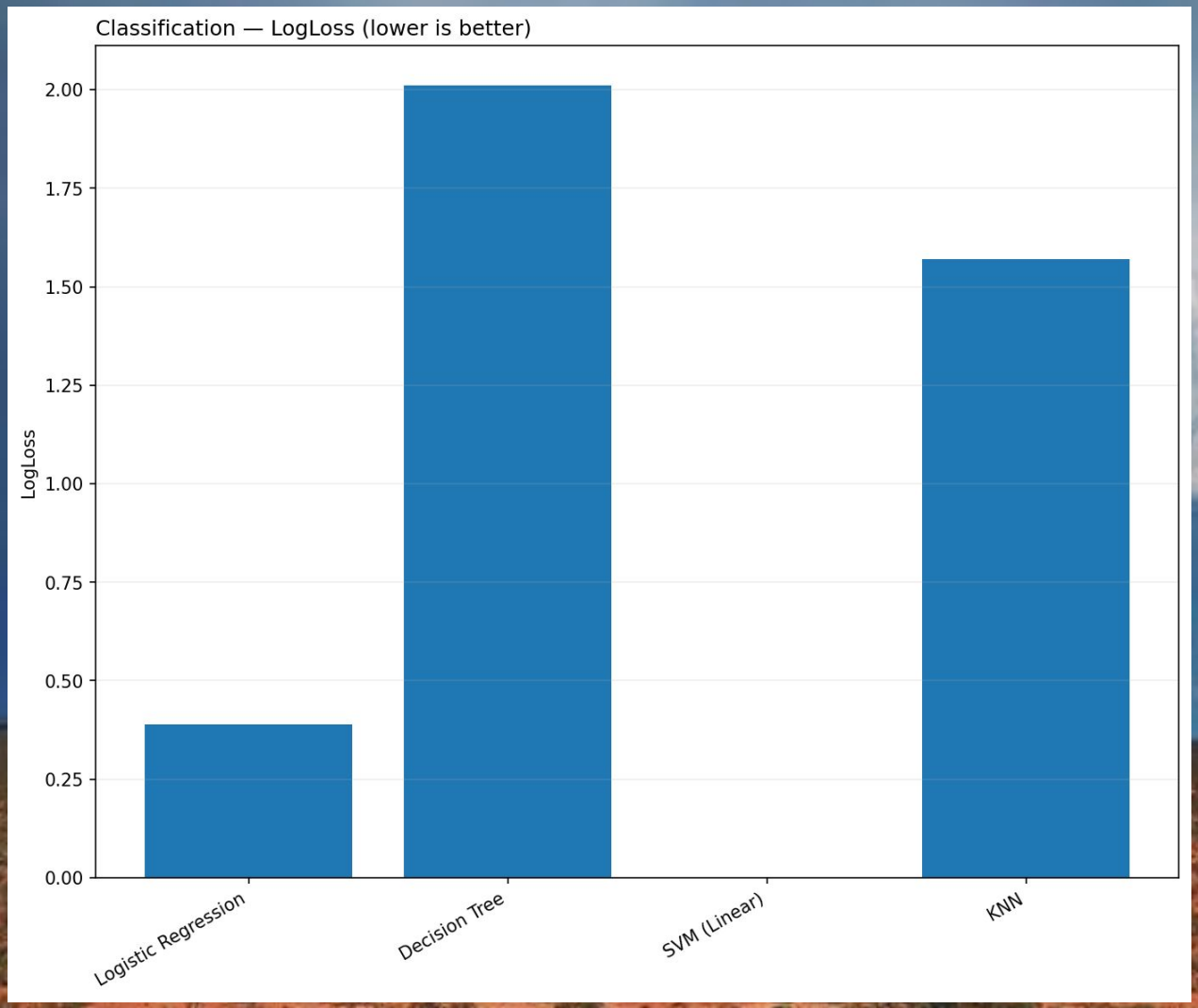
- Classification: Accuracy (overall), F1 (precision-recall balance), Jaccard (overlap), LogLoss (probability quality).
- Regression: MAE (absolute error), MSE (squared error), R^2 (variance explained).



Classification — Accuracy / F1 / Jaccard (from CSV)



Classification — LogLoss (from CSV)



Classification — Evaluation Matrix (from CSV)

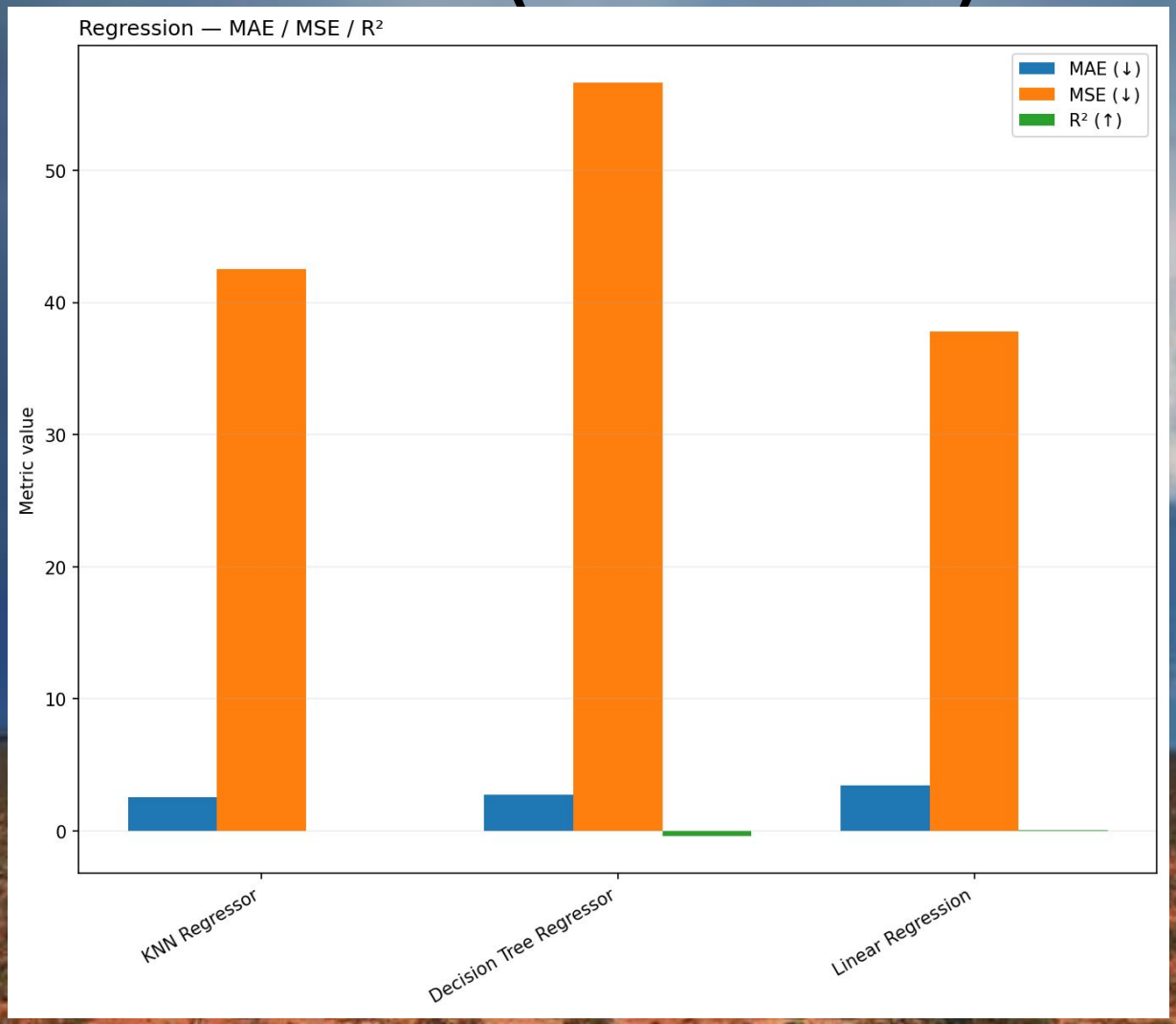
Classification — Evaluation Matrix (from CSV)

Model	Accuracy	Jaccard	F1-Score	LogLoss	LogLoss_filled
Logistic Regression	0.830	0.351	0.519	0.388	0.388
Decision Tree	0.811	0.348	0.516	2.011	2.011
SVM (Linear)	0.832	0.346	0.514	nan	2.011
KNN	0.821	0.319	0.484	1.570	1.570

Classification — Model Choice & Rationale

- Recommended: Logistic Regression (F1=0.519 • Acc=0.830 • LogLoss=0.388)
- Why: F1 prioritizes minority rain events; Accuracy can be misleading under imbalance.
- Action: Tune decision threshold for operational cost (false negatives vs false positives).

Regression — MAE / MSE / R^2 (from CSV)



Regression — Evaluation Matrix (from CSV)

Regression — Evaluation Matrix (from CSV)

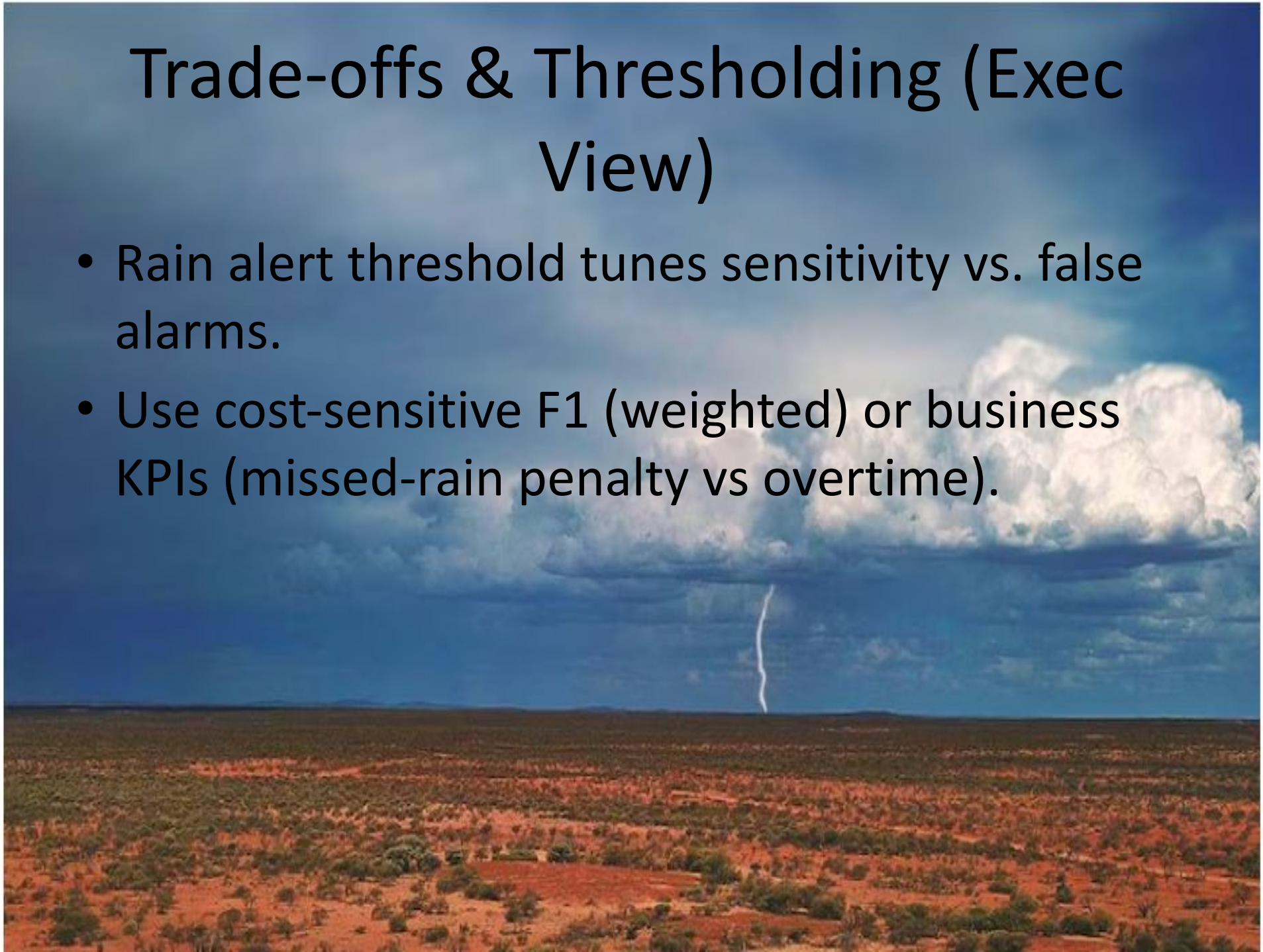
Model	MAE	MSE	R2
KNN Regressor	2.565	42.495	-0.020
Decision Tree Regressor	2.783	56.618	-0.359
Linear Regression	3.432	37.797	0.093

Regression — Model Choice & Rationale

- Recommended: KNN Regressor (MAE=2.57 • MSE=42.49 • R^2=-0.020)
- Why: Minimizes absolute & squared error; R^2 confirms fit on holdout.
- Action: Review error by location/season; consider quantile or zero-inflated models if heavy tails.

Trade-offs & Thresholding (Exec View)

- Rain alert threshold tunes sensitivity vs. false alarms.
- Use cost-sensitive F1 (weighted) or business KPIs (missed-rain penalty vs overtime).



Risks & Mitigations

- Data shift (seasonality, station changes) → set drift monitors; retrain cadence monthly/seasonal.
- Imbalance → monitor F1/recall; recalibrate probabilities where available.
- Operational latency → keep features available early (9am/3pm readings).

Roadmap & Ownership

- Phase 1: Threshold tuning + KPI dashboard (F1, calibration; MAE/R^2).
- Phase 2: Per-city models + lag features; periodic retraining.
- Phase 3: Integrate into scheduling; A/B test cost reduction and service level.

Executive Decision

- Adopt Logistic Regression for rain alerts; operate on F1-optimized threshold.
- Adopt KNN Regressor for rainfall magnitude; use MAE as on-call staffing guardrail.
- Stand up monitoring & governance; iterate quarterly.