

# Rain in Australia

NOM: Ricard López Olivares

NIU: 1571136

Link dataset: <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>

Git Hub: <https://github.com/Ricardlol/Rain-in-Australia>





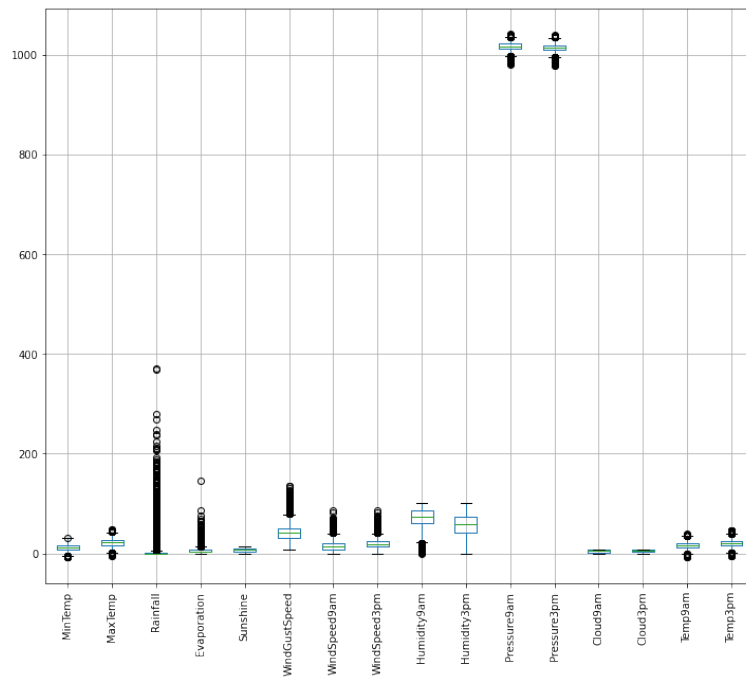
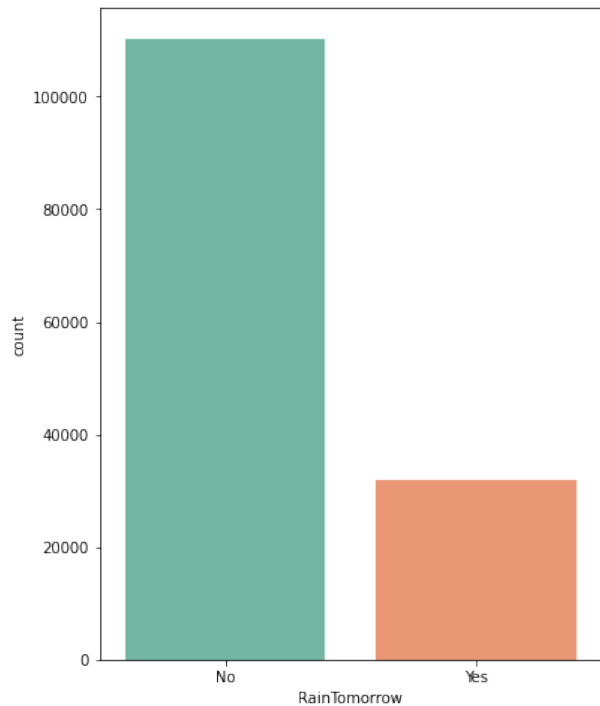
# Dataset

## Rain in Australia

- Predir si demà plourà
- 145460 mostres
- Atributs continus i discrets



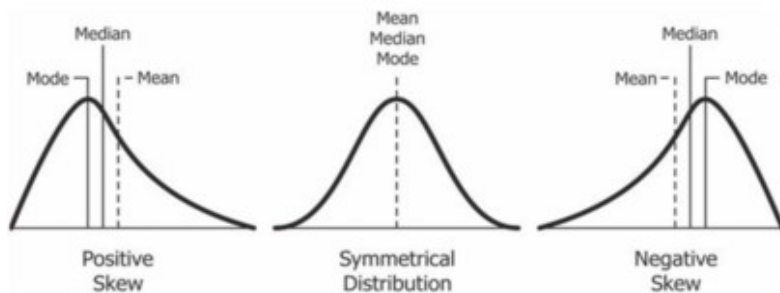
# Processat





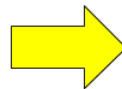
# Metodes utilitzats

## Skew



## OneHot Encoder

Color
Red
Red
Yellow
Green
Yellow



Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1



# Resultats i exemples

## Models per defecte amb dades estandaritzades i oversampling

Model	Hiperparametre	Accuracy	Recall	Precision	F1-score	Temps (s)
Regressió logística	No Parameters	0.787	0.774	0.794	0.787	5.70
Decision Tree	max_depth = 6	0.767	0.775	0.763	0.767	3.41
Random Forest	max_samples = 0.9	0.945	0.977	0.919	0.945	52.40
XGBOOST	objective='binary:logistic' random_state=0	0.838	0.838	0.835	0.838	34.42
KNN	n_neighbors=5	0.817	0.861	0.776	0.888	0.232



## Models sense estandaritzar i oversampling

Model	Hiperparametre	Accuracy	Recall	Precision	F1-score	Temps (s)
Regressió logística	No Parameters	0.777	0.759	0.781	0.774	3.14
SVC	No Parameters	-	-	-	-	no acaba
Decision Tree	max_depth = 6	0.766	0.732	0.787	0.762	2.65
Random Forest	max_samples = 0.9	0.949	0.977	0.919	0.949	406.40
XGBOOST	objective='binary:logistic' random_state=0	0.823	0.832	0.810	0.823	25.70
KNN	n_neighbors=5	0.802	0.810	0.805	0.801	0.263



## Models utilitzant skew amb oversampling

Model	Hiperparametre	Accuracy	Recall	Precision	F1-score	Temps (s)
Regressió logística	No Parameterse	0.761	0.752	0.755	0.761	5.62
Decision Tree	max_depth = 6	0.767	0.745	0.763	0.767	3.41
Random Forest	max_samples = 0.9	0.945	0.977	0.919	0.945	52.40
XGBOOST	objective='binary:logistic' random_state=0	0.838	0.837	0.836	0.838	34.42
KNN	n_neighbors=5	0.817	0.888	0.791	0.817	0.232



### Models utilitzant skew amb oversampling

Model	Hiperparametre	Accuracy	Recall	Precision	F1-score	Temps (s)
Regressió logística	No Parameters	0.787	0.774	0.794	0.787	5.70
Decision Tree	max_depth = 6	0.767	0.775	0.763	0.767	3.41
Random Forest	max_samples = 0.9	0.945	0.977	0.919	0.945	52.40
XGBOOST	objective='binary:logistic' random_state=0	0.838	0.838	0.835	0.838	34.42
KNN	n_neighbors=5	0.817	0.861	0.776	0.888	0.232





# Conclusió i treball futur

Label Encoder

