

Models

Martí Caixal Joaniquet

1563587

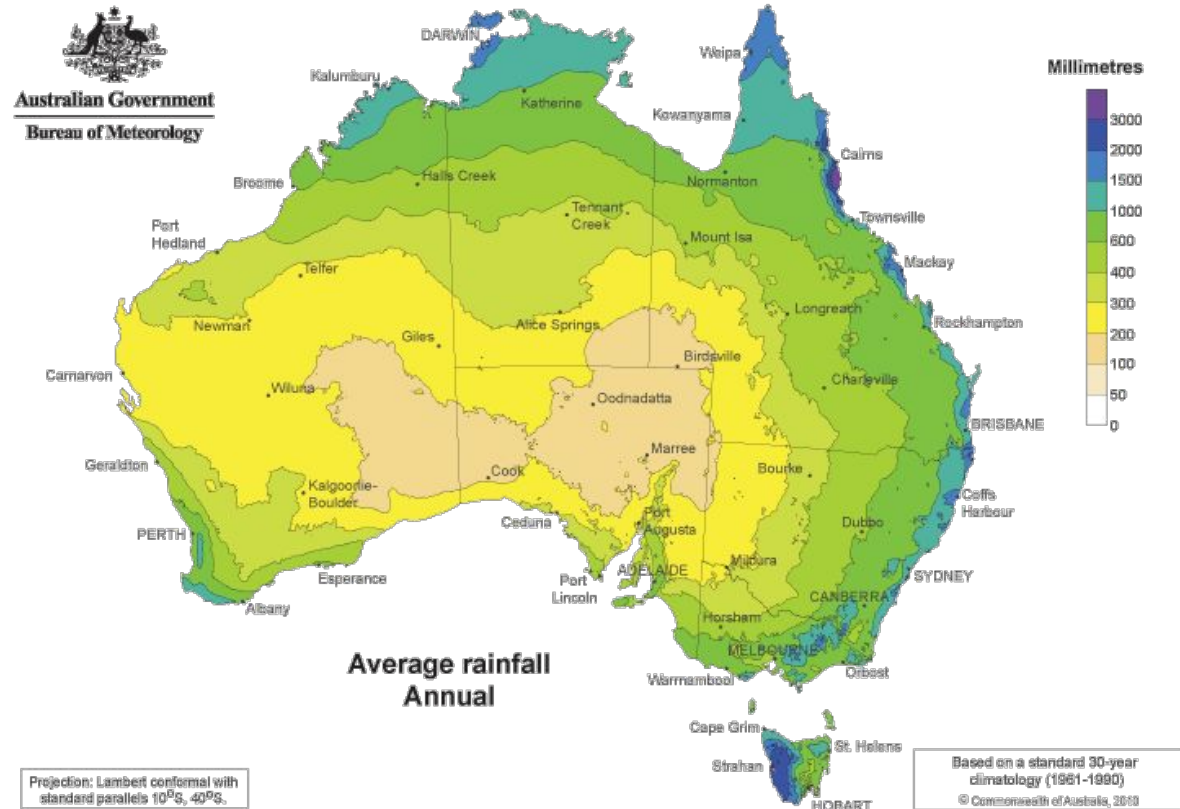
Ricard López Olivares

1571136

Jairo Villarroel Rodriguez

1571069

Base de dados

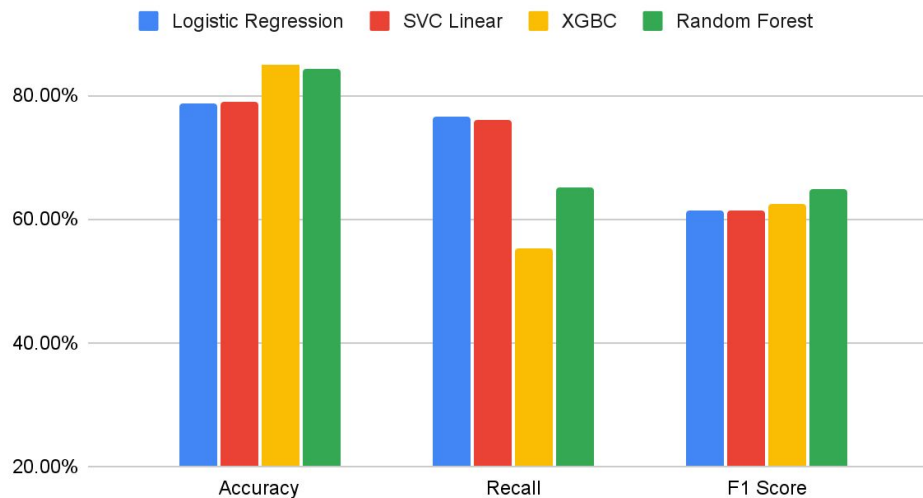


Apartat B

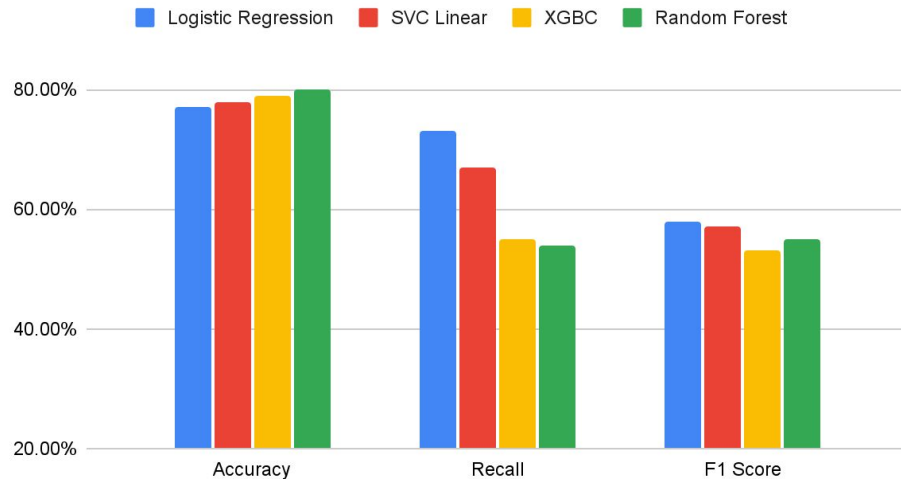


Comparant models

Test Size = 20%



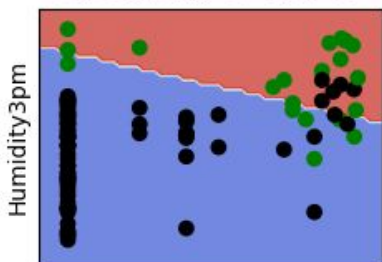
Test Size = 99.9%



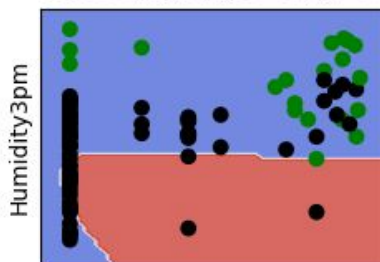


SVM Kernels

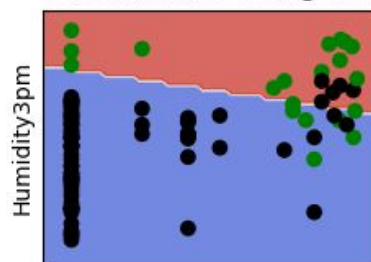
SVC linear, $C=5$, $g=5$



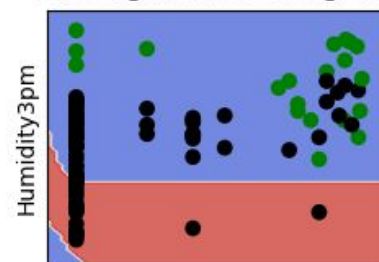
SVC Sigmoid, $C=5$, $g=5$



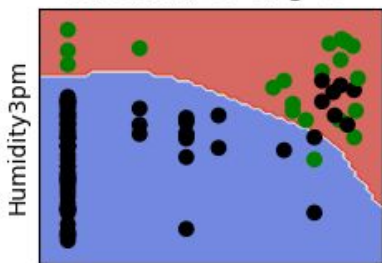
SVC linear, $C=50$, $g=50$



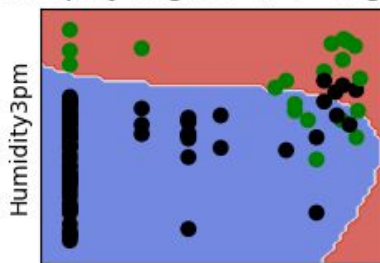
SVC Sigmoid, $C=50$, $g=50$



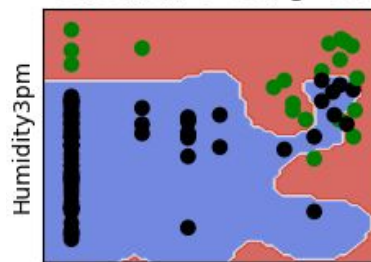
SVC RBF, $C=5$, $g=5$



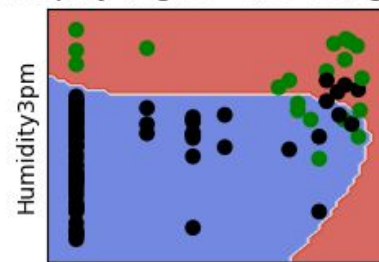
SVC poly degree=3, $C=5$, $g=5$



SVC RBF, $C=50$, $g=50$



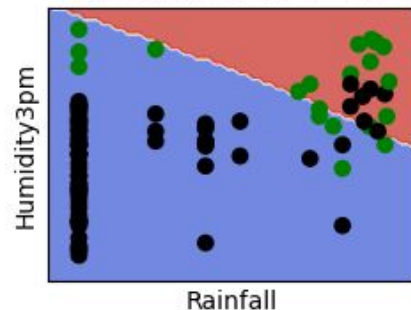
SVC poly degree=3, $C=50$, $g=50$



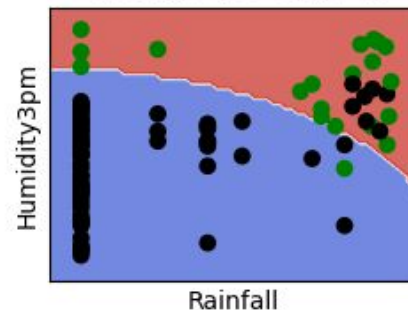


Gamma

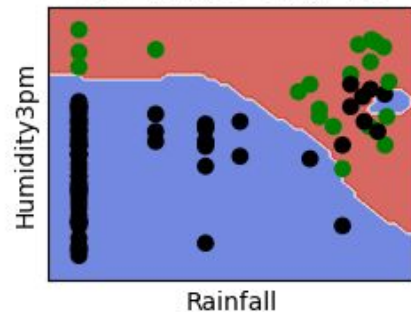
SVC RBF, $C=5$, $g=0.1$



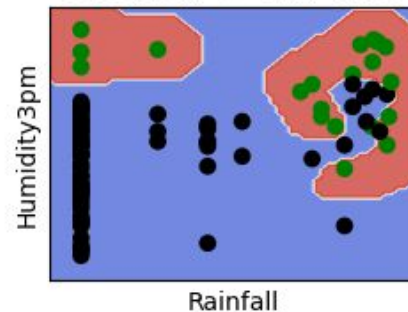
SVC RBF, $C=5$, $g=1$



SVC RBF, $C=5$, $g=10$

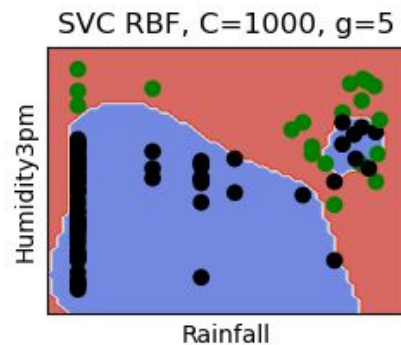
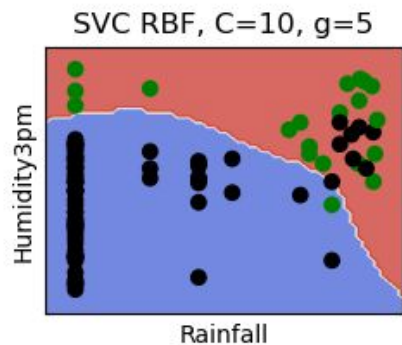
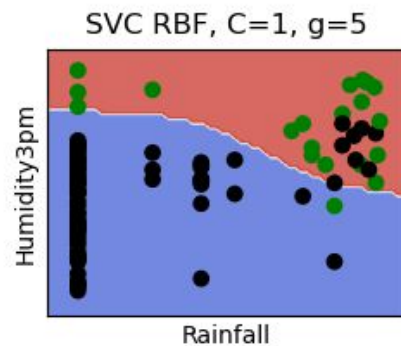
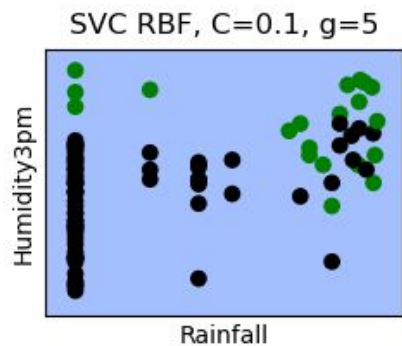


SVC RBF, $C=5$, $g=100$



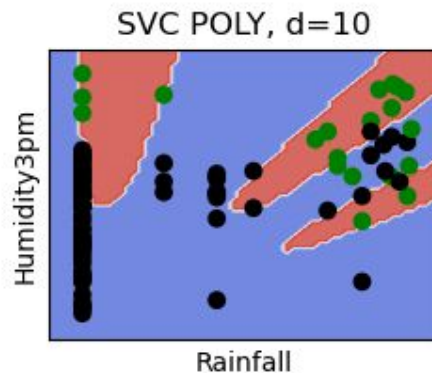
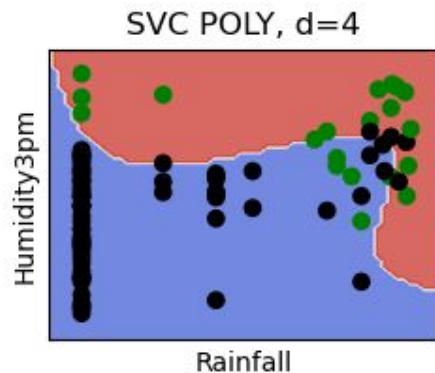
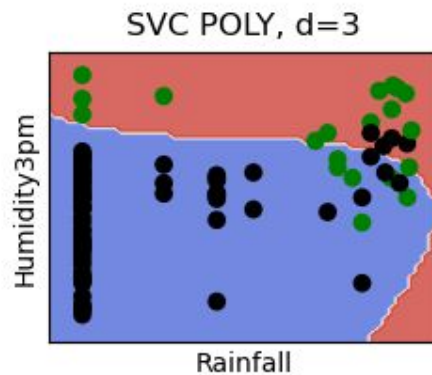
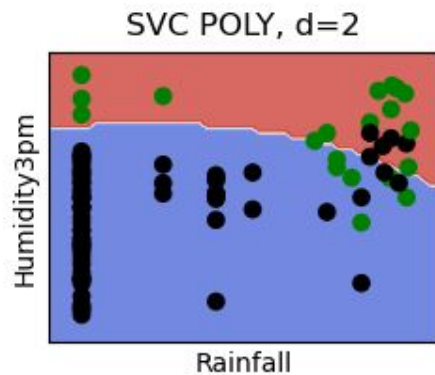


C





Graus del polinomi

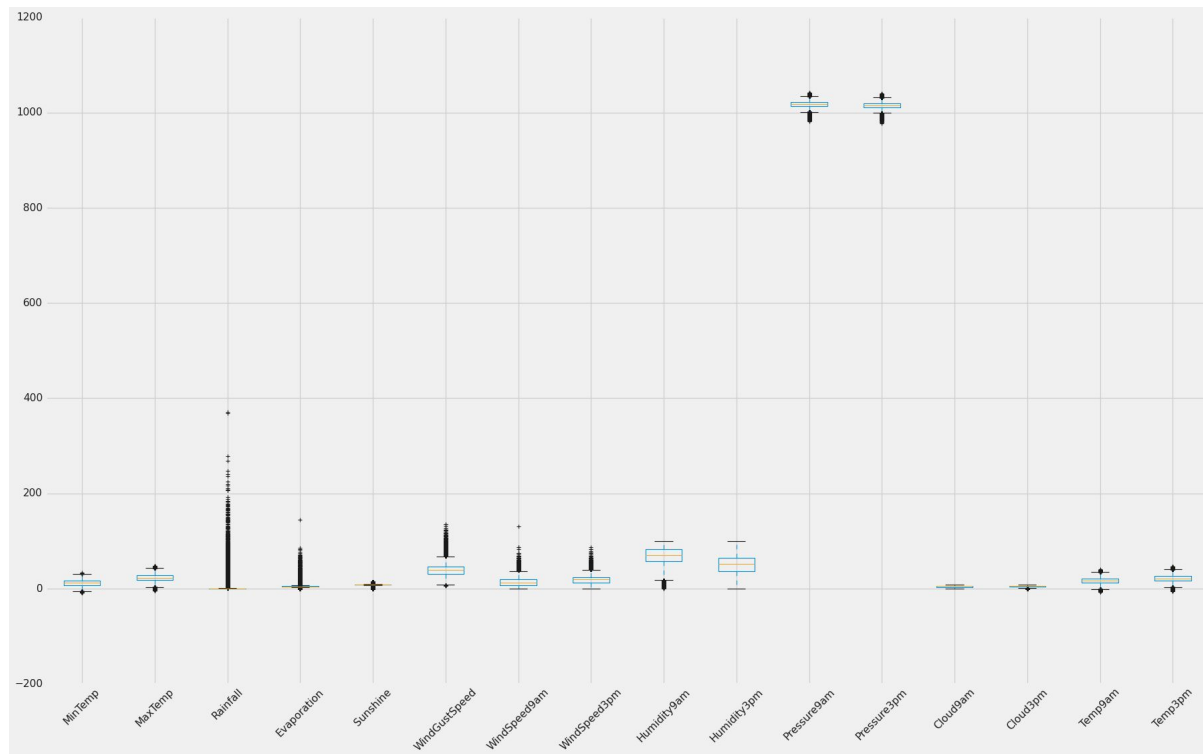




Apatat A



Reescalar dades



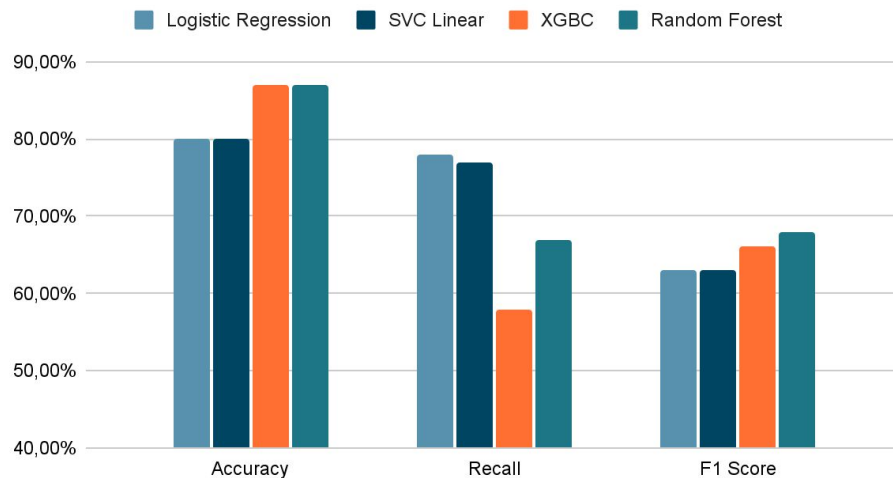


Normalització

Sense modificar



Normalització

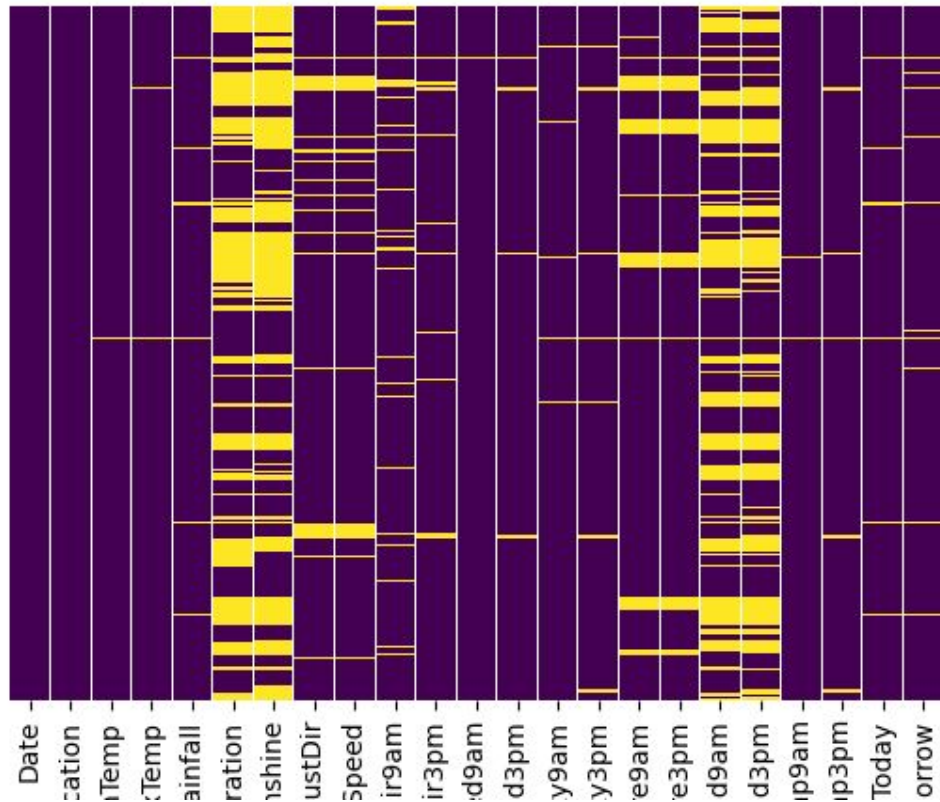




Skew

Sense modifcar

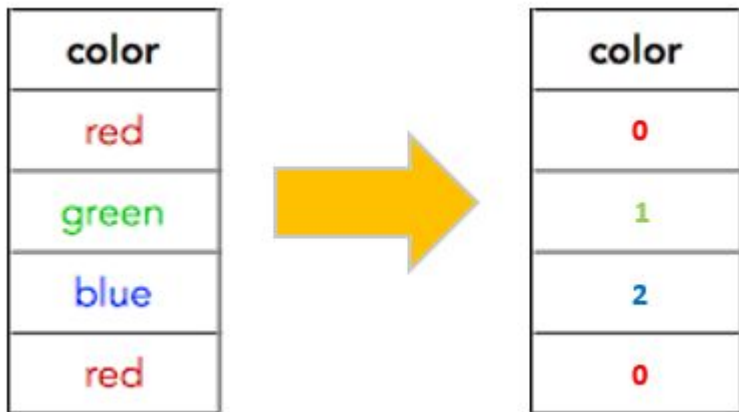




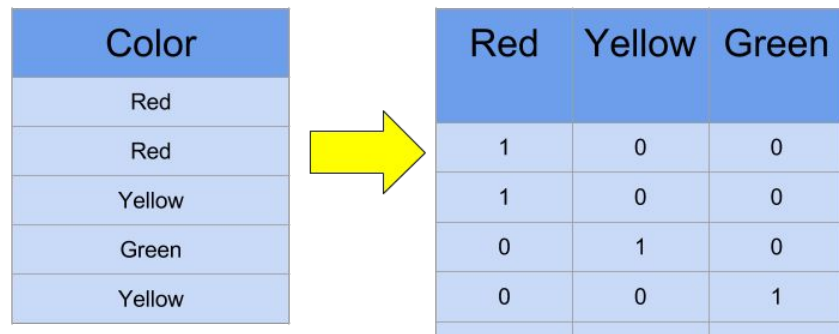


Codificació de dades categòriques

Label Encoder



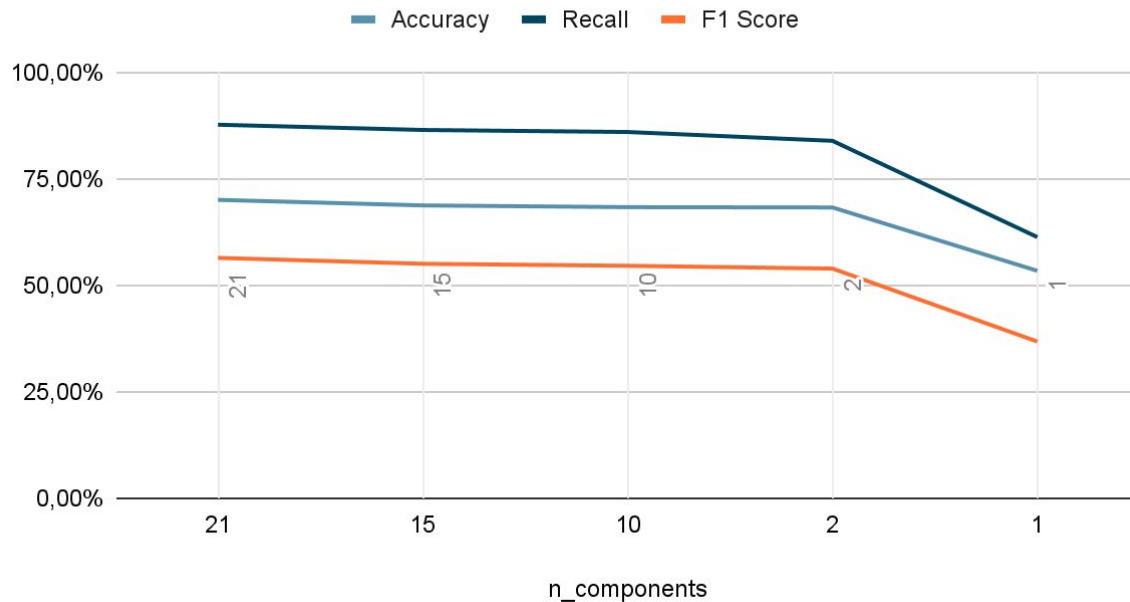
OneHot Encoder





First Component Analysis

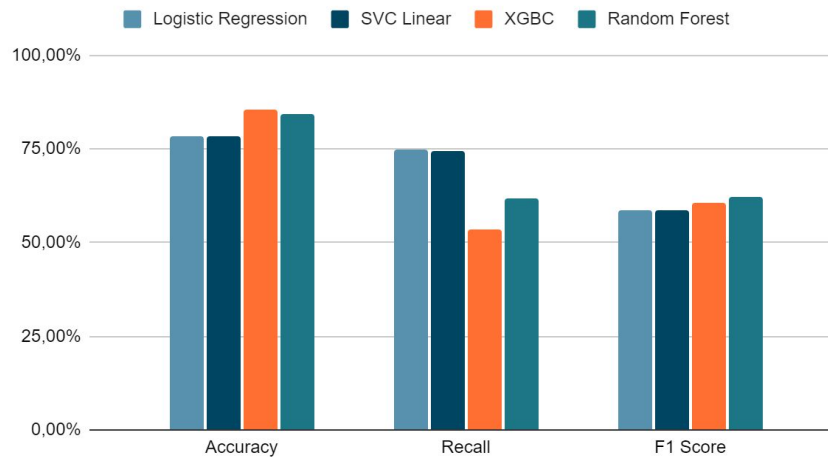
Logistic Regression - Test Size = 50%



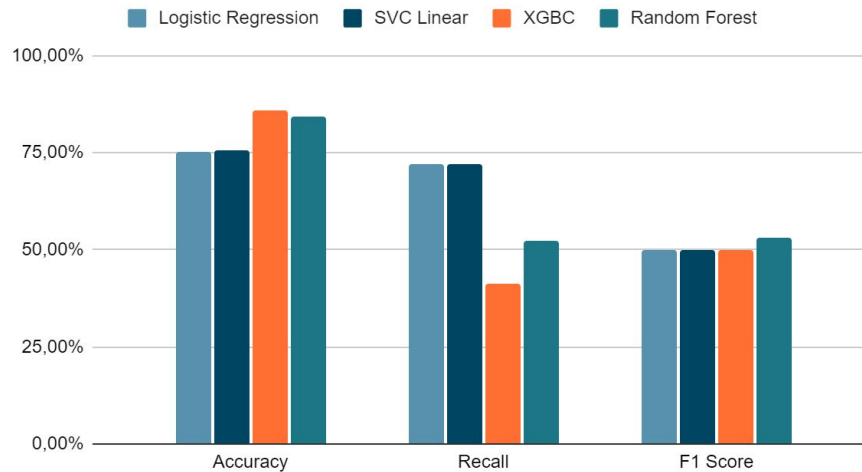


Anomalies

removeOutliers() - Test Size = 50%



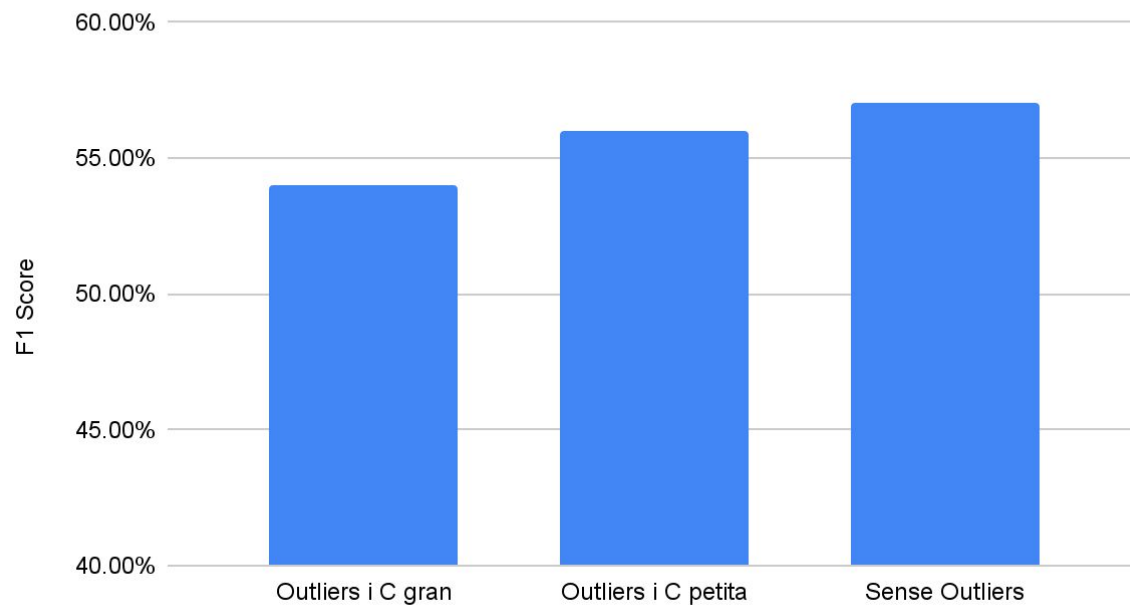
removeOutliers() zScore<2 - Test Size = 50%



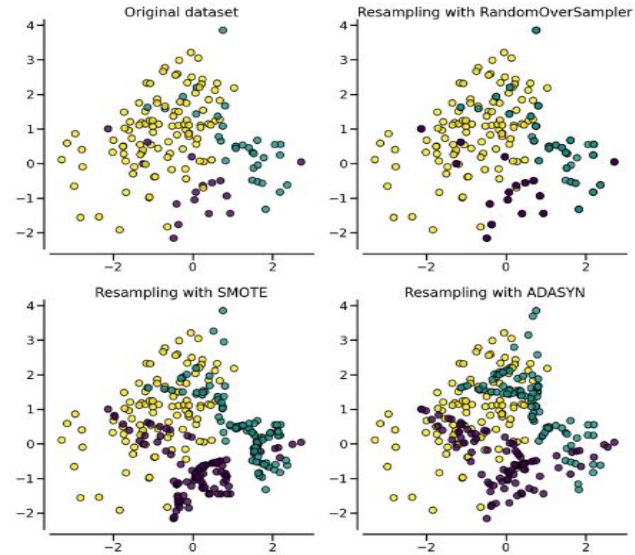
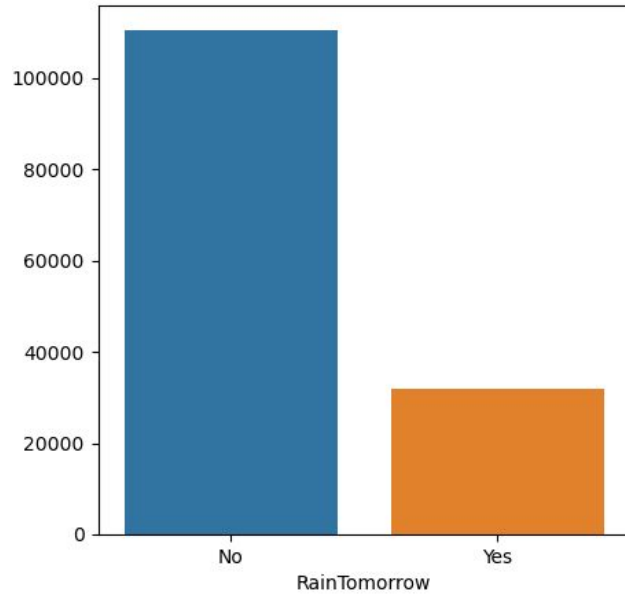


Anomalies (SVM polinomial)

Outliers i SVMs



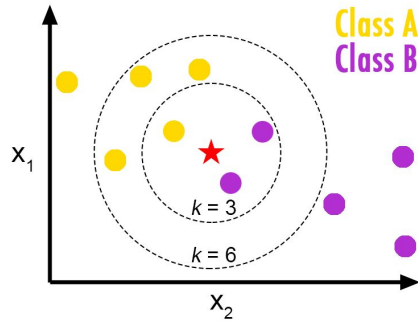
Classes balancejades



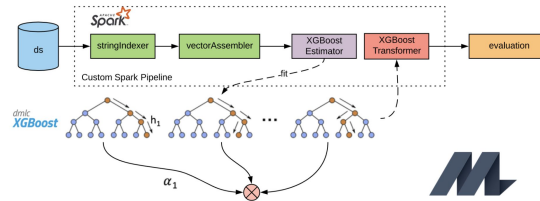


Models aplicats

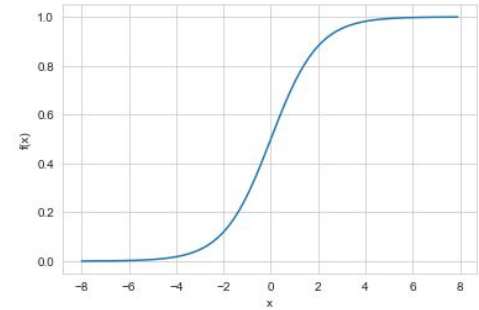
Knn



XGBoost



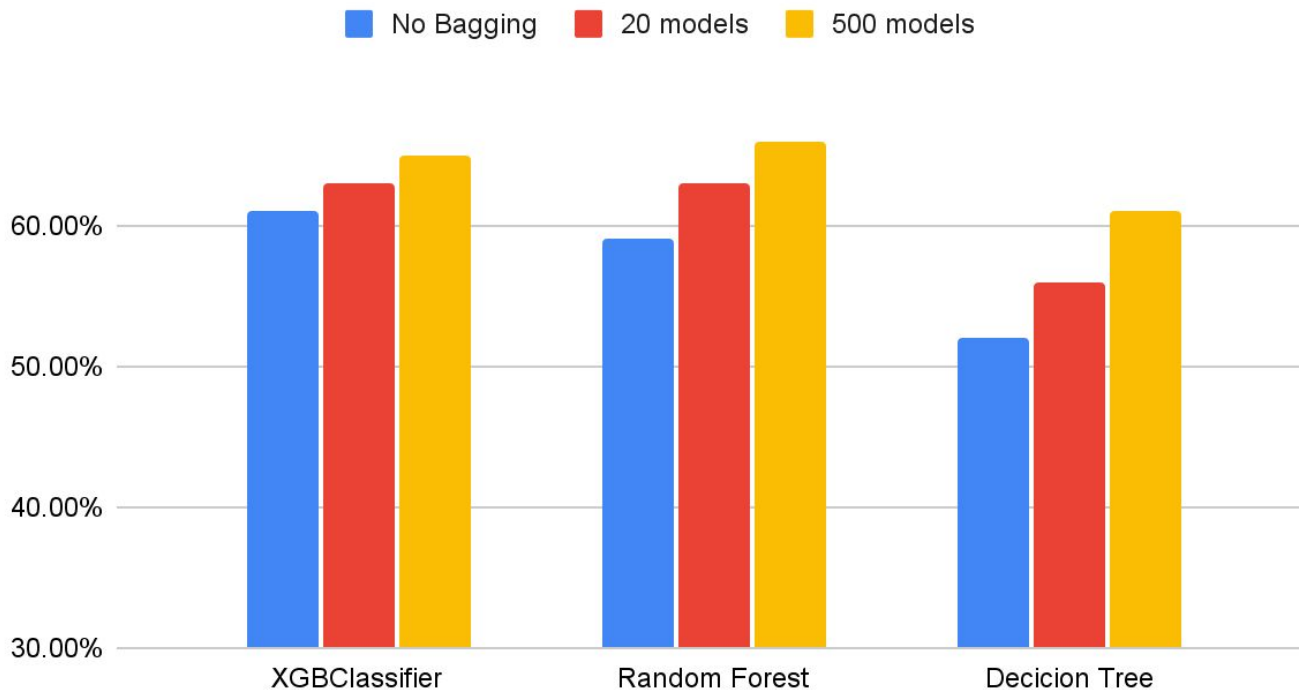
Regresio Logistic





Ensembles

F1 Score - Normal V.S. Bagging





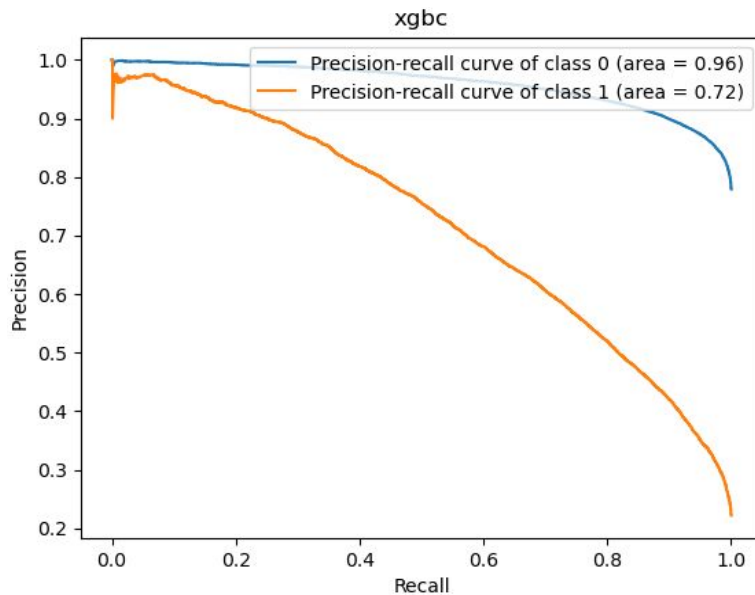
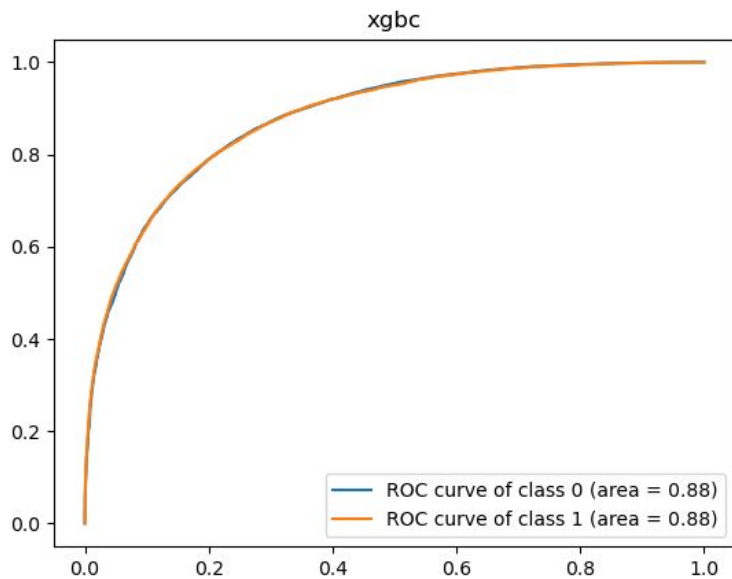
Analisis de mètriques

RAIN?	Actual YES	Actual NO	Total
Predicted YES	TP = 5	FP = 0	5
Predicted NO	FN = 25	TN = 270	295
Total	30	270	300

Accuracy	92%
Precision	100%
Recall	16%
F1 Score	27%



ROC i PR curves





Classification Report

```
XGBC
Accuracy: 0.8103602364911315
Precision: 0.5566428302596914
Recall: 0.6911214953271028
f1 score: 0.616635397123202

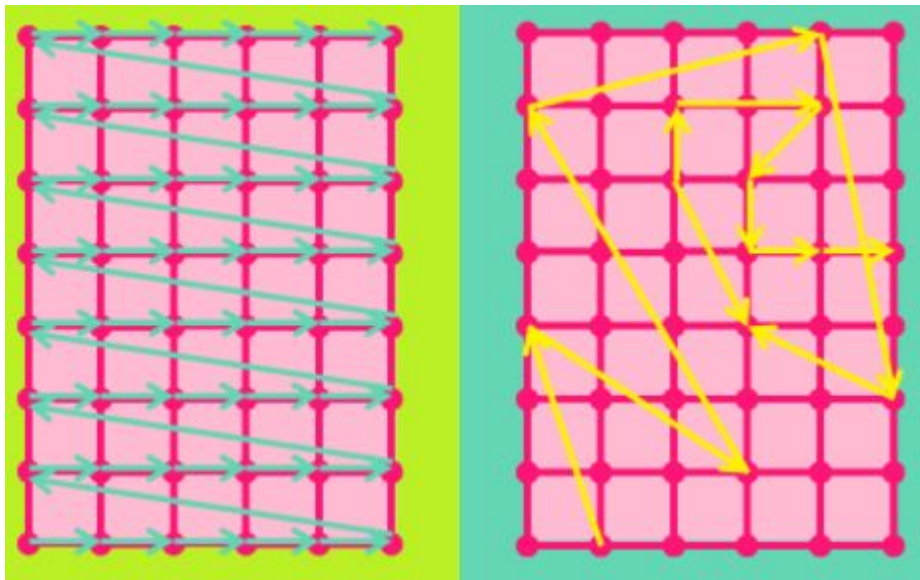
Classification report:

```

	precision	recall	f1-score	support
0	0.91	0.84	0.87	22672
1	0.56	0.69	0.62	6420
accuracy			0.81	29092
macro avg	0.73	0.77	0.75	29092
weighted avg	0.83	0.81	0.82	29092



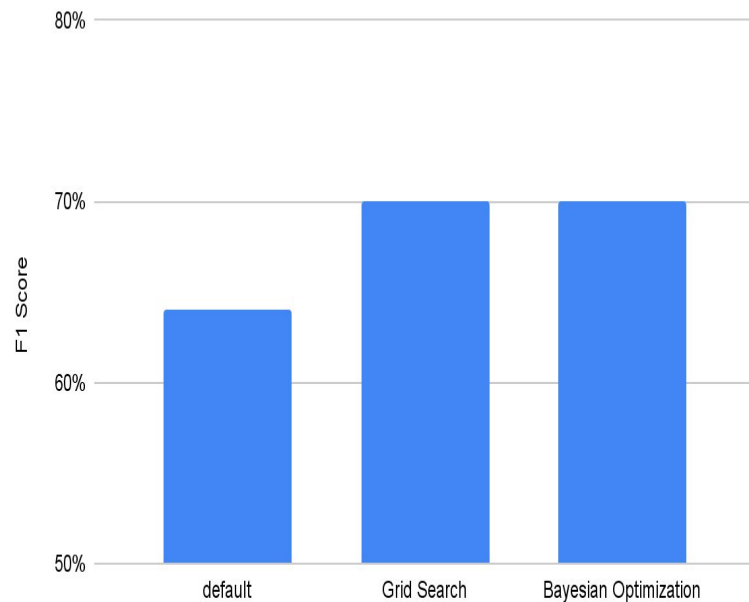
Cerca de l'hiperparàmetre



GRID

BAYESIAN

Random Forest - Hyperparameter Search





Conclusió

