



PRÀCTICA 2:

CLASSIFICACIÓ DE DADES

Implementació de models

classificadors per a dades binàries

Fernández, Mireia (1562636)

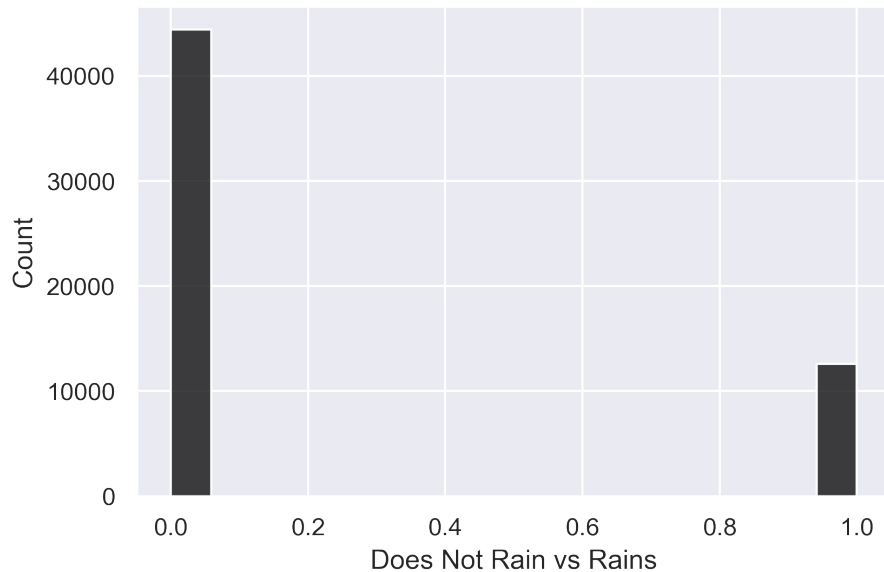
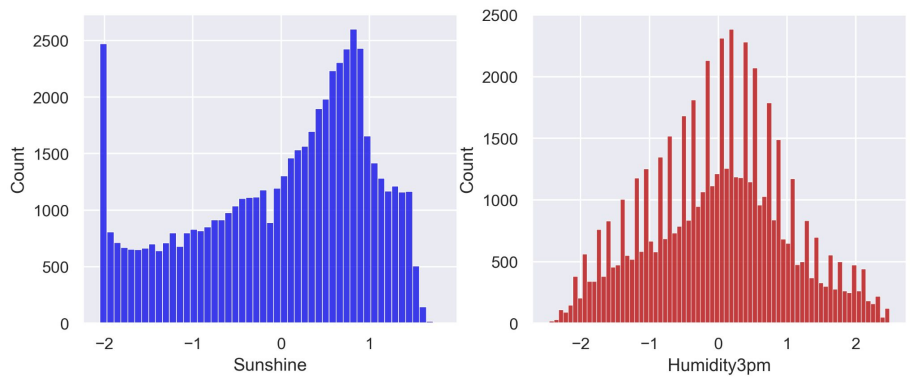
Martín, Marc (1564274)

Vázquez, Marina (1563735)

DATASET:



Rain in Australia



44427 vs 12606



PRECISIÓ DELS CLASSIFICADORS

	Accuracy socre	Temps d'execució (s)
Regressor logístic	0.855308	0.490373
SVM (lineal)	0.829009	69.959064
KNN (7 veïns)	0.844262	193.476523
Random Forest (45 arbres)	0.853774	3.133465



PRECISIÓ DELS DIFERENTS *KERNELS* DEL SVM

	2 atributs		3 atributs	
	Accuracy socre	T (s)	Accuracy socre	T (s)
Lineal	0.829009	69.959064	0.831901	70.501107
Poly	0.828789	1770.248152	-	+7200.0
RBF	0.833041	390.444688	0.832252	506.565012
Sigmoid	0.683659	74.902673	0.683352	146.250606



ENSEMBLE METHODS

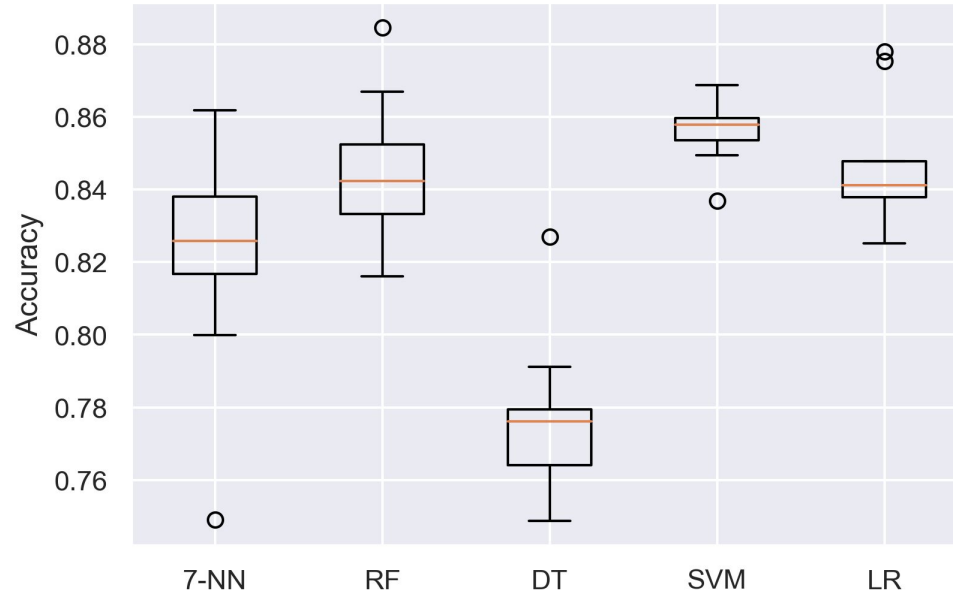
BAGGING

- Random Forest
500 arbres = 0.864338
- BaggingClassifier amb KNN
10 estimadors = 0.843736
- BaggingClassifier amb SVM lineal
10 estimadors = 0.858902

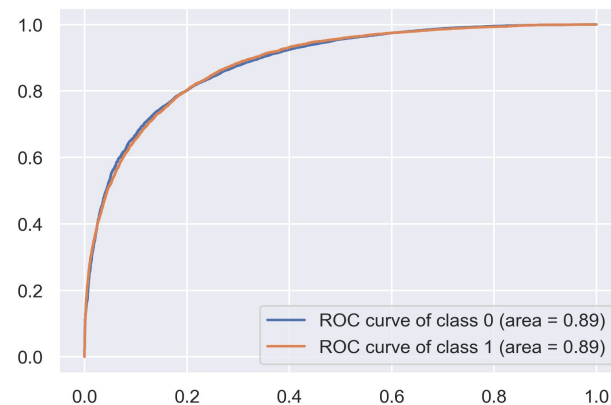
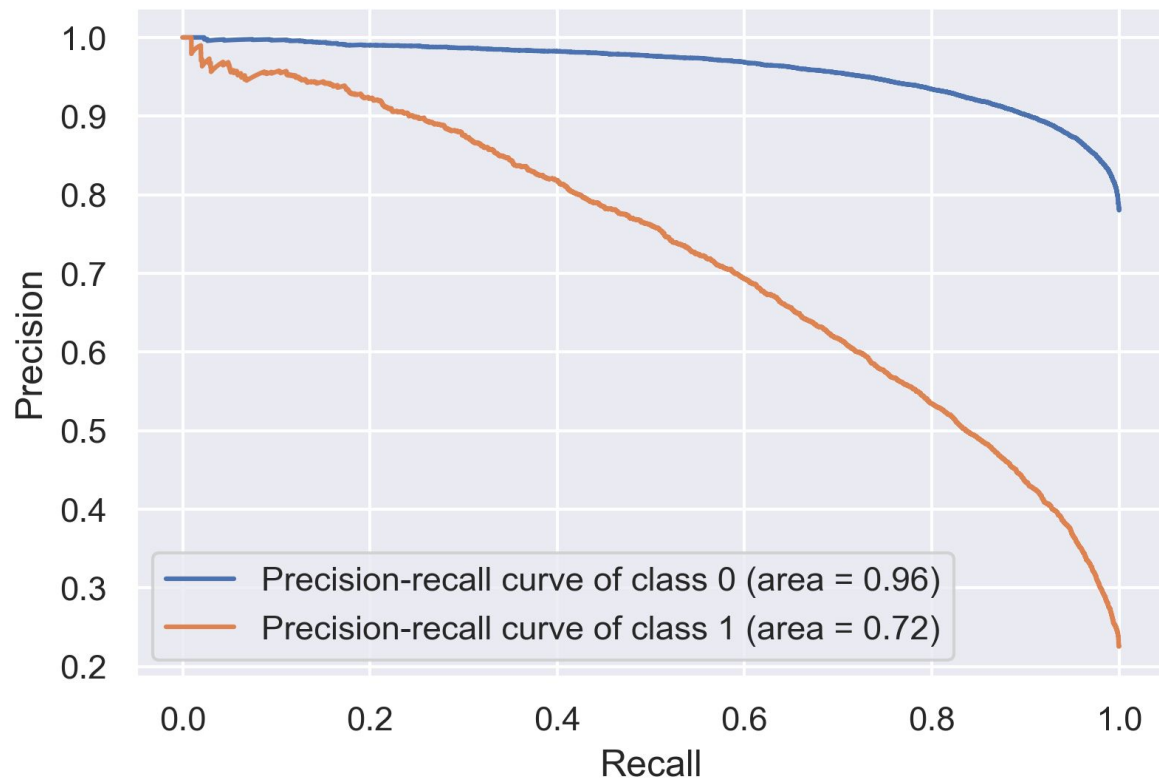
BOOSTING

- Adaptive Boosting (AdaBoost)
100 estimadors = 0.846914
500 estimadors = 0.847071

VALIDACIÓ CREUADA: k-fold amb k=10



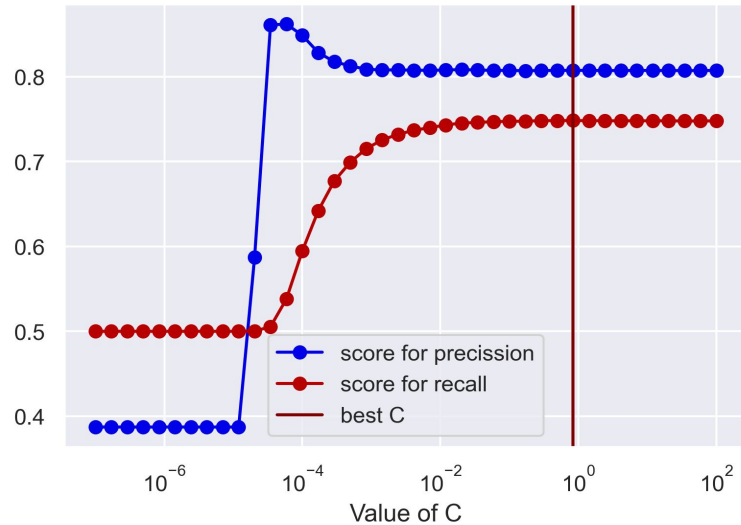
F1 SCORE



	precision	recall	f1-score	support
Yes	0.88	0.95	0.91	17711
No	0.74	0.54	0.62	5103

CERCA D'HIPERPARÀMETRES

LogisticRegression



SVM

