

PLOURÀ DEMÀ?

PREDICCIÓ DE PLUJA A AUSTRÀLIA

NIL FARRÉS SOLER - APRENENTATGE COMPUTACIONAL - UAB



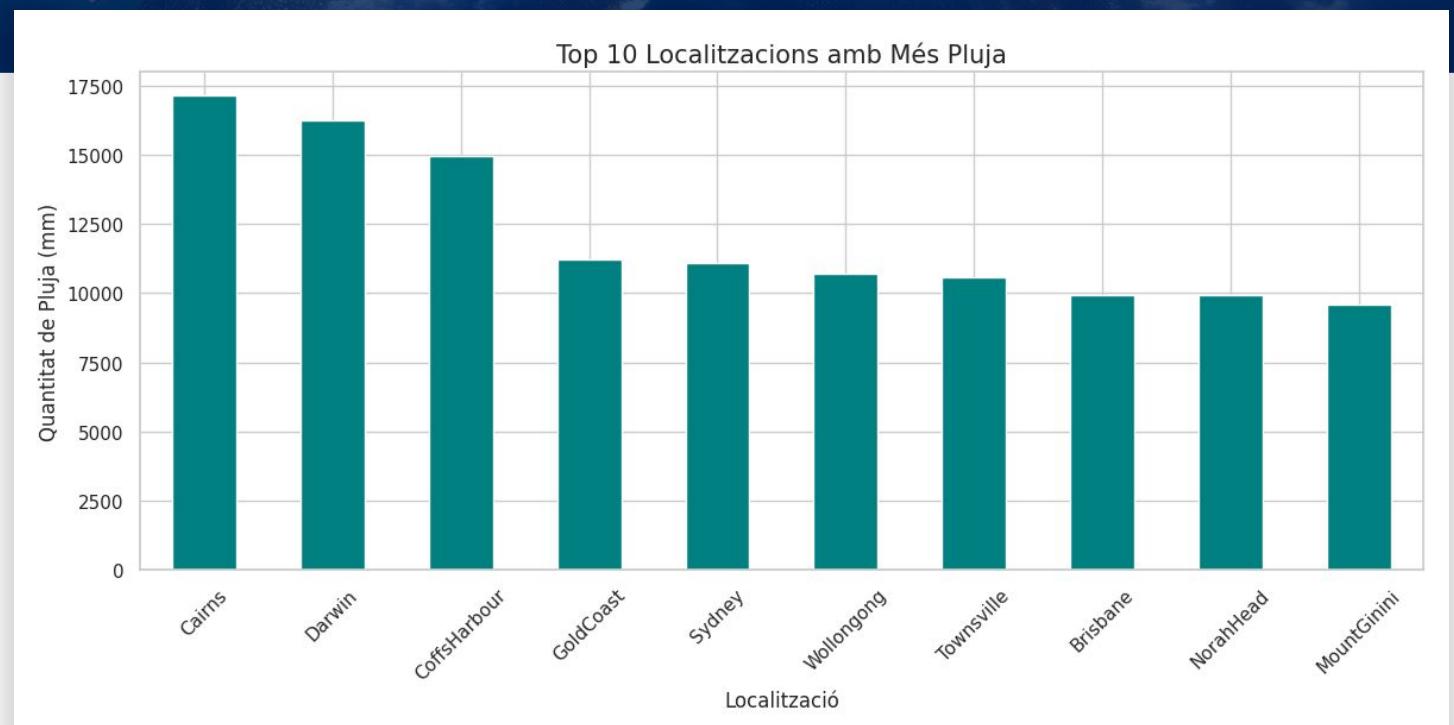
Índex

- **Introducció**
- **Metodologia:**
 - **Neteja i preprocessament de dades.**
 - **Millor mètrica i millor model.**
- **Resultats**
- **Conclusions**



Introducció

Dataset amb 10 anys d'observacions meteorològiques a Austràlia (2007 - 2017)



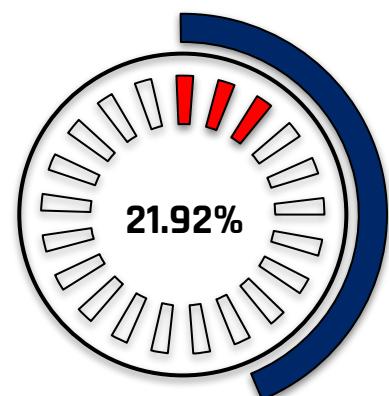
Curiosament, totes són poblacions properes a la costa australiana.

Introducció

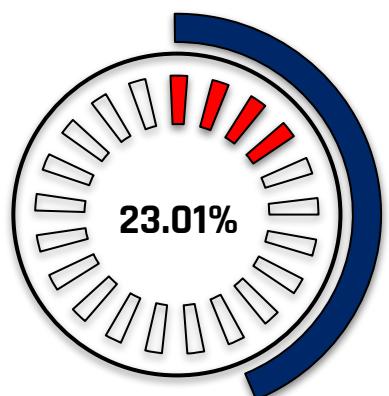
Dataset amb 10 anys d'observacions meteorològiques a Austràlia (2007 - 2017)

Estacions a Austràlia:

- **Estiu:** de desembre a febrer.
- **Tardor:** de març a maig.
- **Hivern:** de juny a agost.
- **Primavera:** de setembre a novembre.

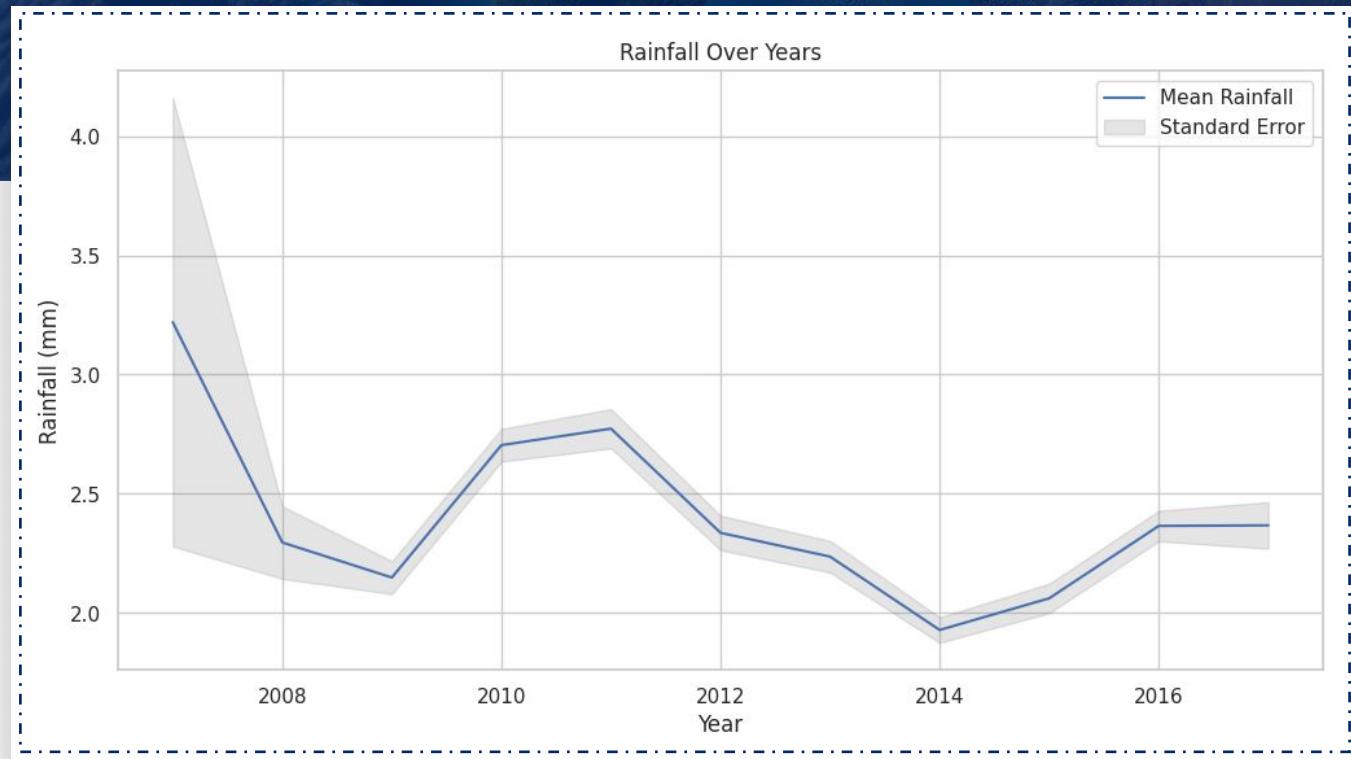


Dies plujosos

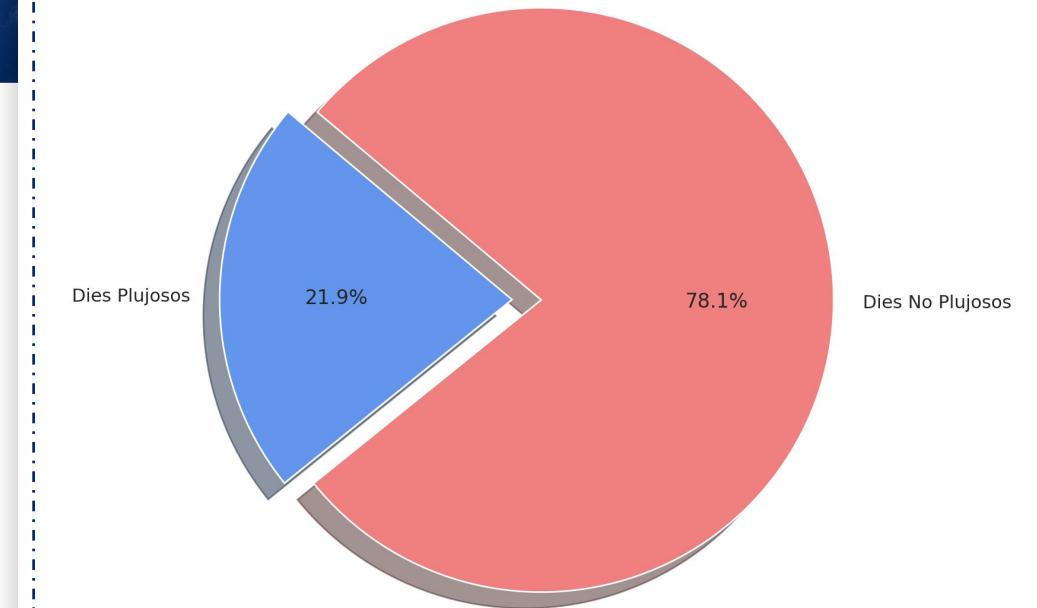


Disminució de pluja
(2007-2017)

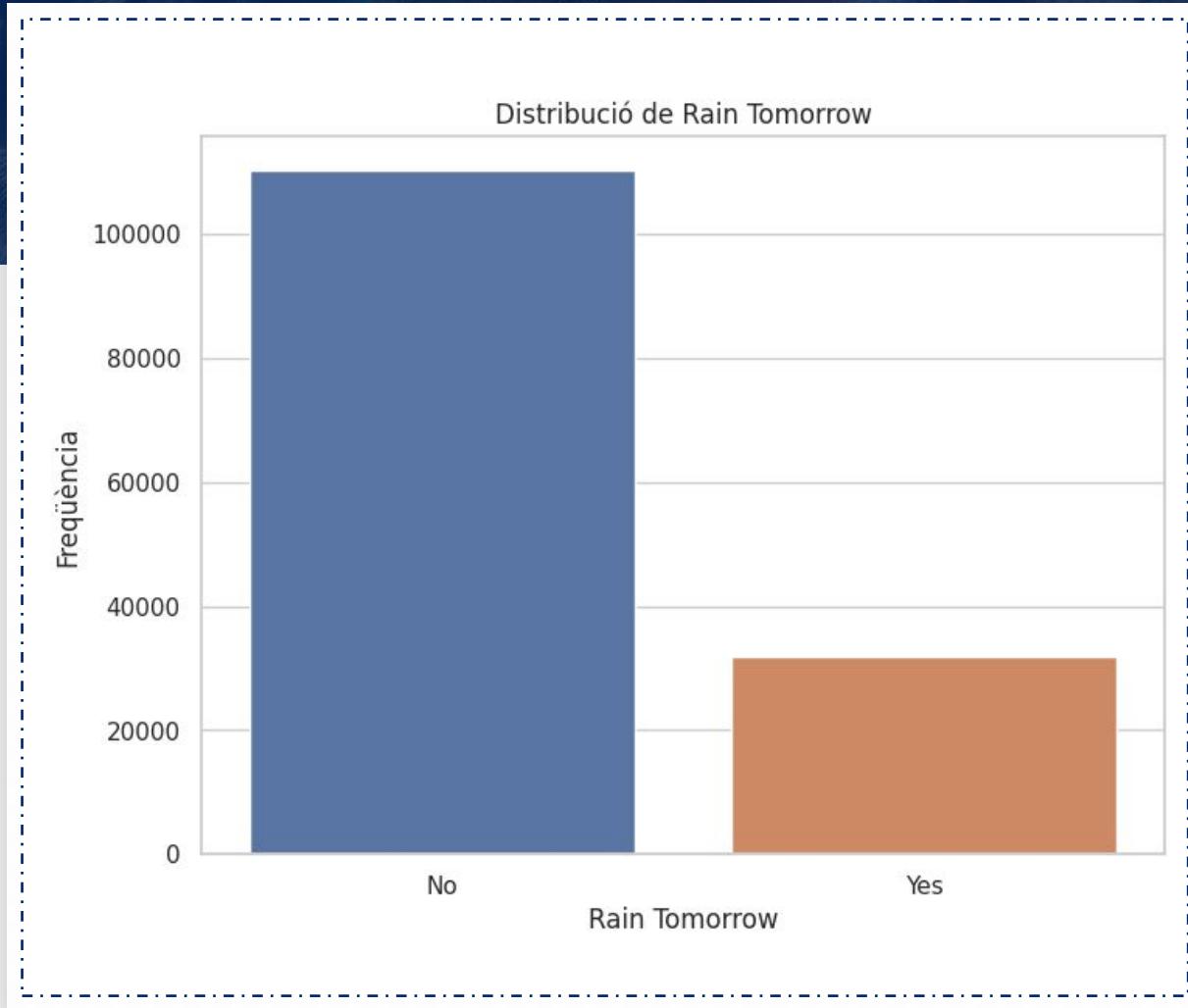
Introducció



Percentatge de Dies Plujosos vs Dies No Plujosos a Austràlia



Introducció

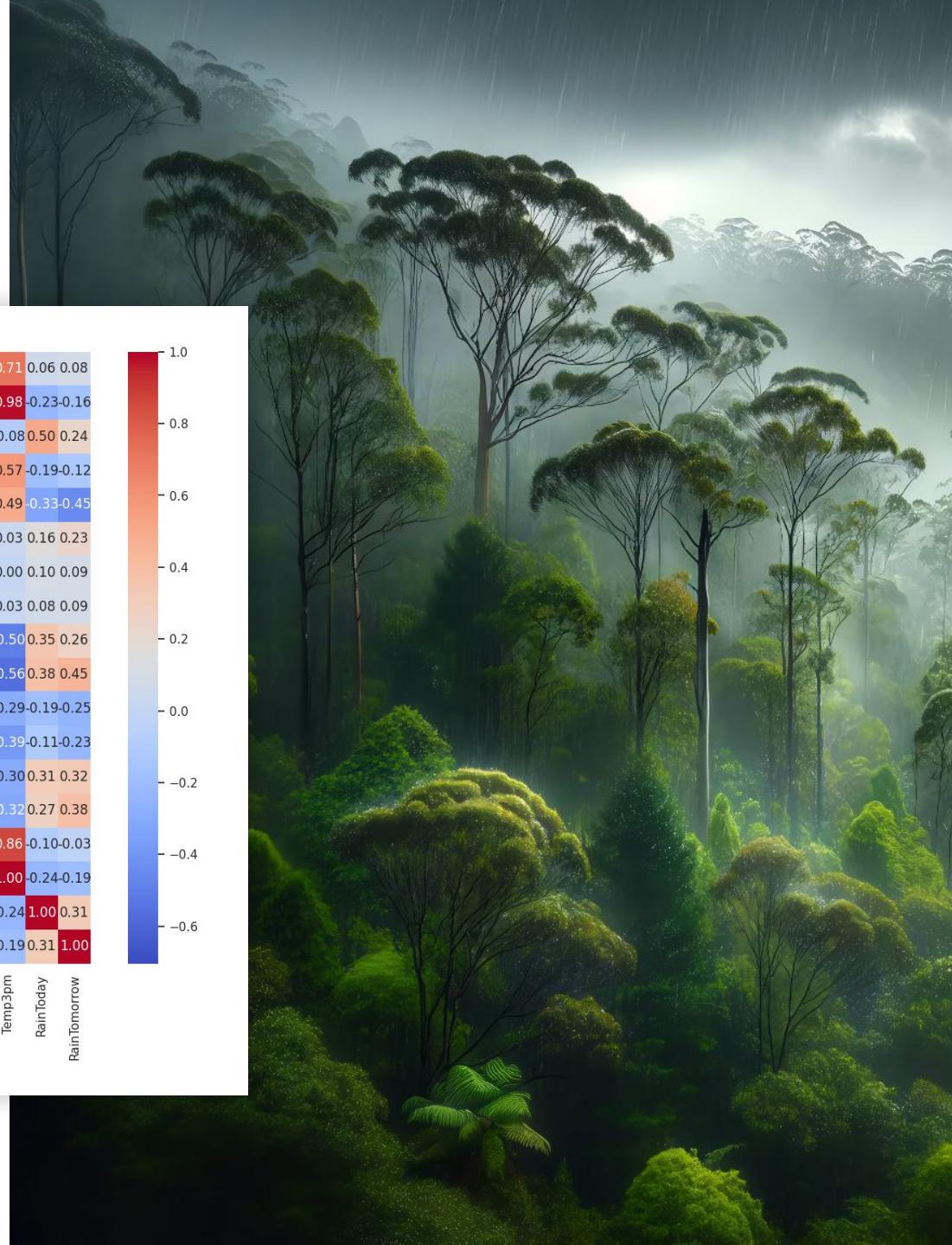
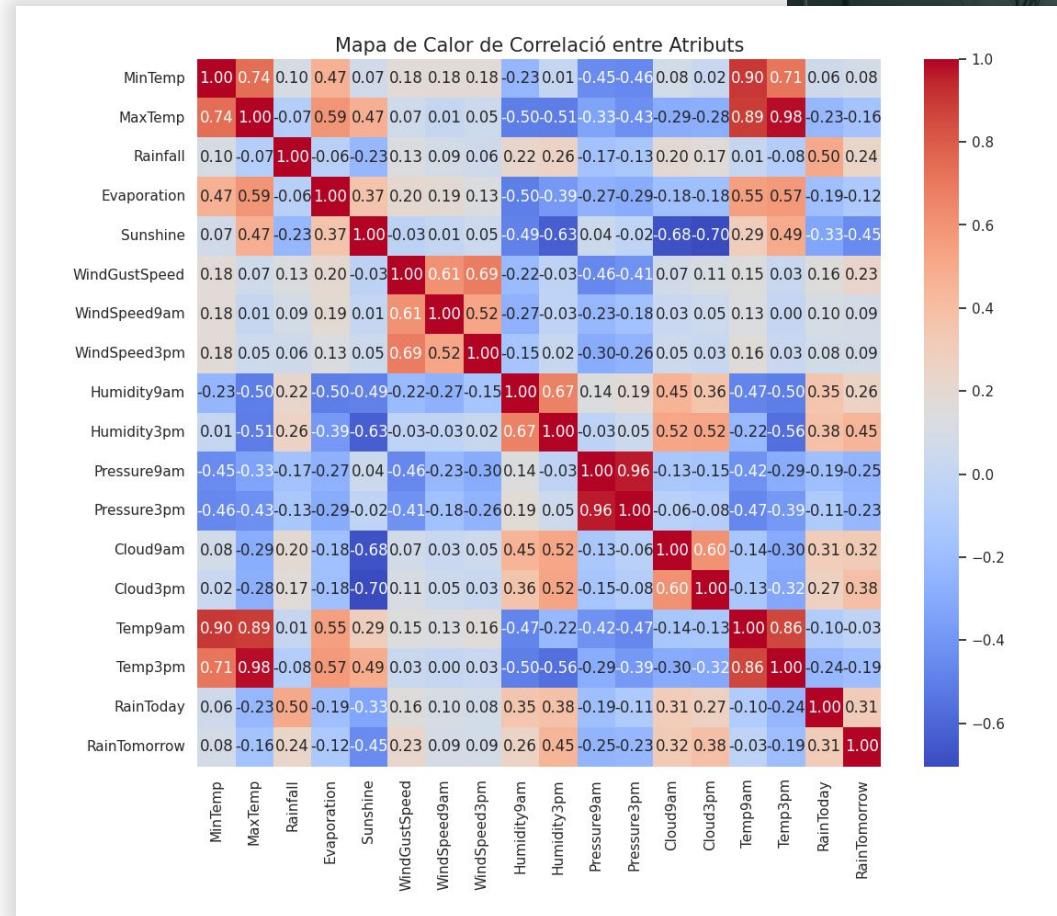
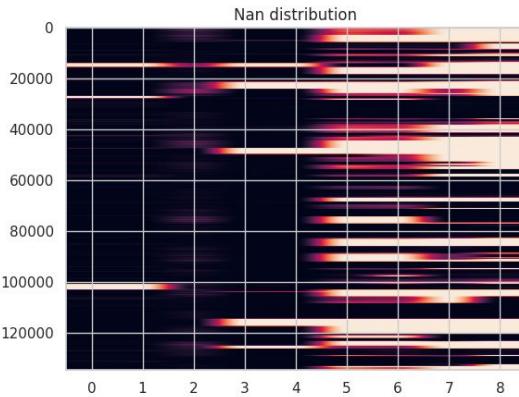
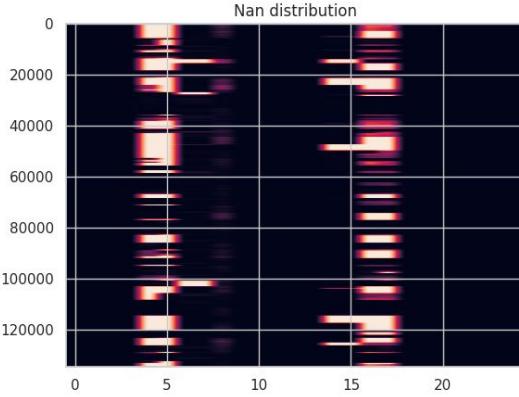


Objectiu:

Entrenar un model que predigui si plourà o no demà.

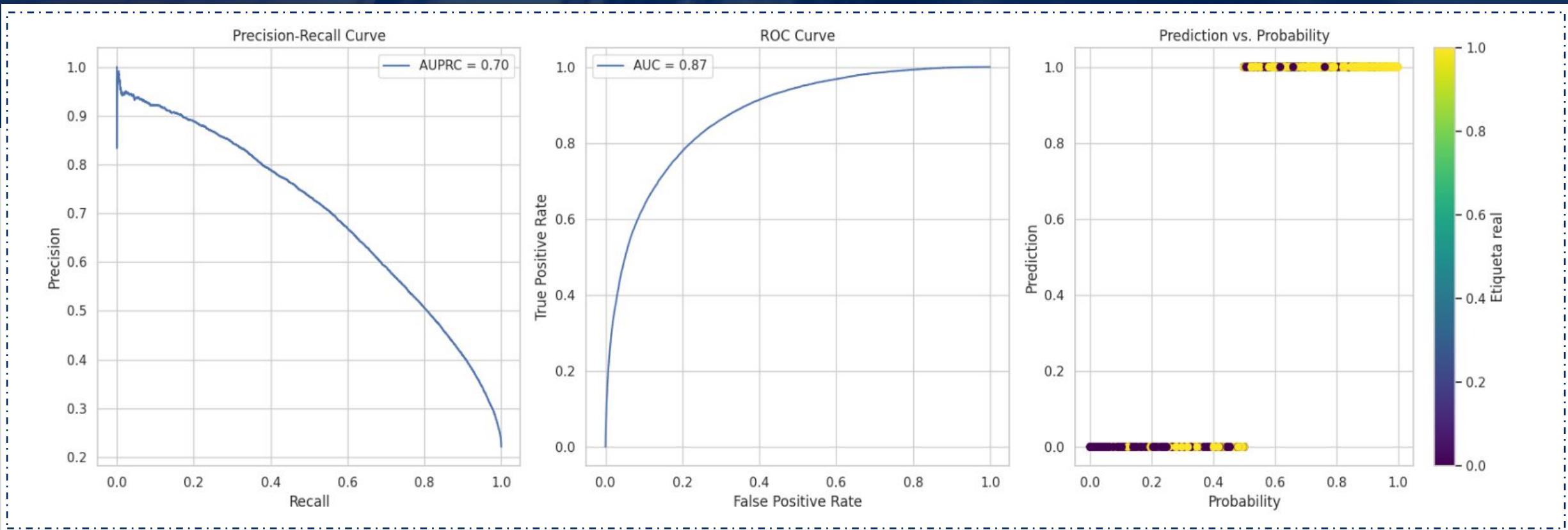
Metodologia

Neteja i preprocessament de dades



Metodologia

Millor mètrica



Mètrica escollida: AUPRC de la corba PR (Precision-Recall)

Metodologia

Millor model

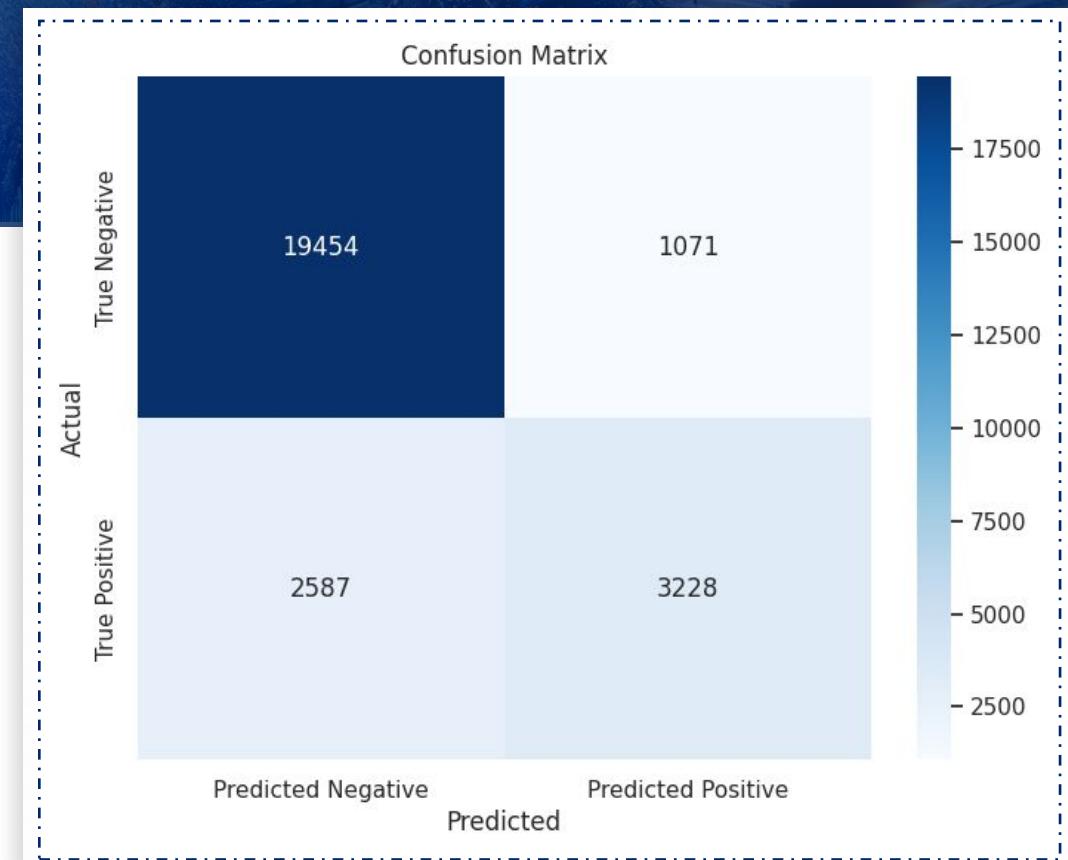
```
Millors hiperparàmetres per a LogisticRegression: {'C': 10}
Millors hiperparàmetres per a RandomForest: {'max_depth': 5, 'n_estimators': 200}
Millors hiperparàmetres per a KNN: {'n_neighbors': 7, 'weights': 'distance'}
Millors hiperparàmetres per a LGBM: {'max_depth': 5, 'n_estimators': 200}
Millors hiperparàmetres per a XGB: {'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200}
Millors hiperparàmetres per a AdaBoostClassifier: {'learning_rate': 1, 'n_estimators': 200}

Resultats AUPRC per a cada model: {'LogisticRegression': 0.6970009585777069,
Temps d'execució per a cada model: {'LogisticRegression': 75.98532748222351,
'RandomForest': 0.6829087545293079, 'KNN': 0.6695353831886269,
'RandomForest': 737.1510136127472, 'KNN': 273.4061703681946,
'LGBM': 0.7503989009941167, 'XGB': 0.756380227878523, 'AdaBoostClassifier': 0.7134465739682005}
'LGBM': 143.52469396591187, 'XGB': 335.1453778743744, 'AdaBoostClassifier': 1733.791310787201}
```

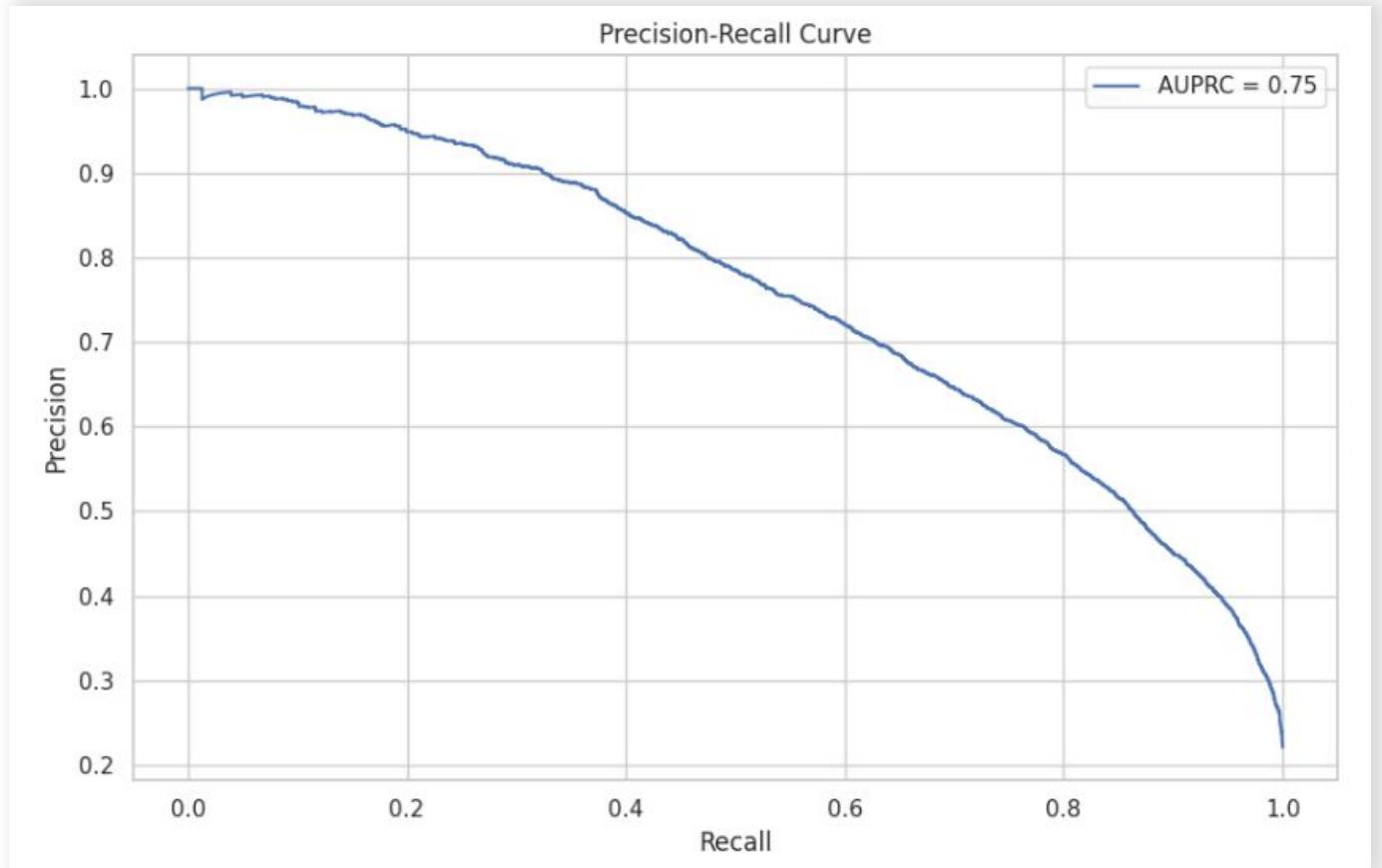
El millor model segons la mètrica AUPRC és: XGB amb una puntuació de **76%**

Results

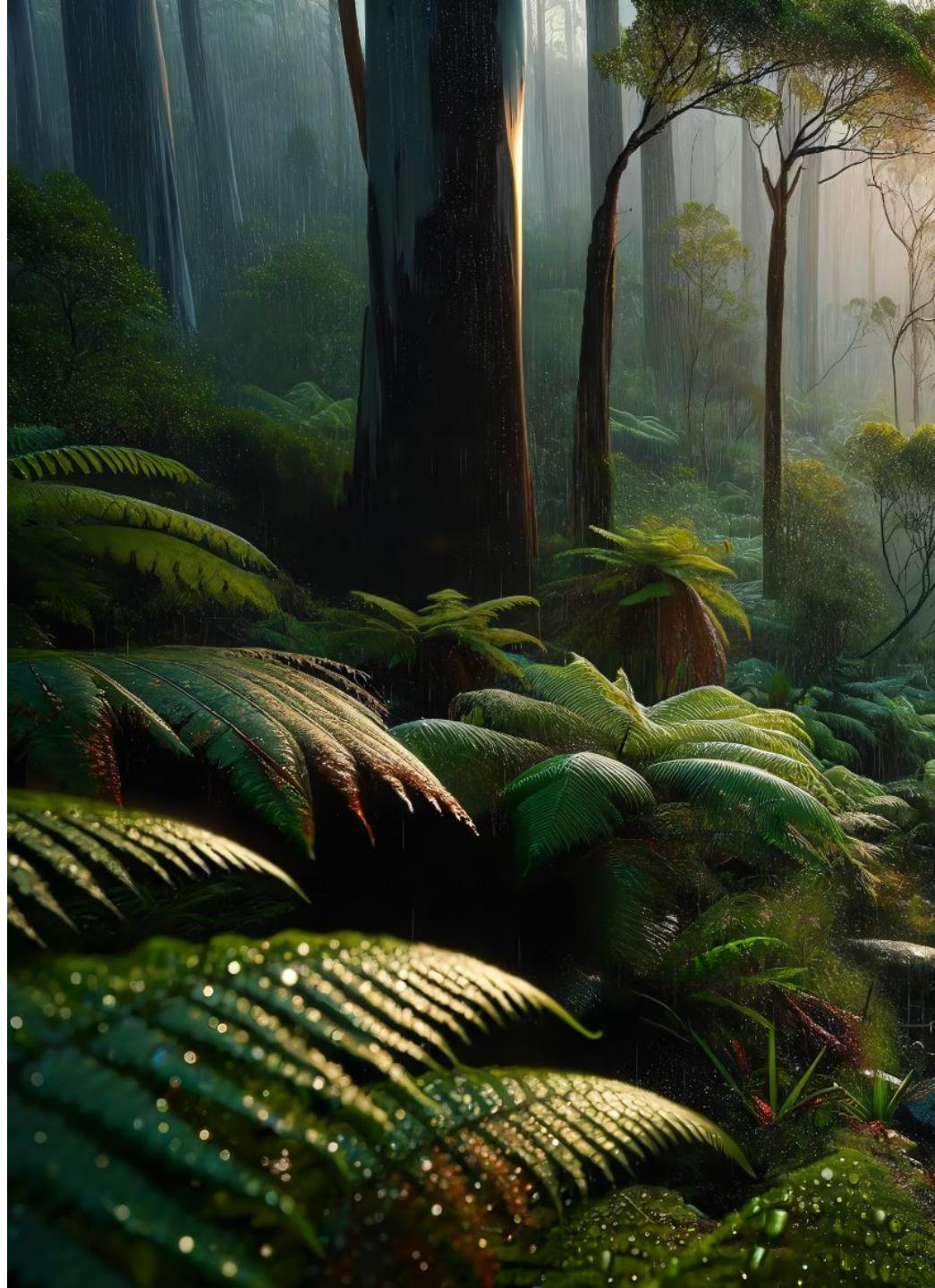
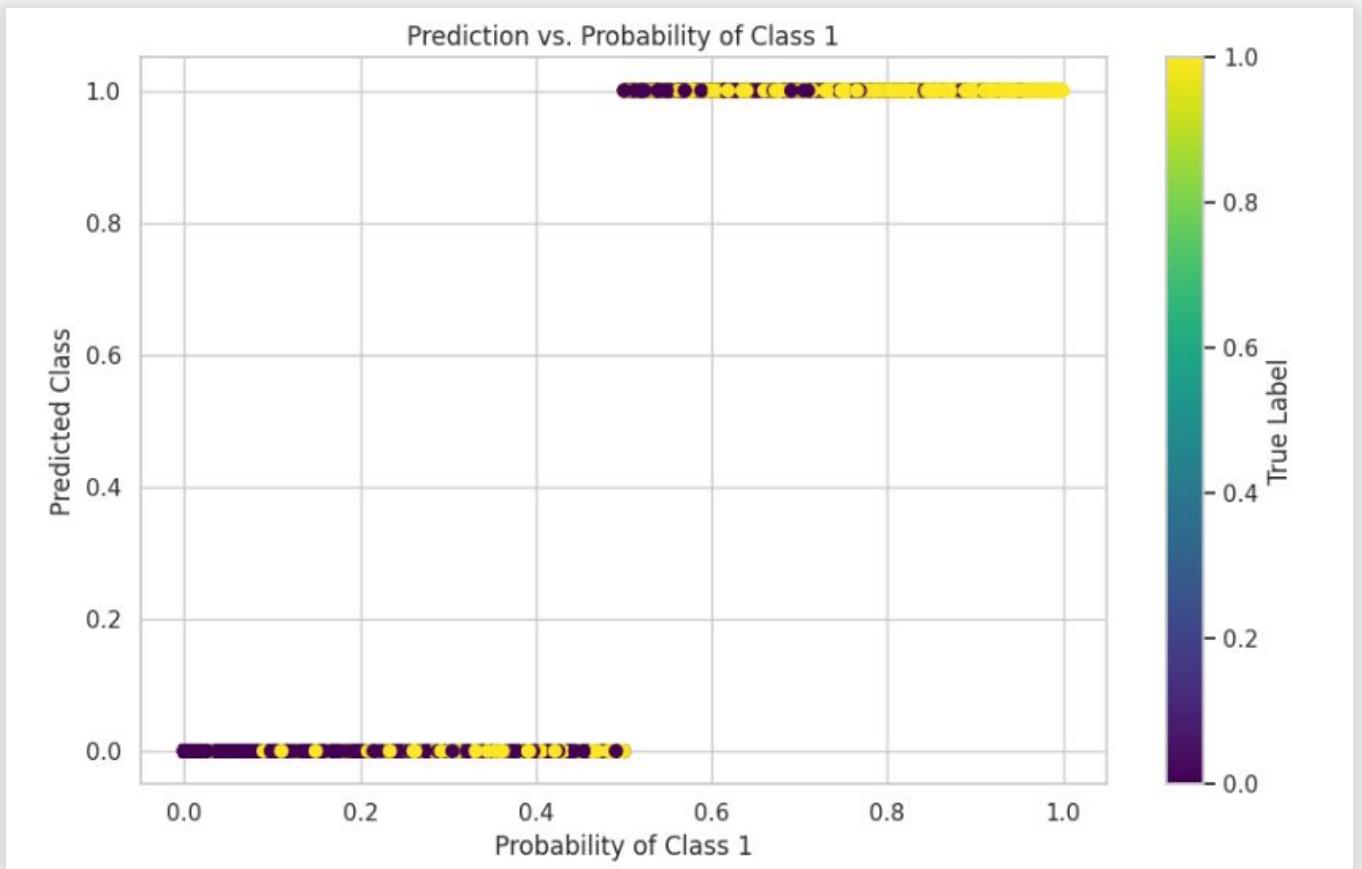
	precision	recall	f1-score	support
False	0.88	0.95	0.91	20525
True	0.75	0.56	0.64	5815
accuracy			0.86	26340
macro avg	0.82	0.75	0.78	26340
weighted avg	0.85	0.86	0.85	26340



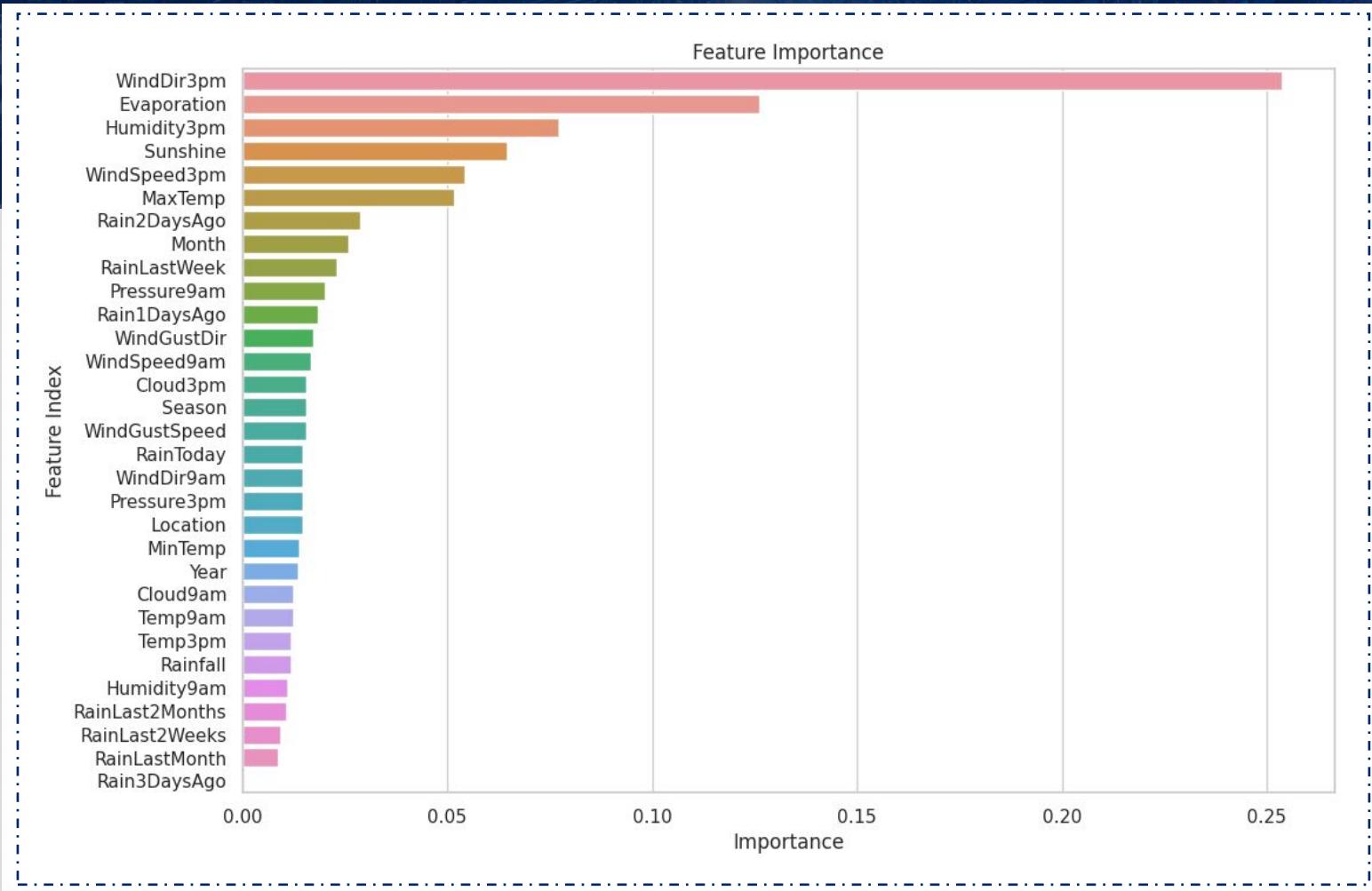
Résultats



Results

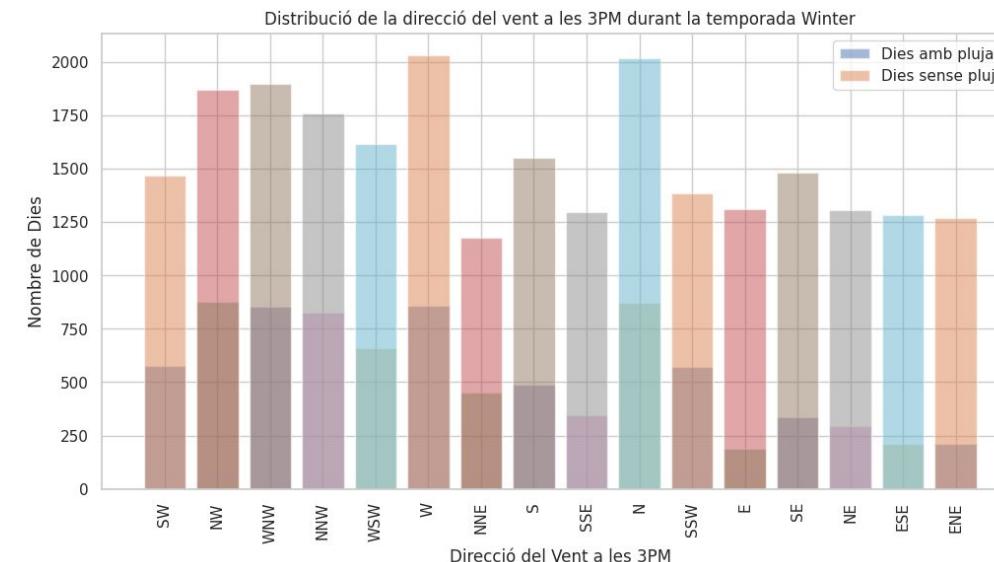
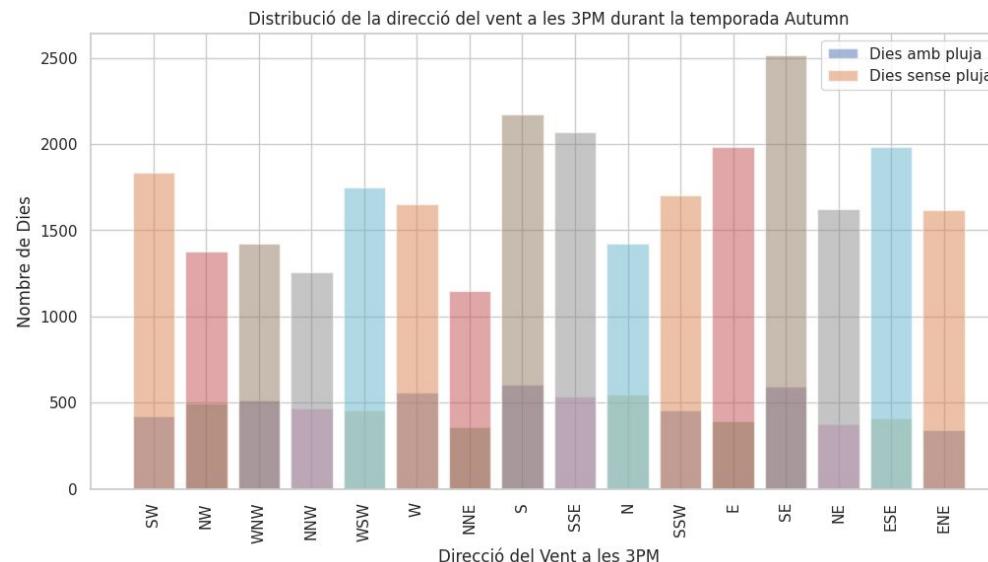
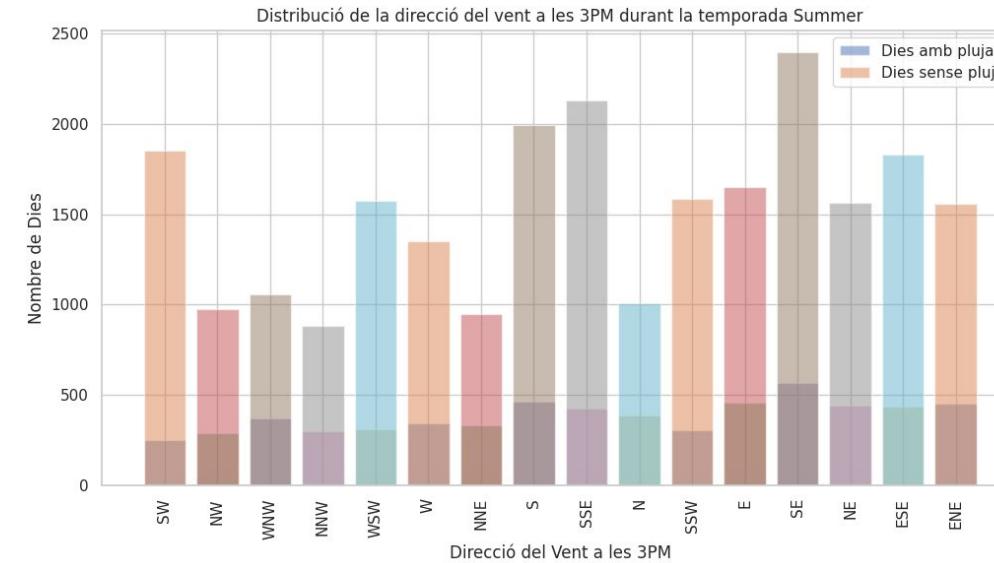
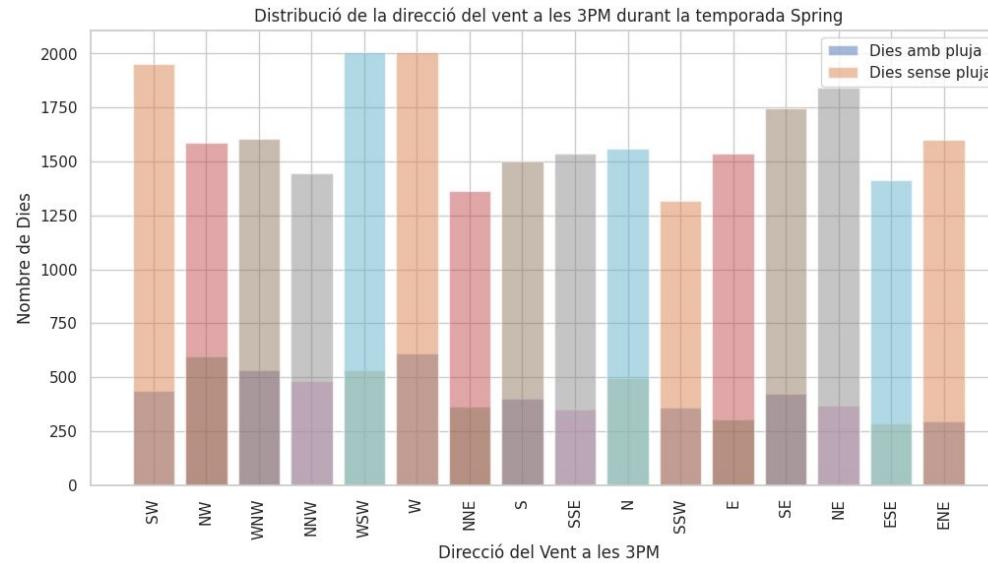


Results



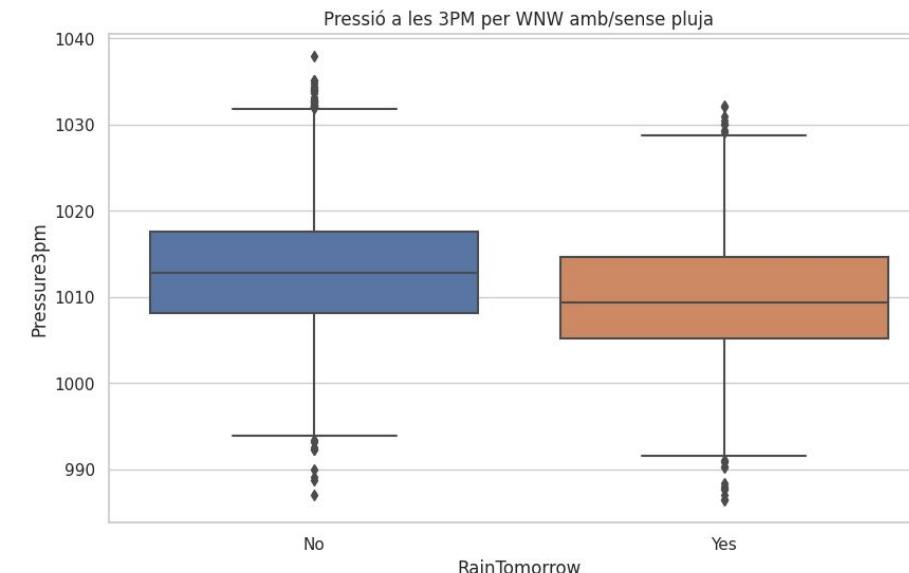
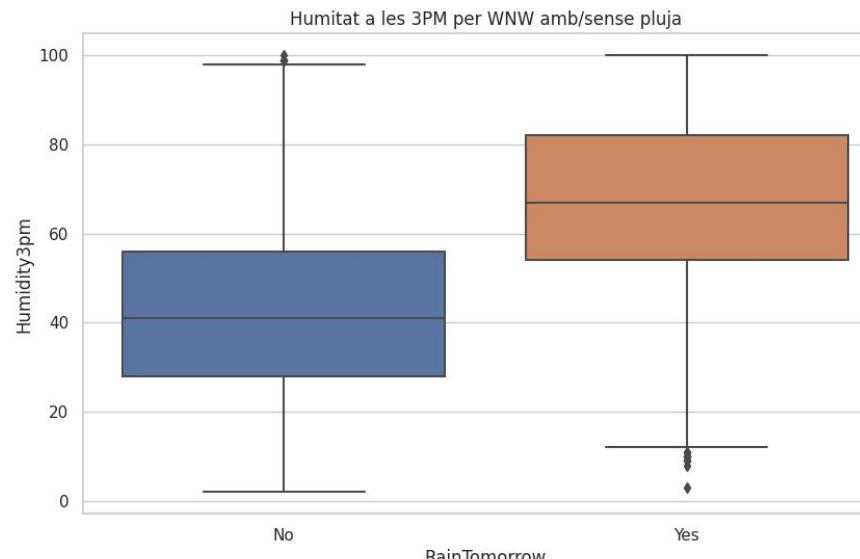
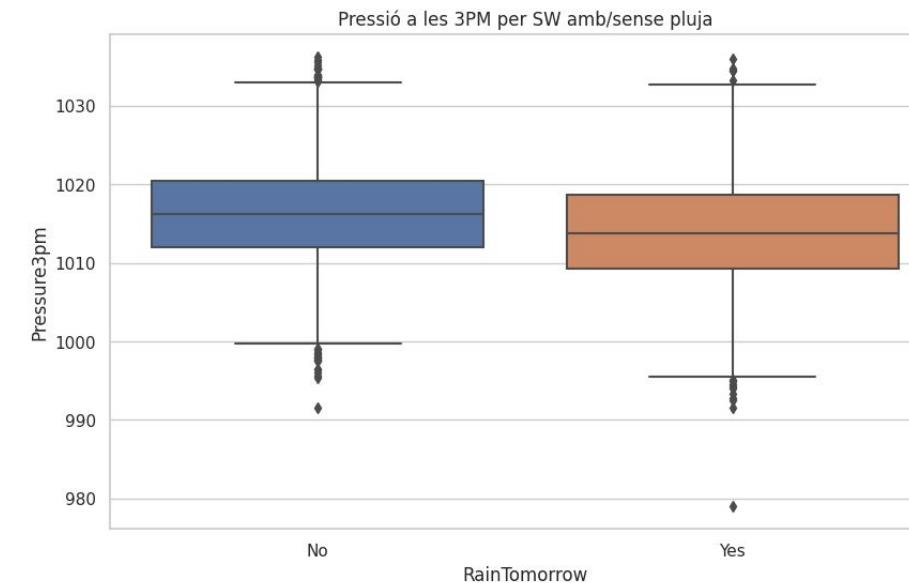
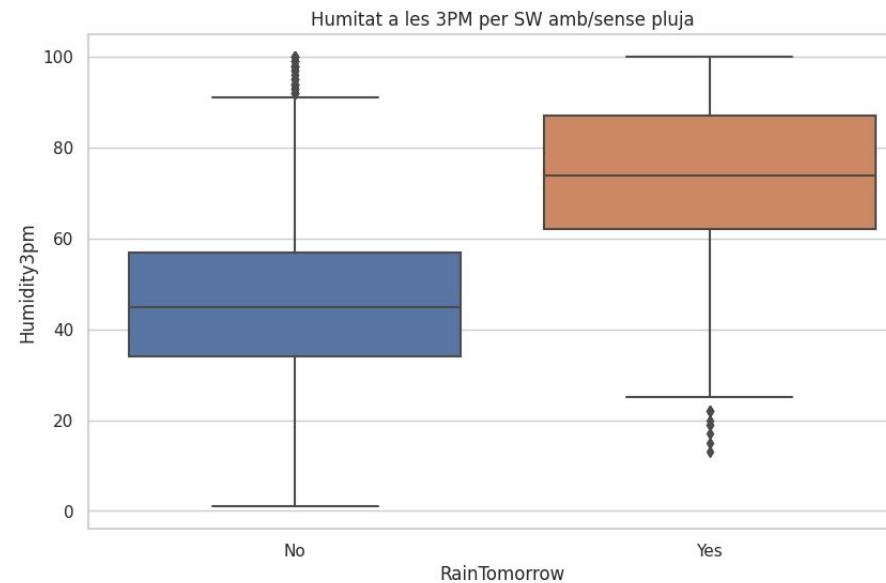
Resultats

A Austràlia, l'estació humida comença a finals de primavera i s'allarga tot l'estiu fins a l'inici de la tardor.



Resultats

Els vents del sud-oest i oest-nord-oest soLEN estar associats amb l'arribada de sistemes frontals que porten aire més fred i humit des de l'oceà.



01

XGBClassifier demostra un bon rendiment predint de pluja amb la mètrica AUPRC de 0.75, amb una exactitud general del 86% i tendeix a predir millor la classe negativa (dies sense pluja).

02

La **matriu de confusió** mostra una quantitat significativa de falsos negatius, el que reafirma la dificultat del model per capturar tots els dies plujosos.

03

La **direcció del vent a les 3PM, l'evaporació, la humitat a les 3PM i la llum solar** són les variables que tenen més influència en les decisions del model.

04

Les direccions del vent SW i WNW a les 3PM semblen tenir relació amb l'augment de la humitat i la disminució de la pressió atmosfèrica, dos indicadors clau de la probabilitat de pluja al dia següent.

05

Les direccions del vent SW i WNW semblen ser més freqüents en estacions plujosos com l'estiu.

Conclusions



MOLTES GRÀCIES!

