

# Classificació tiroidisme

Lluís Panal  
1668072



# Dataset

31 atributs:

- 20 booleans
- 8 numèrics (patient\_id)
- 3 categòrics (target)

9172 observacions

# Target

Possibles valors: '-', 'S', 'F', 'AK', 'R', 'I', 'M', 'N', 'G', 'K', 'A',  
'KJ', 'L', 'MK', 'Q', 'J', 'C|I', 'O', 'LJ', 'H|K', 'D', 'GK', 'MI', 'P',  
'FK', 'B', 'GI', 'C', 'GKJ', 'OI', 'D|R', 'E'

'A', 'B', 'C', 'D' ---> Hipertiroidisme (2)

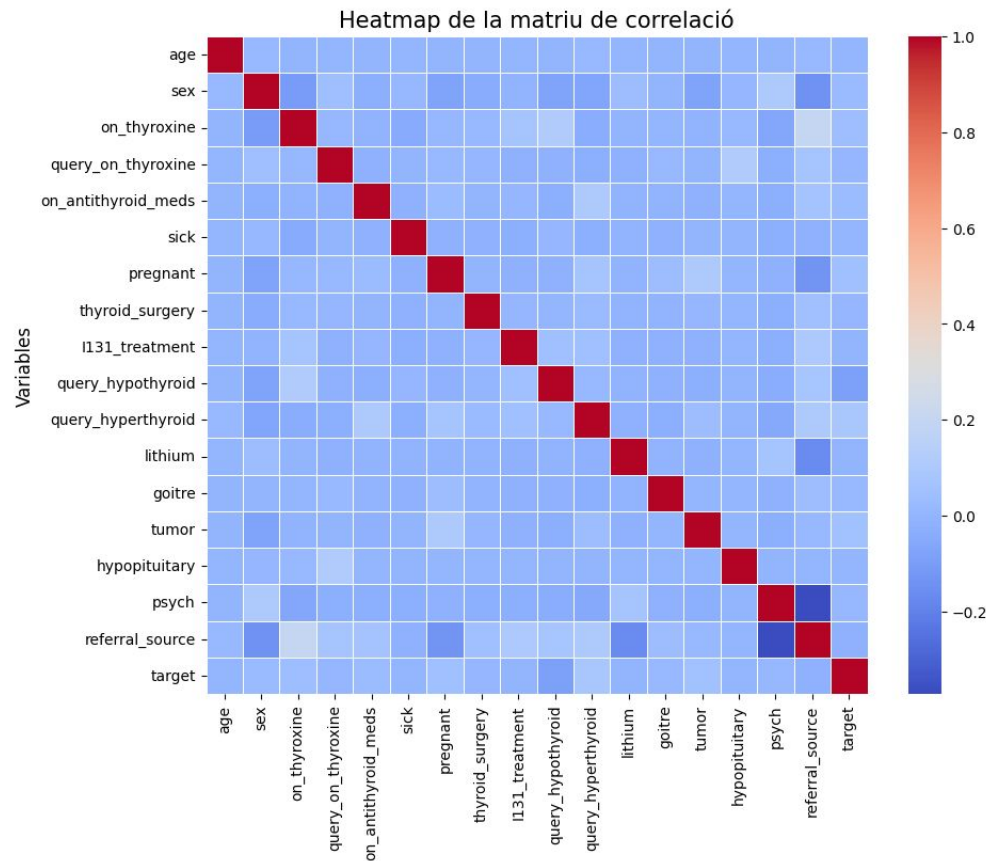
'E', 'F', 'G', 'H' ---> Hipotiroidisme (0)

Altres ---> Sa

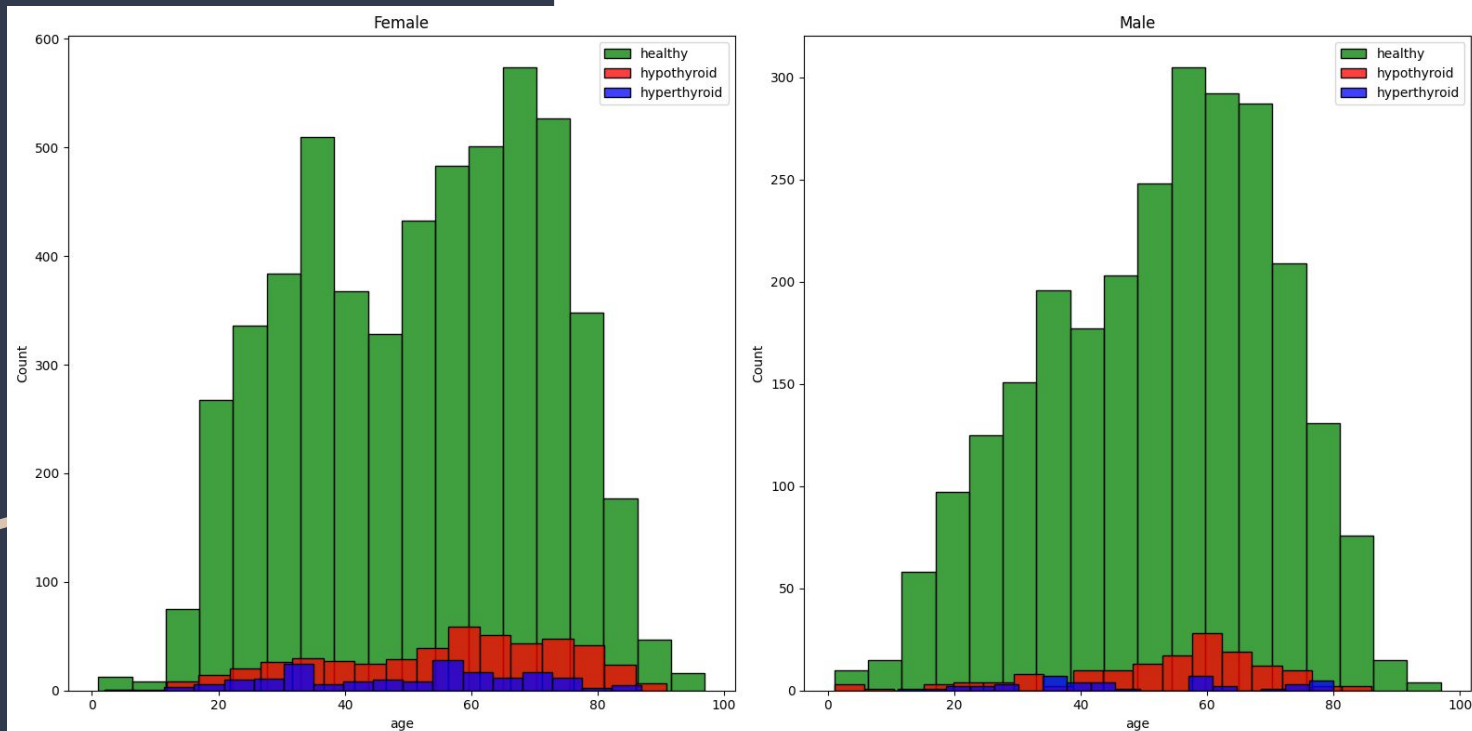
# Hormones

TSH, T3, TT4, T4U, FTI i TBG

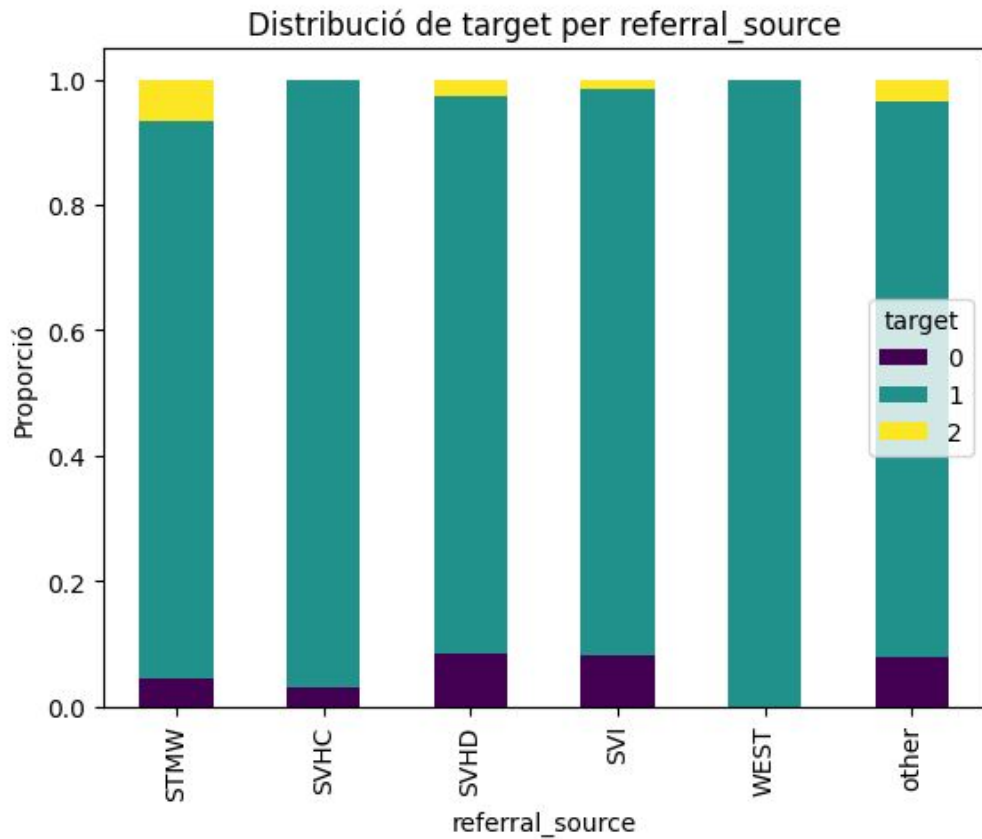
TSH	TSH_mesured
0.3	1
1.6	1
NaN	0



# Distribució del target



# Distribució del target



# Edat i font de referència

Número de mostra	Age
8105	65526
6392	65512
5710	65511
2976	455
790	97

Fonts de referència: STMW, SVHC, SVHD, SVI, WEST, other

Referral _source	STMW	SVHC	SVHD	SVI	other
'STMW'	1	0	0	0	0
'other'	0	0	0	0	1
'WEST'	0	0	0	0	0



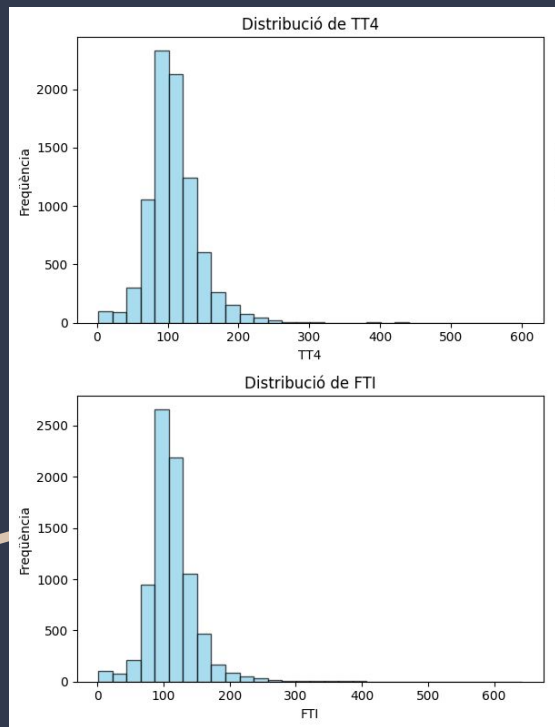
# Sex

3,3% de NaNs

Embarassat ---> dona

Eliminació

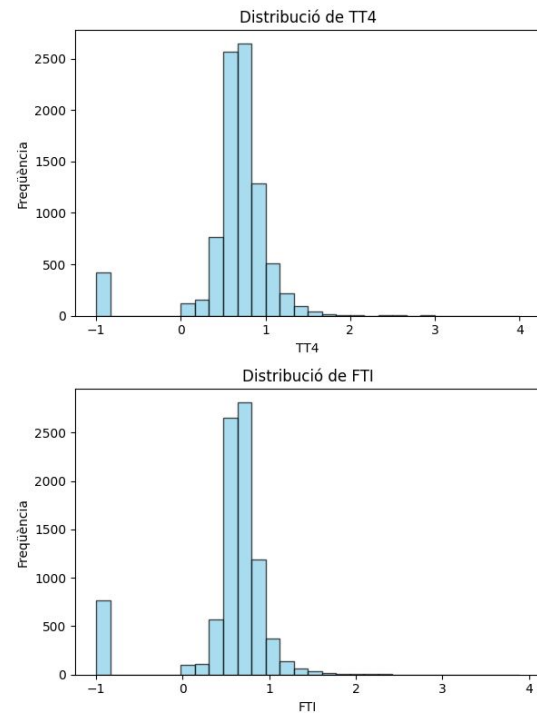
# Hormones I: omplir els NaNs



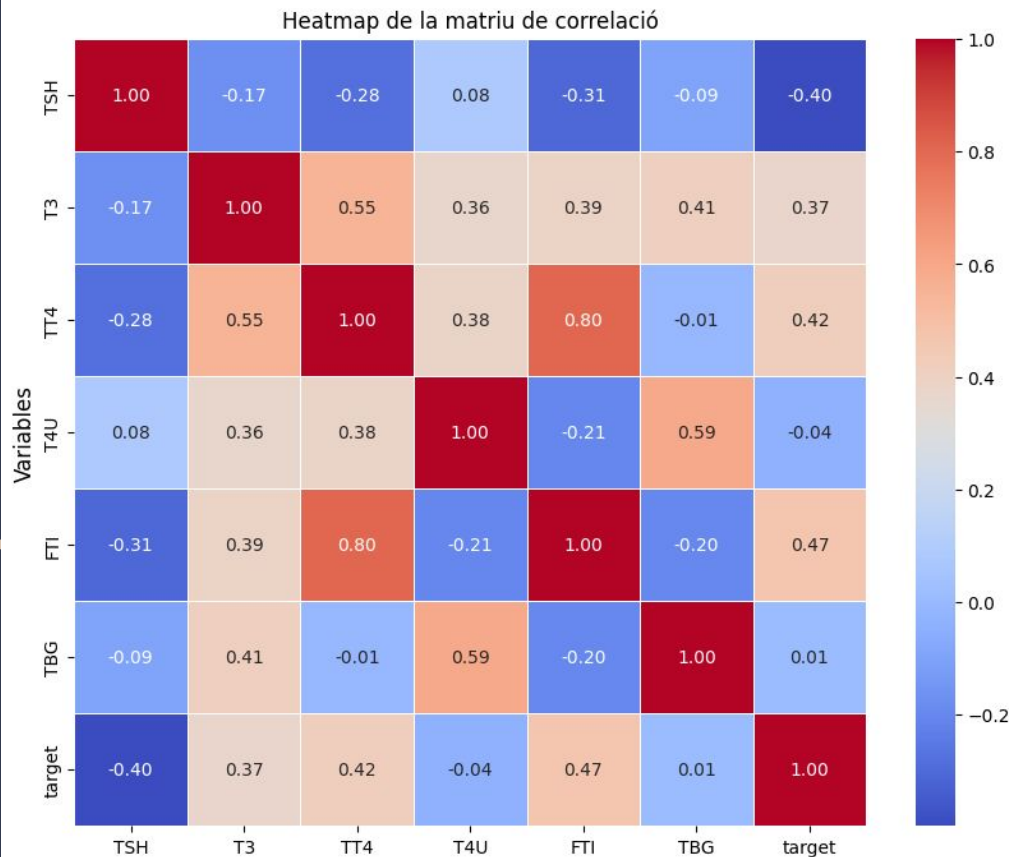
Escalar variables

Omplir amb -1

Age/100



# Hormones I: omplir els NaNs



# Hormones: discretització

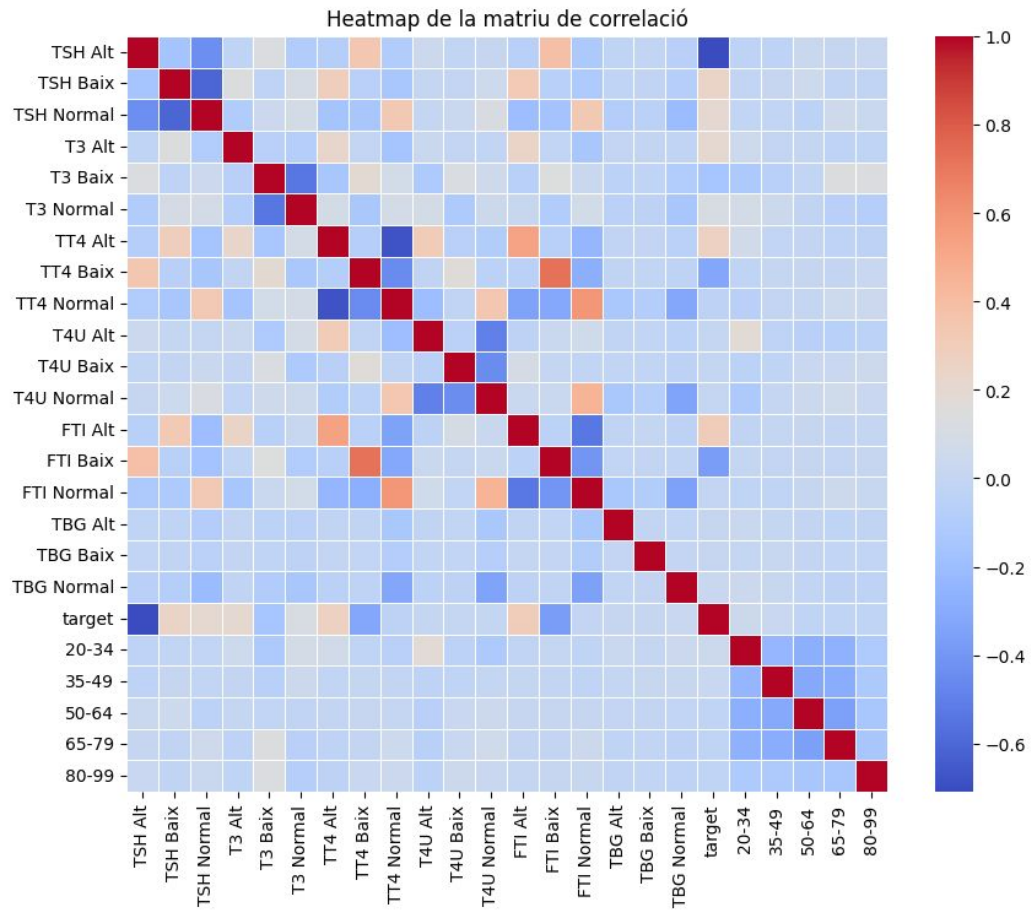
1 Hormona ---> 3 intervals (baix, normal, alt)

Age ---> 6 intervals

TSH	Baix	Normal	Alt
0.2	1	0	0
1.6	0	1	0
9.8	0	0	1
NaN	0	0	0

Valors normals pel TSH: 0.26-5.6

# Hormones: discretització



# Rendiment inicial

Amb un gradient boosting

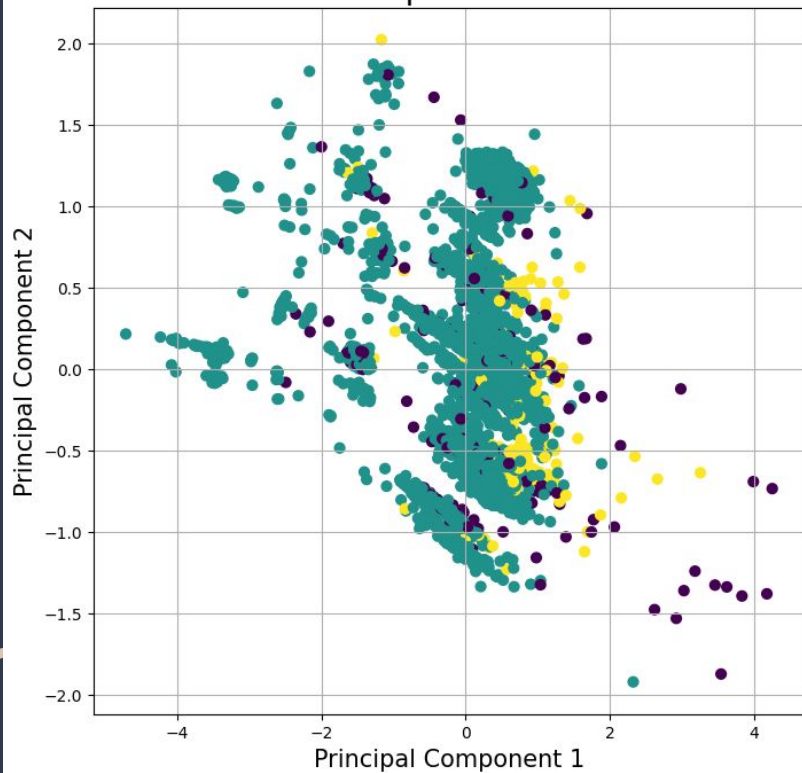
5 splits

Estratègia	Train	Test
Omplir amb -1	0.97	0.9'
Intervals	0.91	0.87

# PCA

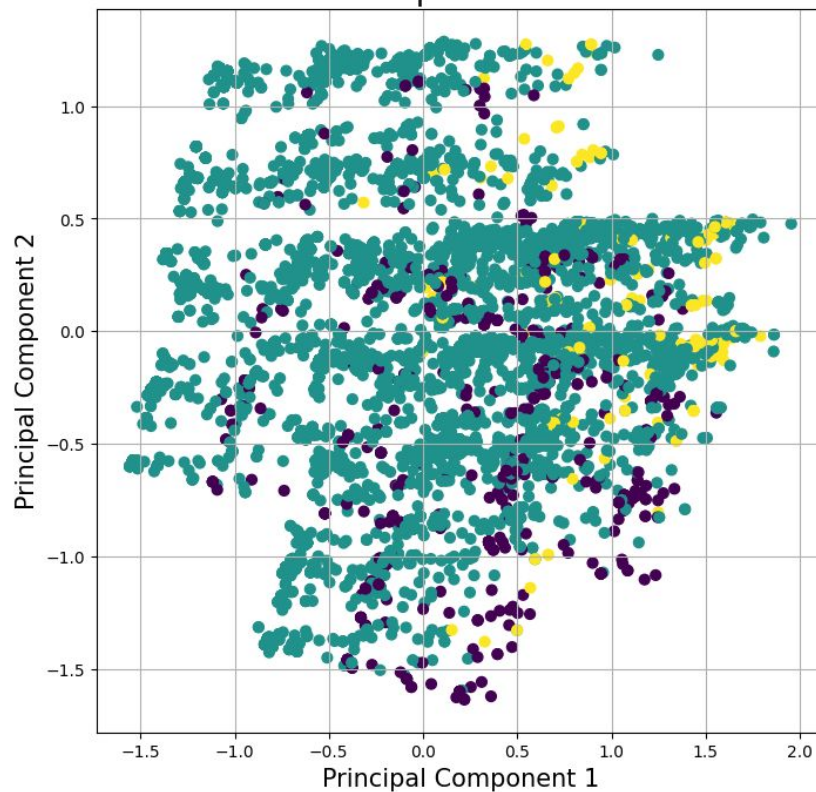
Omplint NaNs:

2 component PCA



Intervals:

2 component PCA



# Mètrica

Dataset molt desbalancejat

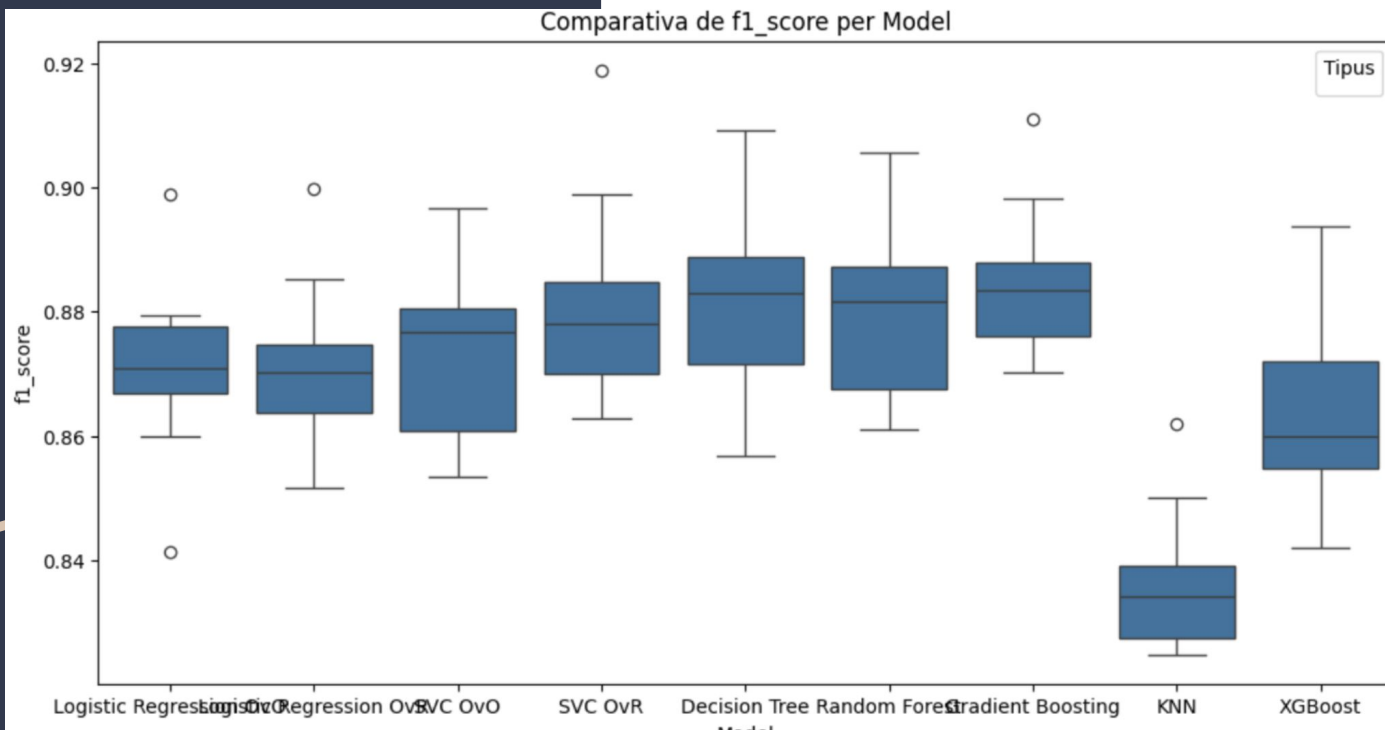
3 classes



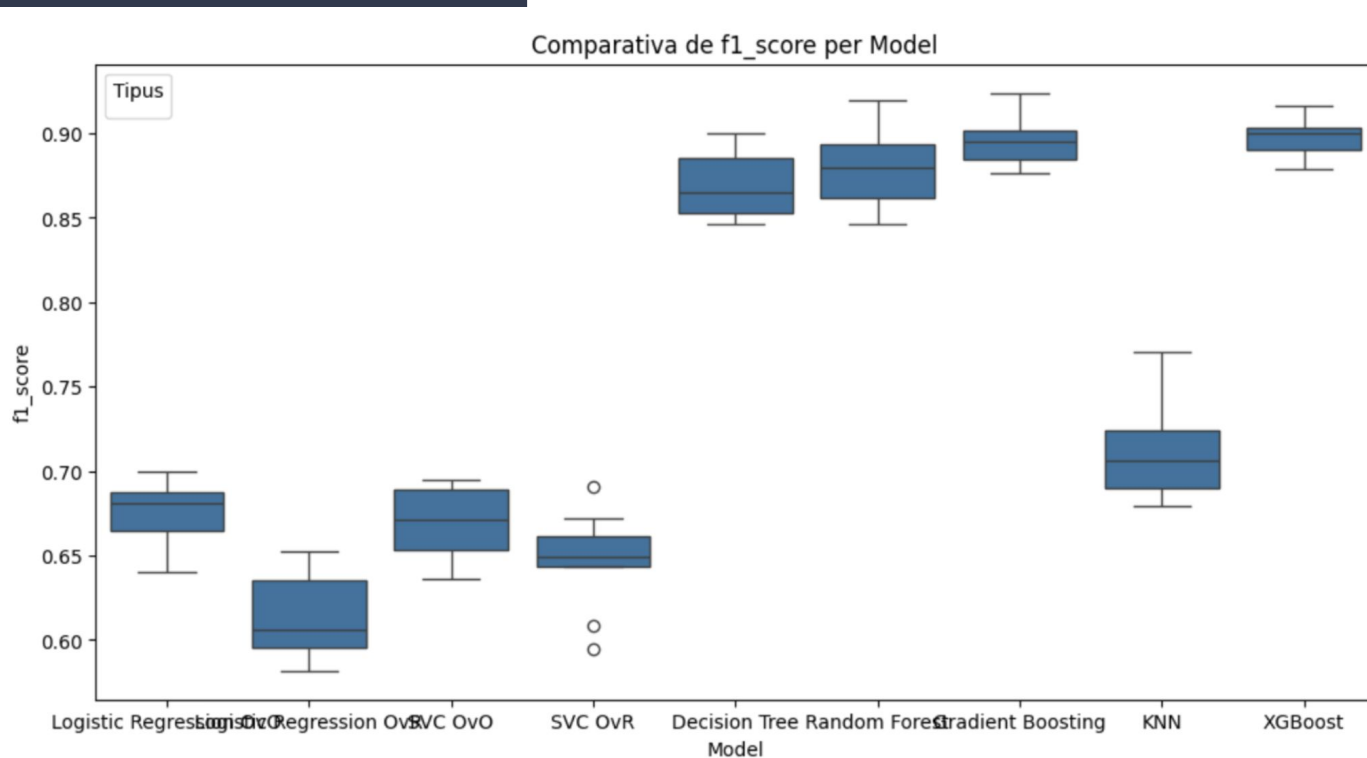
F1 score  
(average='macro')



# Models per defecte: omplint NaNs



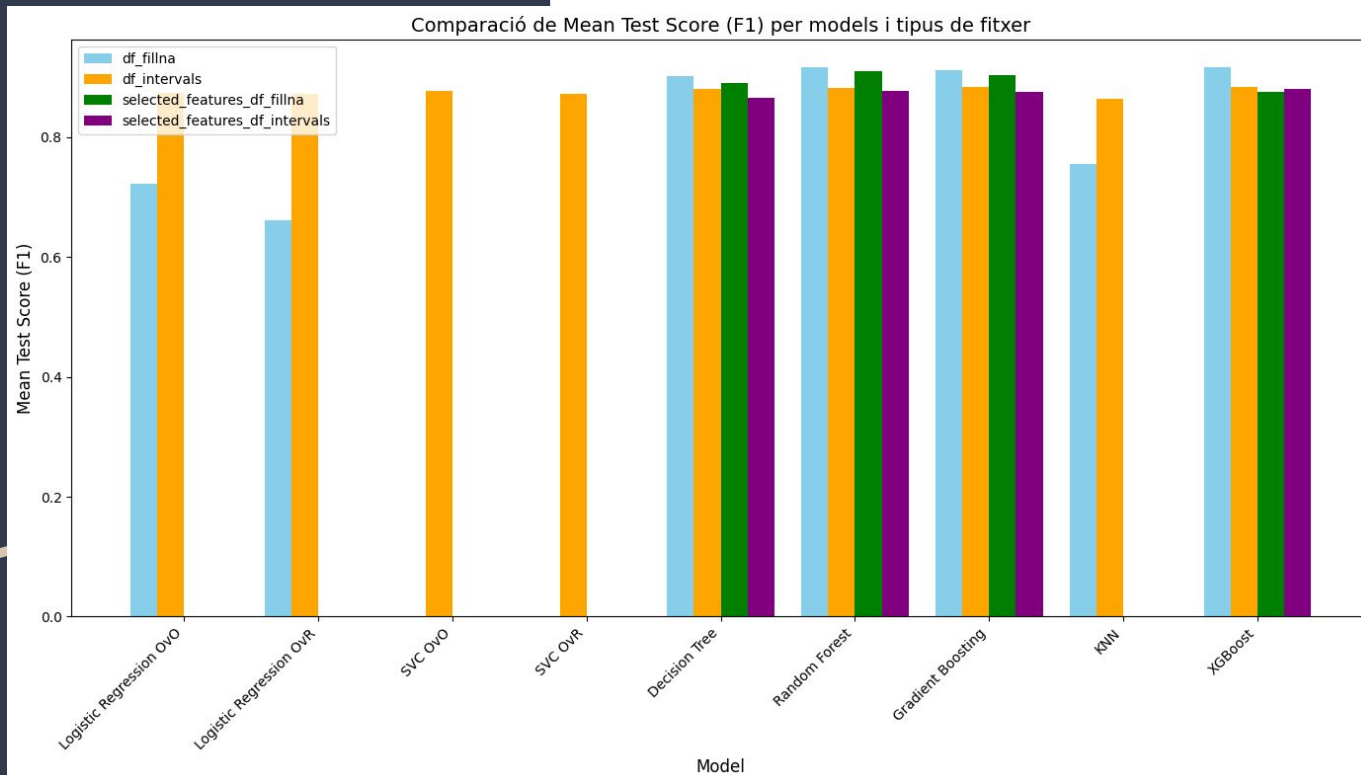
# Models per defecte: intervals



## Models

## Caractéristiques

# Cerca d'hiperparàmetres



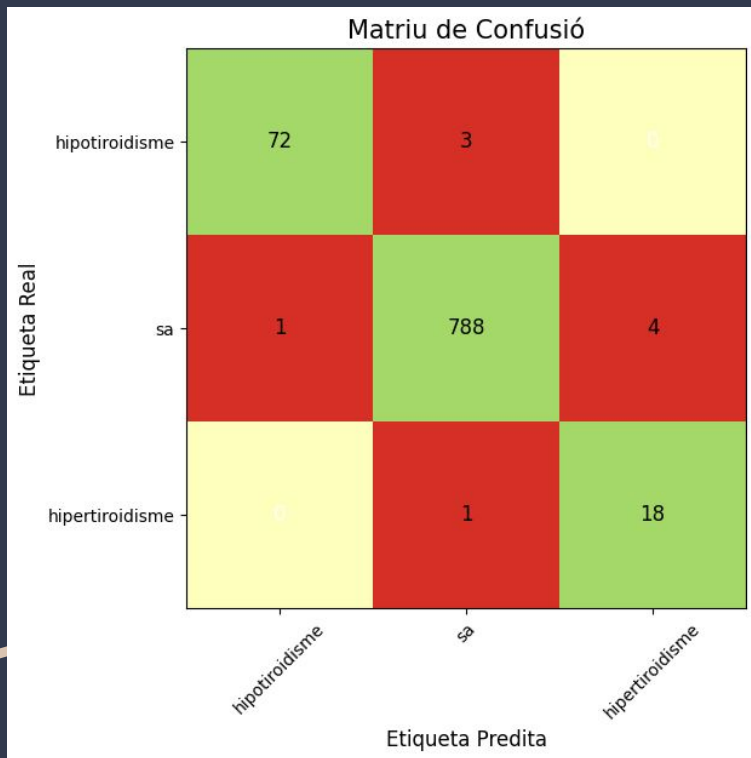
# Millor model

Dataset: omplir NaNs

Selected features: No

```
RandomForestClassifier  
RandomForestClassifier(criterion='entropy', max_features=None,  
                        min_samples_leaf=2, min_samples_split=10,  
                        random_state=20)
```

# Resultat final



• F1 score macro: 0.949

• F1 score micro: 0.988

• Accuracy: 0.988

Classe	Precision	Recall	F1-Score	Support
Hipotiroidisme	0.99	0.96	0.97	75
Sa	0.99	0.99	0.99	793
Hipertiroidisme	0.82	0.95	0.88	19