

Avaluació de models d'AC per a la detecció anticipada d'ictus

Lucia Garrido, Adrià Muro i David Morillo

ÍNDEX

01

Introducció

02

Dades i processament

03

Metodologia i
arquitectura dels models

04

Resultats

05

Conclusions i
possibles millores

A decorative graphic on the left side of the slide, consisting of several overlapping squares in different shades of gray. The squares are arranged in a way that creates a sense of depth and movement, with some squares appearing to be in front of others. The overall effect is a modern, geometric design.

01

Introducció

INTRODUCCIÓ

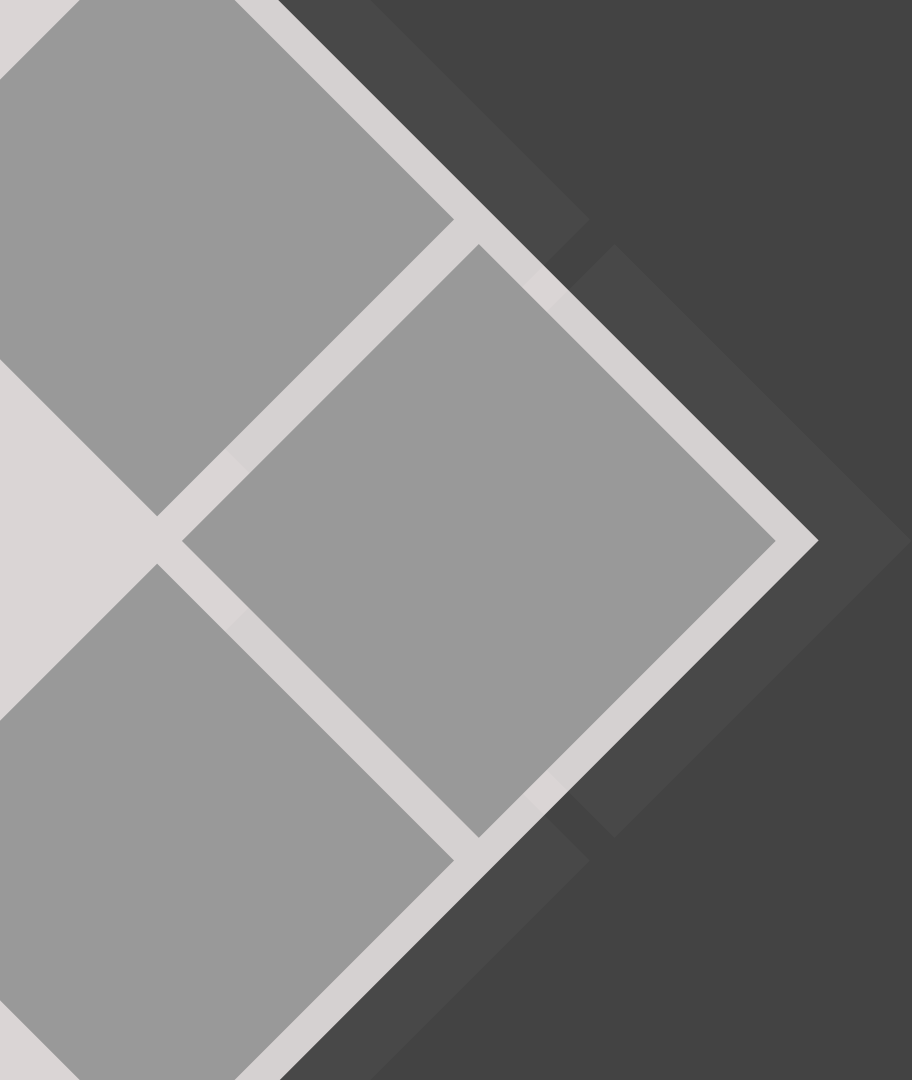
Motivació:

- Prevenció proactiva i reducció de factors de risc
- Accessibilitat per a persones sense recursos mèdics

Es pot predir un ictus a partir del dataset triat?

Es pot predir un ictus sense saber les característiques mèdiques dels pacients?





02

Dades i
processament

DATASET

- “Stroke Prediction Dataset” (Kaggle)
- 4908 entrades
- 20 % de test (982) i 80% d’entrenament (3926)

| | id | gender | age | hypertension | heart_disease | ever_married | work_type | Residence_type | avg_glucose_level | bmi | smoking_status | stroke |
|---|-------|--------|------|--------------|---------------|--------------|---------------|----------------|-------------------|------|-----------------|--------|
| 0 | 9046 | Male | 67.0 | 0 | 1 | Yes | Private | Urban | 228.69 | 36.6 | formerly smoked | 1 |
| 1 | 51676 | Female | 61.0 | 0 | 0 | Yes | Self-employed | Rural | 202.21 | NaN | never smoked | 1 |
| 2 | 31112 | Male | 80.0 | 0 | 1 | Yes | Private | Rural | 105.92 | 32.5 | never smoked | 1 |
| 3 | 60182 | Female | 49.0 | 0 | 0 | Yes | Private | Urban | 171.23 | 34.4 | smokes | 1 |
| 4 | 1665 | Female | 79.0 | 1 | 0 | Yes | Self-employed | Rural | 174.12 | 24.0 | never smoked | 1 |

PROBLEMES AMB EL DATASET

Problema principal:

- Desequilibri de dades: 209/4908 files positives → 4,26%

Conseqüència:

- El model tendeix a etiquetar totes les dades com a negatives



Solució inicial:

- Retall de dades per forçar el mateix nombre de positius i negatius
 - Es perd la major part de les dades (91,48%)

SOLUCIÓ (SMOTE)

Generar exemples sintètics de la classe minoritària en lloc de duplicar dades existents.

- Per cada mostra de la classe minoritària, es busquen els seus veïns més propers.
- Es creen noves mostres combinant aleatòriament la mostra original i els seus veïns.
- Només aplicat a les dades d'entrenament

A decorative graphic on the left side of the slide, consisting of several overlapping diamond shapes in various shades of gray, creating a geometric pattern.

03

Metodologia i arquitectura dels models

MODELS APLICATS

| Models simples | Models avançats |
|---------------------|-----------------|
| Logistic Regression | Random Forest |
| Naive Bayes | AdaBoost |
| K-Nearest Neighbors | XGBoost |

MÈTRIQUES MÉS IMPORTANTS

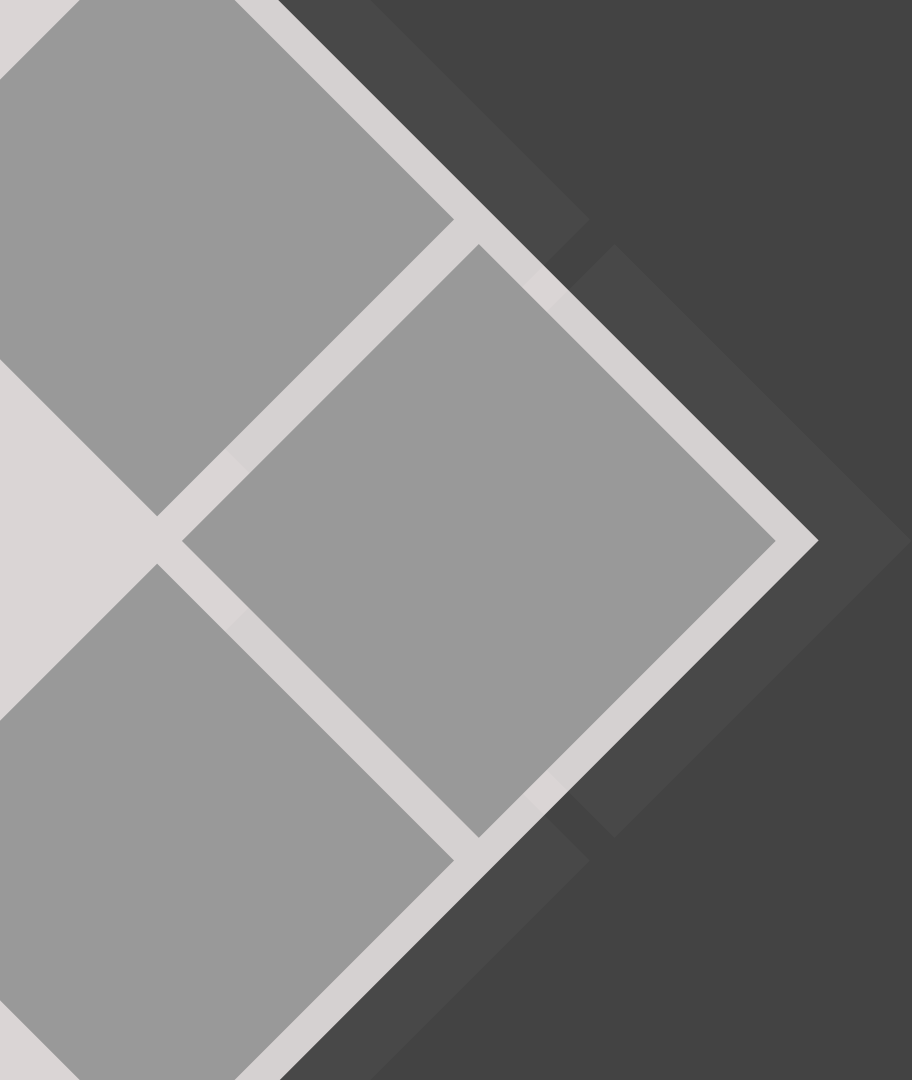
- **Recall**
 - Millor falsos positius que falsos negatius
- **Accuracy**
 - Confiança del model
 - Balanceig global

Accuracy + Recall

EXTRACCIÓ DE CARACTERÍSTIQUES NO MESURABLES A CASA

Explicar quines característiques hem tret i quines queden

Treiem avg_glucose, bmi, obesity, hyperthension, heart_disease

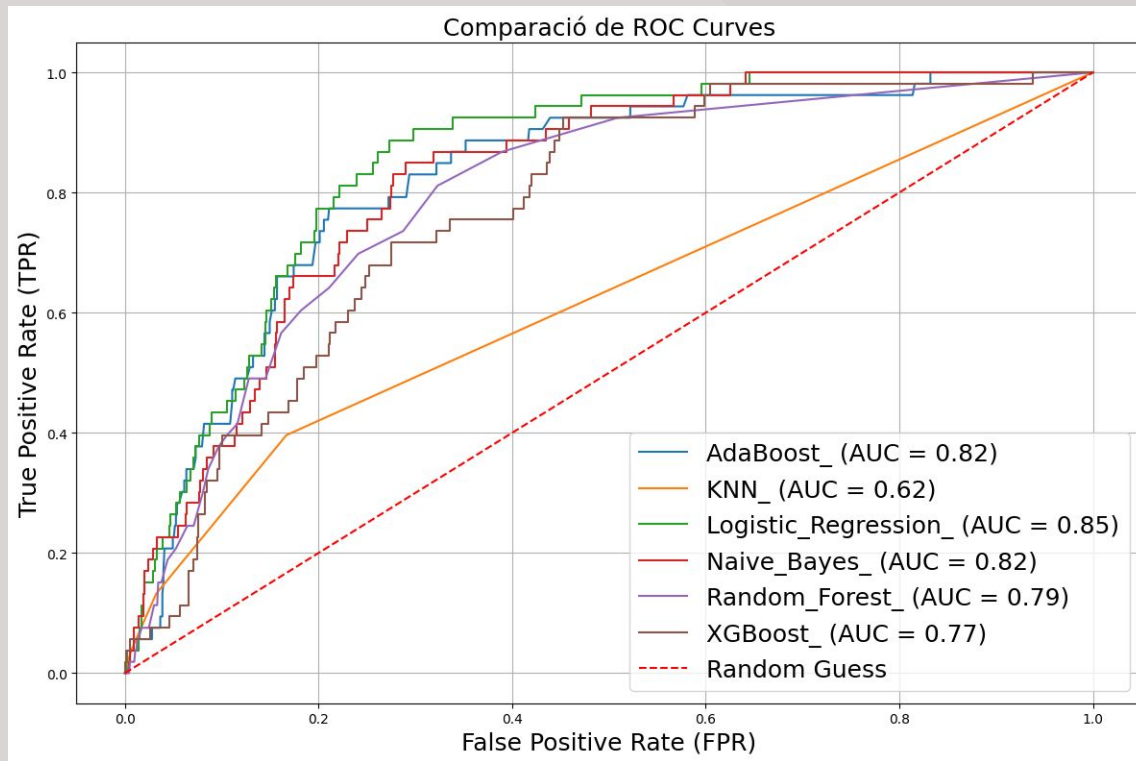


04

Resultats

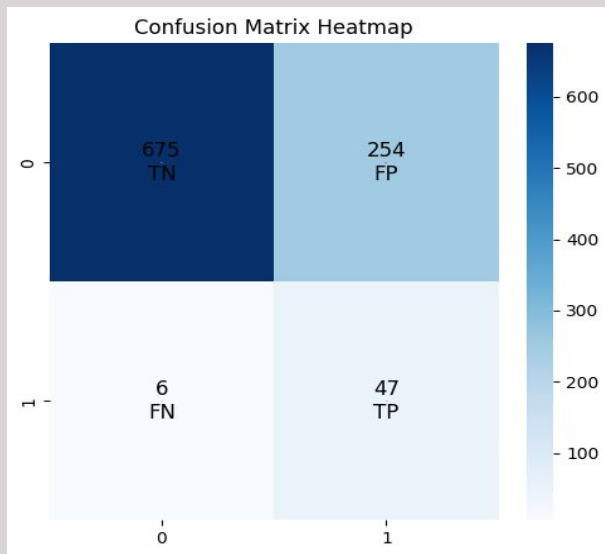
**Es pot predir un ictus a partir
del dataset triat?**

ROC CURVE DE TOTS ELS MODELS

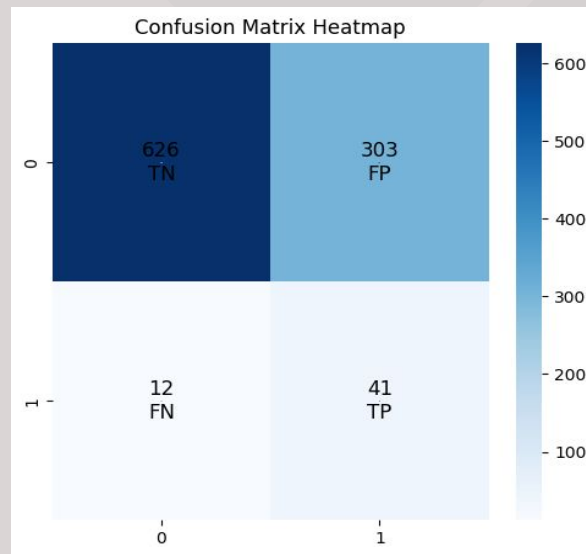


MATRIUS DE CONFUSIÓ

Logistic Regression



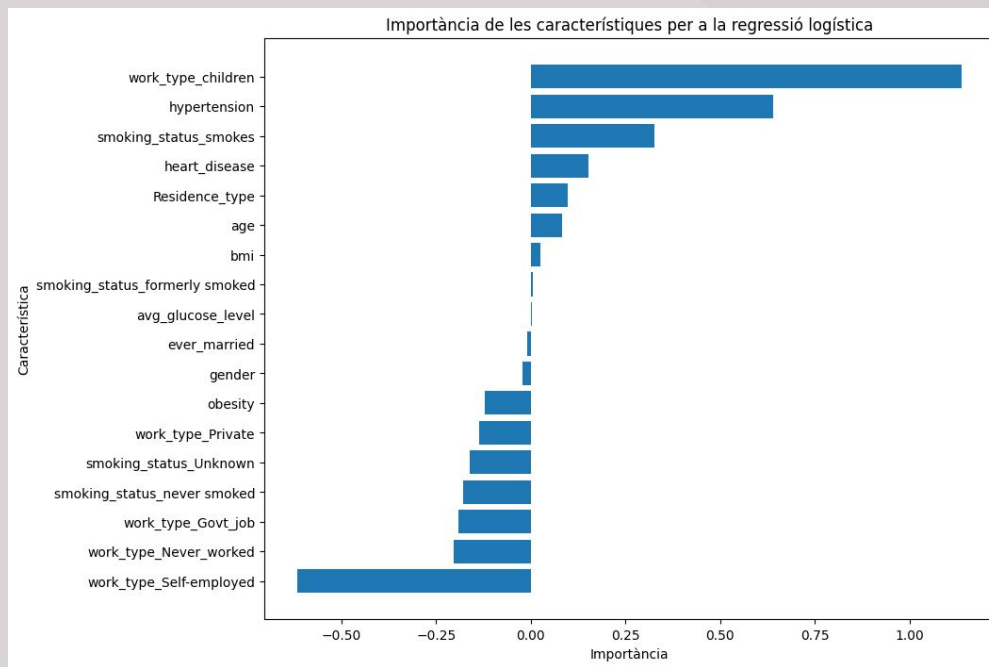
XG Boost



MATRIUS DE CONFUSIÓ

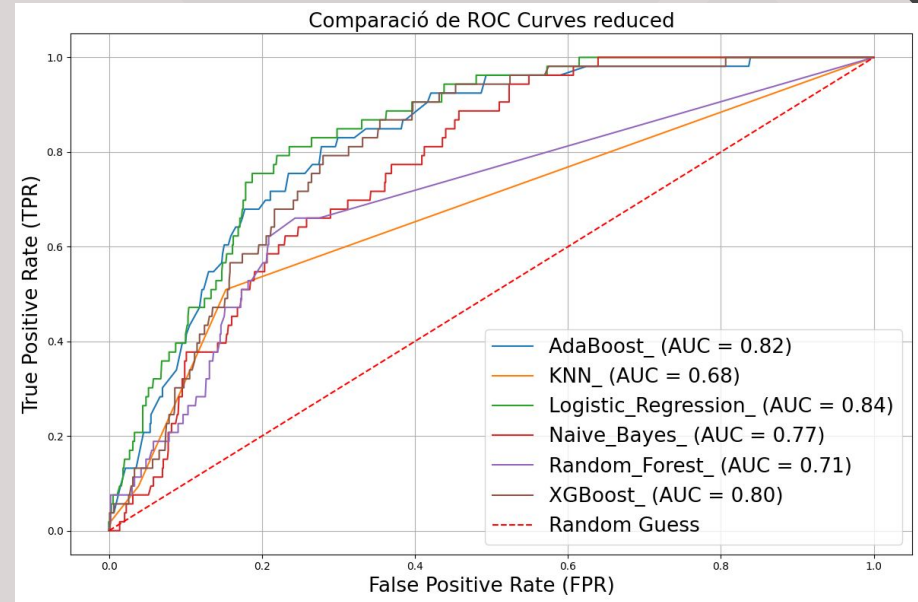
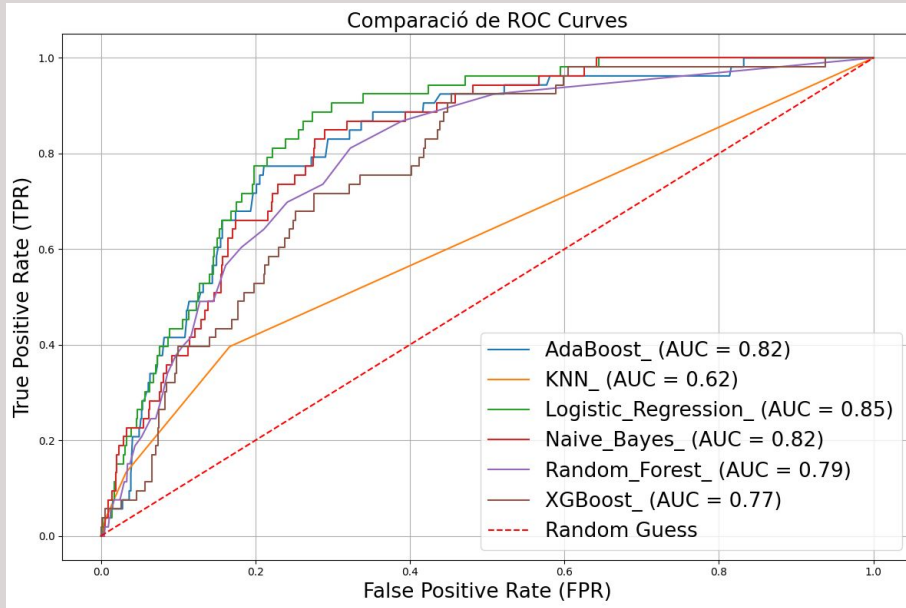
| | Logistic Regression | XGBoost |
|-----------|---------------------|---------|
| Precision | 0.1561 | 0.1192 |
| Accuracy | 0.7352 | 0.6792 |
| Recall | 0.8868 | 0.7736 |
| F1 Score | 0.2655 | 0.2065 |

CARACTERÍSTIQUES MÉS IMPORTANTS

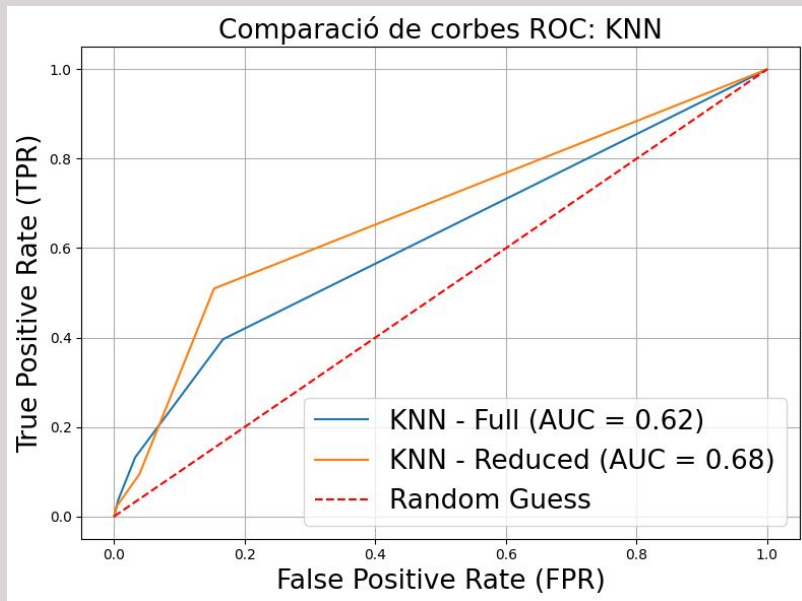


**Es pot predir un ictus sense
saber les característiques
mèdiques dels pacients?**

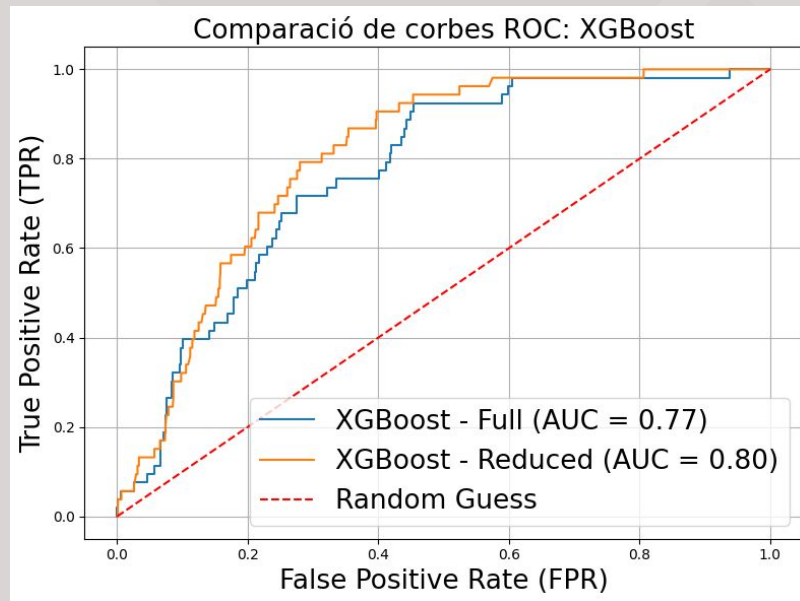
ROC CURVES TOTS ELS MODELS



COMPARACIÓ DE CADA MODEL

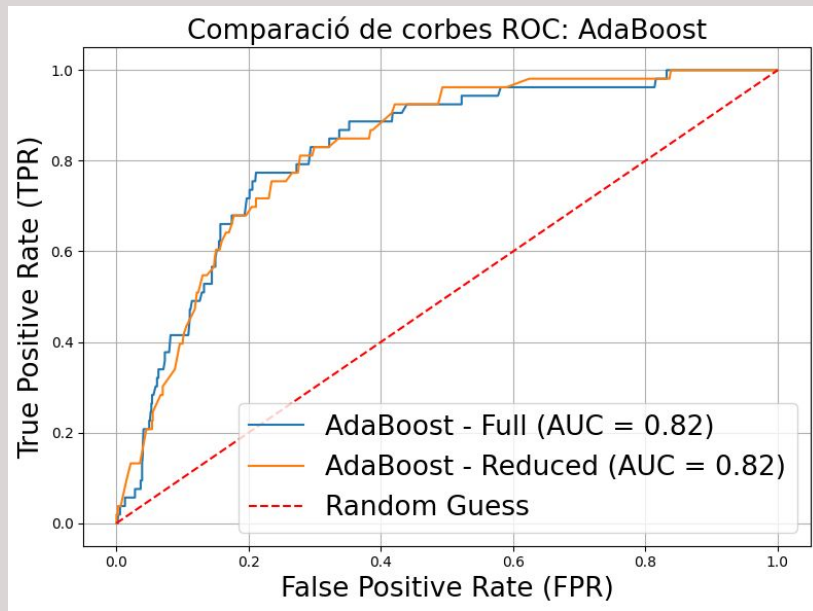


KNN

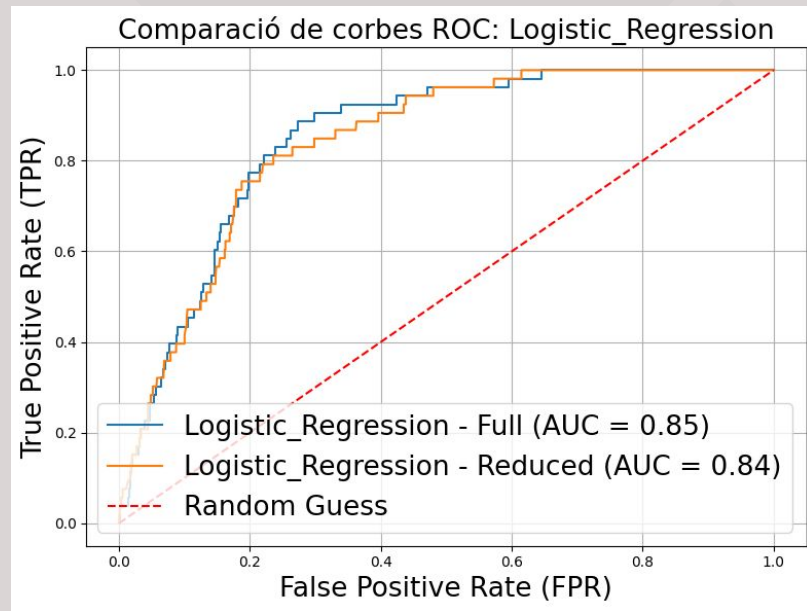


XGBoost

COMPARACIÓ DE CADA MODEL

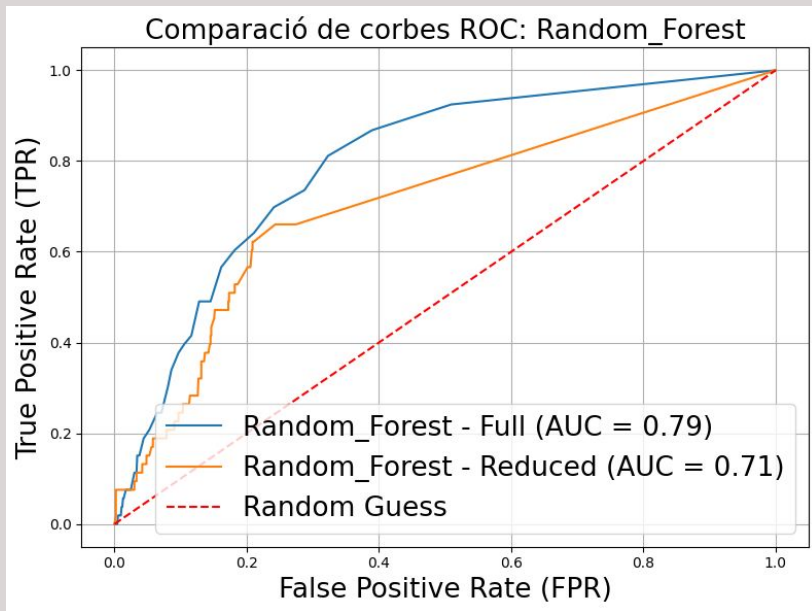


Adaboost

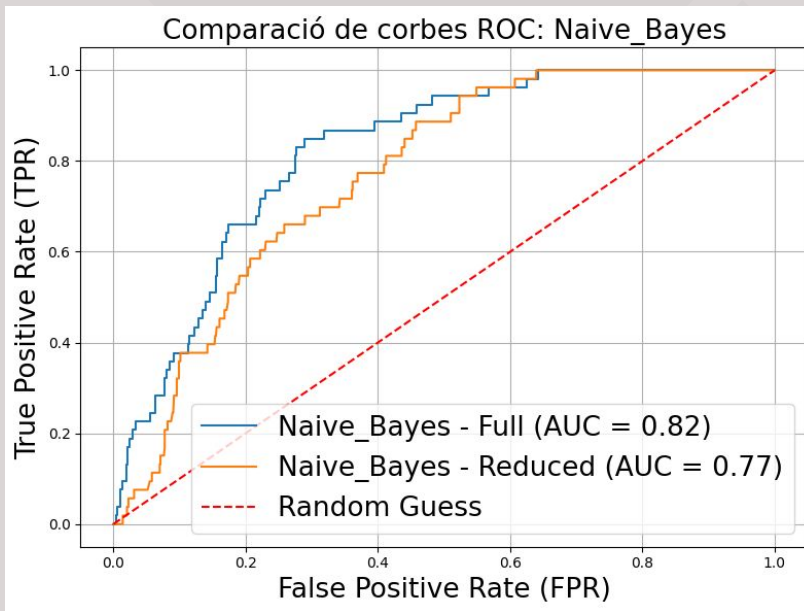


Logistic Regression

COMPARACIÓ DE CADA MODEL



Random Forest



Naive Bayes

A decorative graphic on the left side of the slide, consisting of several overlapping diamond shapes. The diamonds are in various shades of gray, with some having white outlines. They are arranged in a way that creates a sense of depth and movement, pointing towards the right.

05

Conclusions i
possibles millores

CONCLUSIONS GENERALS

- S'ha pogut estudiar diversos models que satisfan la pregunta inicial
 - Millors models:
 - Regressió Logística
 - XGBoost
 - Valors de Recall i Accuracy decents
- Es poden fer prediccions sense dades estrictament mèdiques

MILLORES O TREBALL FUTUR

- Recopilació de més dades
 - Tipus de dades
 - Activitat física
 - Hàbits alimentaris
 - Quantitat
 - Més dades de casos positius

A decorative graphic on the left side of the slide, consisting of several overlapping diamond shapes. The diamonds are in various shades of gray, with some having white outlines. They are arranged in a way that creates a sense of depth and movement, pointing towards the right.

Gràcies!