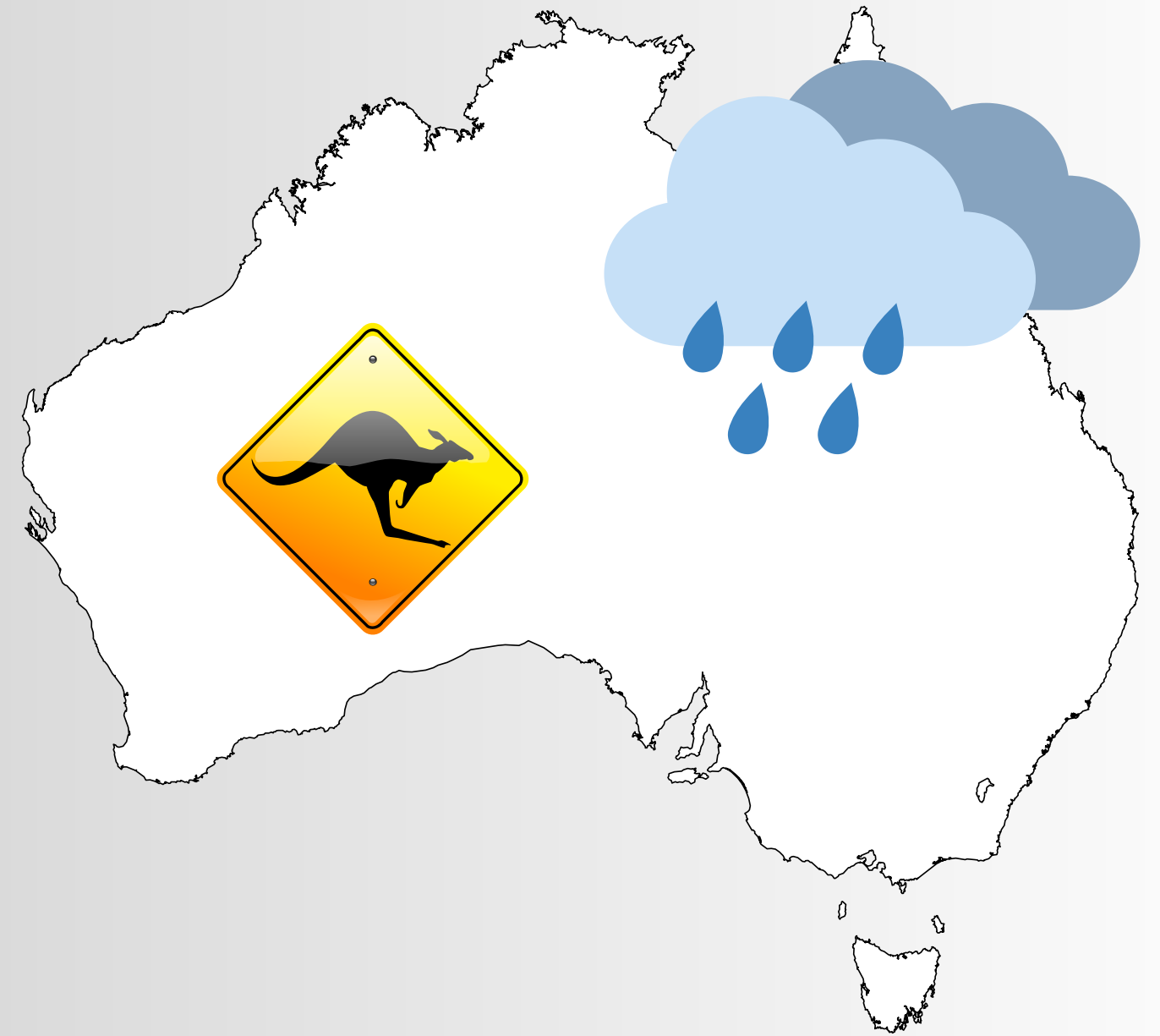


# Predicció de precipitacions a Austràlia

Aprenentatge Computacional



Youssef Cahouach Guella Ikhlef 1638618

**UAB**  
Universitat Autònoma  
de Barcelona

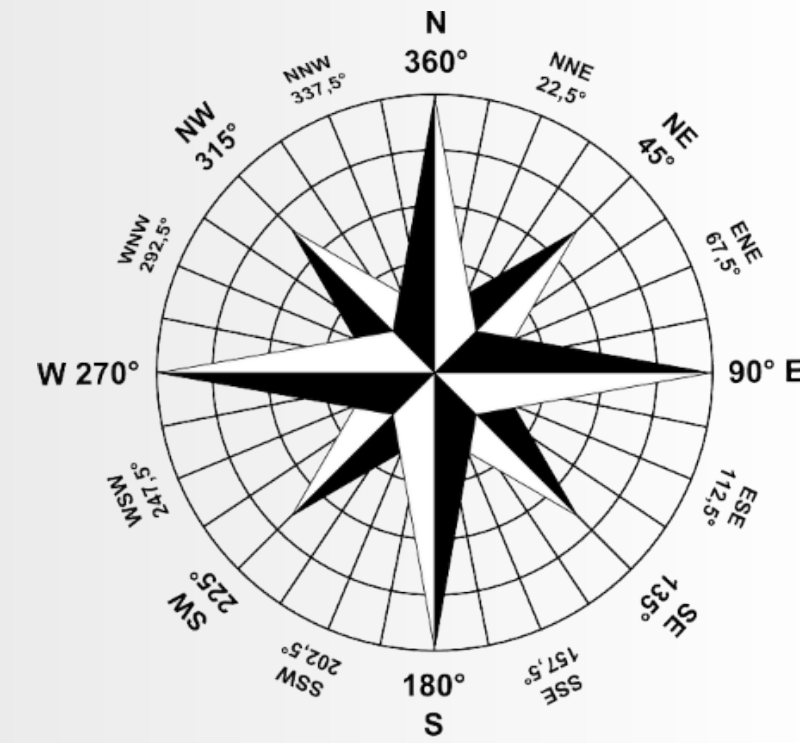
# Introducció

Aquest ‘dataset’ anomenat “Rain in Australia” proporciona un conjunt de dades exhaustives que recull observacions meteorològiques diàries de diverses localitzacions d’Austràlia, que es troba en un clima particularment variable.

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	WNW	20.0	24.0	71.0	22.0	1007.7	1007.1	8.0	NaN	16.9	21.8	No	No
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	WSW	4.0	22.0	44.0	25.0	1010.6	1007.8	NaN	NaN	17.2	24.3	No	No
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	46.0	W	WSW	19.0	26.0	38.0	30.0	1007.6	1008.7	NaN	2.0	21.0	23.2	No	No
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	24.0	SE	E	11.0	9.0	45.0	16.0	1017.6	1012.8	NaN	NaN	18.1	26.5	No	No
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	NW	7.0	20.0	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	No	No
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
145455	2017-06-21	Uluru	2.8	23.4	0.0	NaN	NaN	E	31.0	SE	ENE	13.0	11.0	51.0	24.0	1024.6	1020.3	NaN	NaN	10.1	22.4	No	No
145456	2017-06-22	Uluru	3.6	25.3	0.0	NaN	NaN	NNW	22.0	SE	N	13.0	9.0	56.0	21.0	1023.5	1019.1	NaN	NaN	10.9	24.5	No	No
145457	2017-06-23	Uluru	5.4	26.9	0.0	NaN	NaN	N	37.0	SE	WNW	9.0	9.0	53.0	24.0	1021.0	1016.8	NaN	NaN	12.5	26.1	No	No
145458	2017-06-24	Uluru	7.8	27.0	0.0	NaN	NaN	SE	28.0	SSE	N	13.0	7.0	51.0	24.0	1019.4	1016.5	3.0	2.0	15.1	26.0	No	No
145459	2017-06-25	Uluru	14.9	NaN	0.0	NaN	NaN	NaN	NaN	ESE	ESE	17.0	17.0	62.0	36.0	1020.2	1017.9	8.0	8.0	15.0	20.9	No	NaN

145460 rows × 23 columns

# Tractament de dades



## Tractament de NaNs:

- PER A LES COLUMNES NUMÈRIQUES, OMPLIM ELS VALORS NAN UTILITZANT PRIMERAMENT EL MÈTODE 'BACKWARD FILL', SEGUIT DEL 'FORWARD FILL' I FINALMENT LA MITJANA.
- PER A LES COLUMNES NO NUMÈRIQUES, OMPLIM ELS VALORS NAN AMB EL VALOR MÉS FREQUENT, ÉS A DIR, EL MODE.

## Noves dades:

- GENEREM NOVES COLUMNES COM 'RAINYESTERDAY' I 'RAINLASTWEEK', QUE INDIQUEN SI VA PLOURE EL DIA ANTERIOR O LA SETMANA PASSADA, RESPECTIVAMENT.

## Encoding:

- PER A LES COLUMNES QUE INDIQUEN LA DIRECCIÓ DEL VENT, UTILITZEM UN CODI PERSONALITZAT BASAT EN LES DIRECCIONS CARDINALS MESURADES EN GRAUS.
- PEL QUE FA A LA DATA, QUE ÉS ÚNICA PER A CADA FILA DEL CONJUNT DE DADES, UTILITZEM UNA TÈCNICA BASADA EN LA FUNCIO SINUSOIDAL PER CONVERTIR CADA DATA EN UN VALOR NUMÈRIC.

# Dataset tractat

	MinTemp	MaxTemp	Rainfall	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow	DateSin	RainedYesterday	RainedLastWeek
45593	6.1	18.2	0.2	0.707107	43.0	0.707107	0.923880	19.0	26.0	63.0	47.0	1024.6	1022.2	4.0	6.0	12.4	17.3	0.0	0.0	-0.803460	0.0	1.0
45594	8.3	17.0	0.0	0.999998	41.0	0.707107	0.999998	11.0	24.0	65.0	57.0	1026.2	1024.2	6.0	7.0	12.1	15.5	0.0	0.0	-0.793100	0.0	1.0
45595	8.8	19.5	0.0	-0.001745	48.0	0.999998	0.923880	19.0	17.0	70.0	48.0	1026.1	1022.7	7.0	7.0	14.1	18.9	0.0	1.0	-0.782505	0.0	1.0
45596	8.4	22.8	16.2	0.999998	31.0	-0.001745	0.923880	7.0	6.0	82.0	32.0	1024.1	1020.7	7.0	1.0	13.3	21.7	1.0	0.0	-0.771679	0.0	1.0
45597	9.1	25.2	0.0	0.001745	30.0	0.707107	-0.707107	6.0	9.0	74.0	34.0	1024.4	1021.1	1.0	2.0	14.6	24.0	0.0	0.0	-0.760624	1.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
145454	3.5	21.8	0.0	0.999998	31.0	0.923880	0.999998	15.0	13.0	59.0	27.0	1024.7	1021.2	3.0	2.0	9.4	20.9	0.0	0.0	0.198569	0.0	0.0
145455	2.8	23.4	0.0	0.999998	31.0	0.707107	0.923880	13.0	11.0	51.0	24.0	1024.6	1020.3	3.0	2.0	10.1	22.4	0.0	0.0	0.181680	0.0	0.0
145456	3.6	25.3	0.0	-0.382683	22.0	0.707107	0.001745	13.0	9.0	56.0	21.0	1023.5	1019.1	3.0	2.0	10.9	24.5	0.0	0.0	0.164737	0.0	0.0
145457	5.4	26.9	0.0	0.001745	37.0	0.707107	-0.923880	9.0	9.0	53.0	24.0	1021.0	1016.8	3.0	2.0	12.5	26.1	0.0	0.0	0.147746	0.0	0.0
145458	7.8	27.0	0.0	0.707107	28.0	0.382683	0.001745	13.0	7.0	51.0	24.0	1019.4	1016.5	3.0	2.0	15.1	26.0	0.0	0.0	0.130711	0.0	0.0

102182 rows × 22 columns

# Classificadors: millors paràmetres

StratifiedKFold = 10

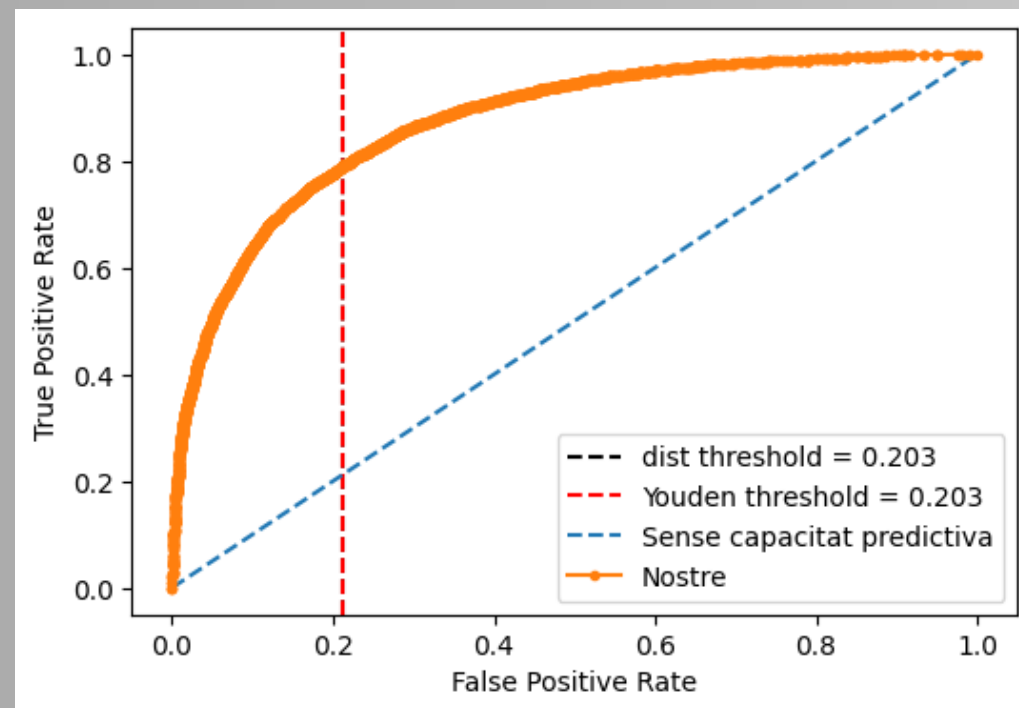
Regressió Logística	KNN	XGBoost
C: 0.03359818286283781 penalty: l2 solver: lbfgs	n_neighbors: 8 p: 1 weights: distance	learning_rate: 0.2 max_depth: 10 min_child_weight: 1 n_estimators: 500

# Classificadors: resultats

Train – CV = 10		
Model	F1-Score	
	Per Defecte	Millor paràmetres
Regressió Logística	0.74966722156244	0.749914586215
KNN	0.73438586866161	0.744864761043
XGBoost	0.76811363952401	0.775600887485

# Resultats finals

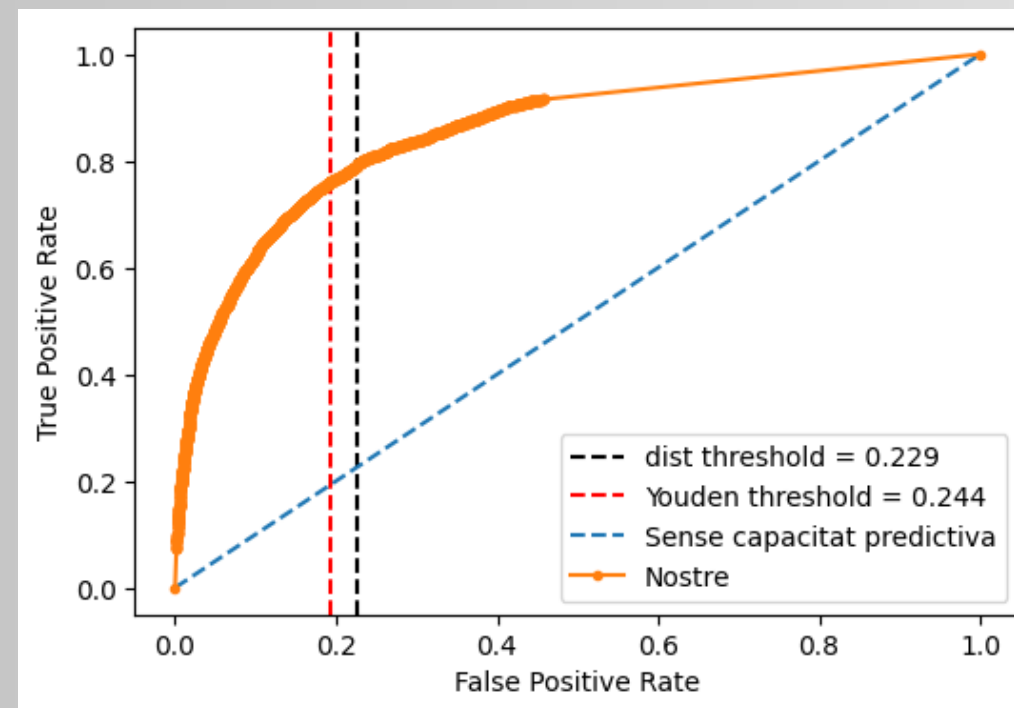
## Regressió Logística



Accuracy: 0.8535  
Precision: 0.8015  
Recall: 0.7280  
F1-Score: 0.7543

ROC AUC = 0.874  
Best Distance thr: 0.203  
Best Youden thr: 0.203

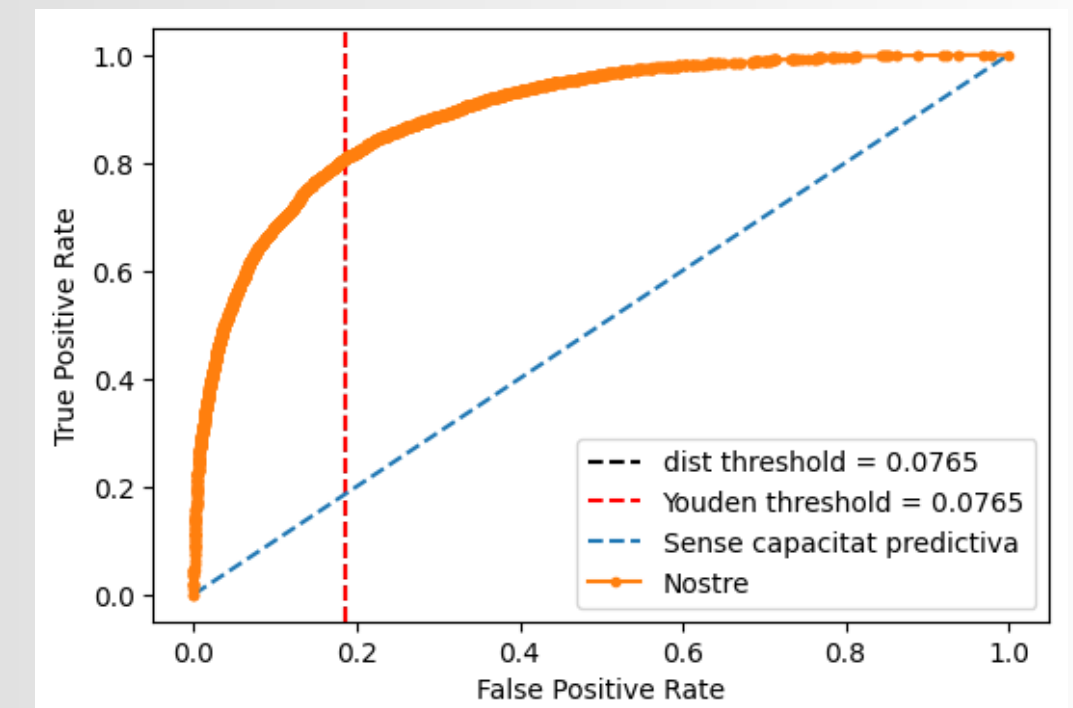
## KNN



Accuracy: 0.8497  
Precision: 0.7985  
Recall: 0.7155  
F1-Score: 0.7435

ROC AUC = 0.854  
Best Distance thr: 0.229  
Best Youden thr: 0.244

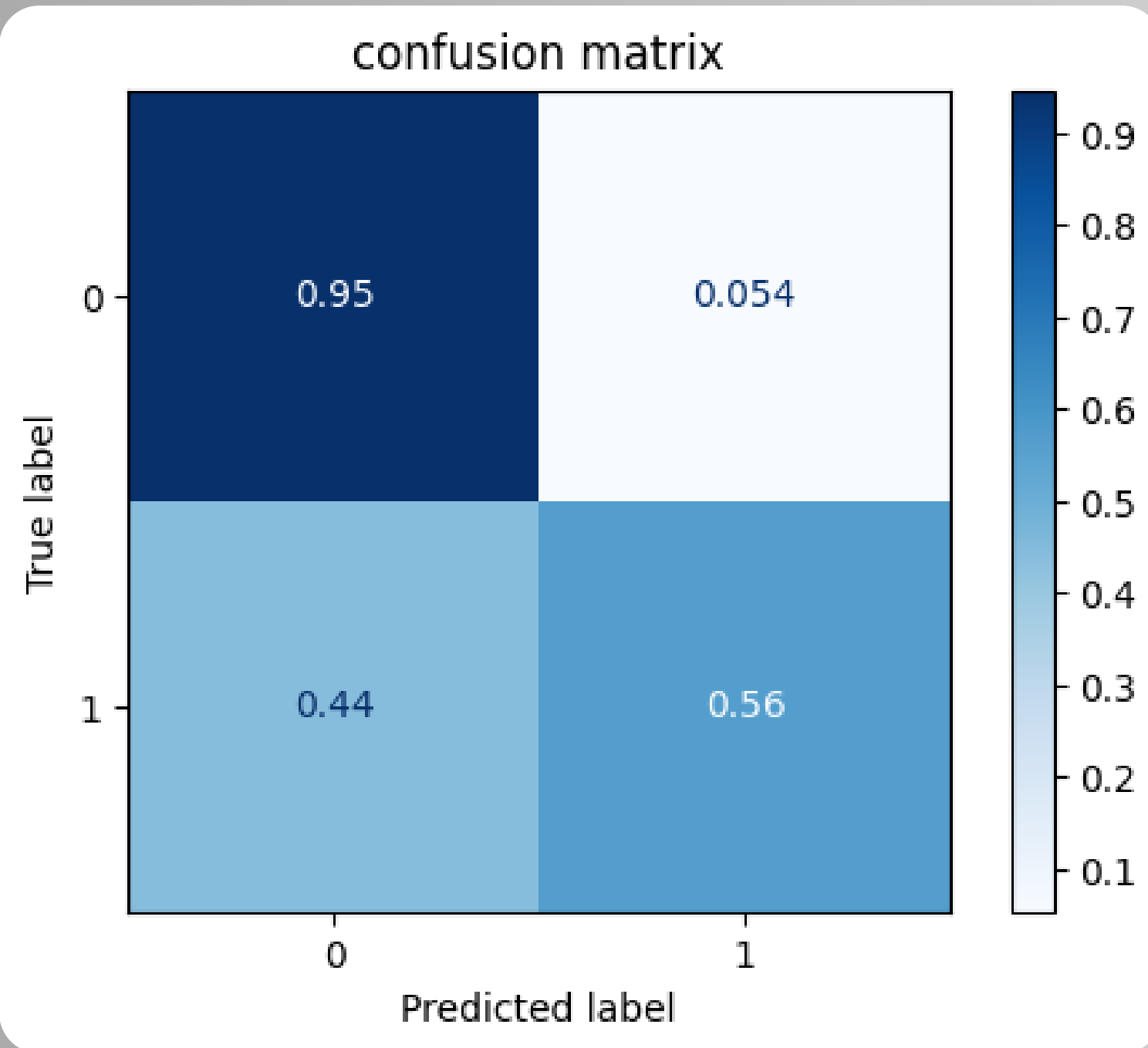
## XGBoost



Accuracy: 0.8637  
Precision: 0.8144  
Recall: 0.7538  
F1-Score: 0.7774

ROC AUC = 0.894  
Best Distance thr: 0.0769  
Best Youden thr: 0.0769

# Conclusions



XGBoost

Accuracy: 0.8637764838283506

Precision: 0.8144731965984963

Recall: 0.7538899855632417

F1-score: 0.7774820289404807

ROC AUC= 0.894





**Gràcies**