

CARONTE CHALLENGE

Datathon 2024

Arnau Busquets, David Franch, Marc Ugas, Jordi Ventosa

COM HO HEM ORGANITZAT

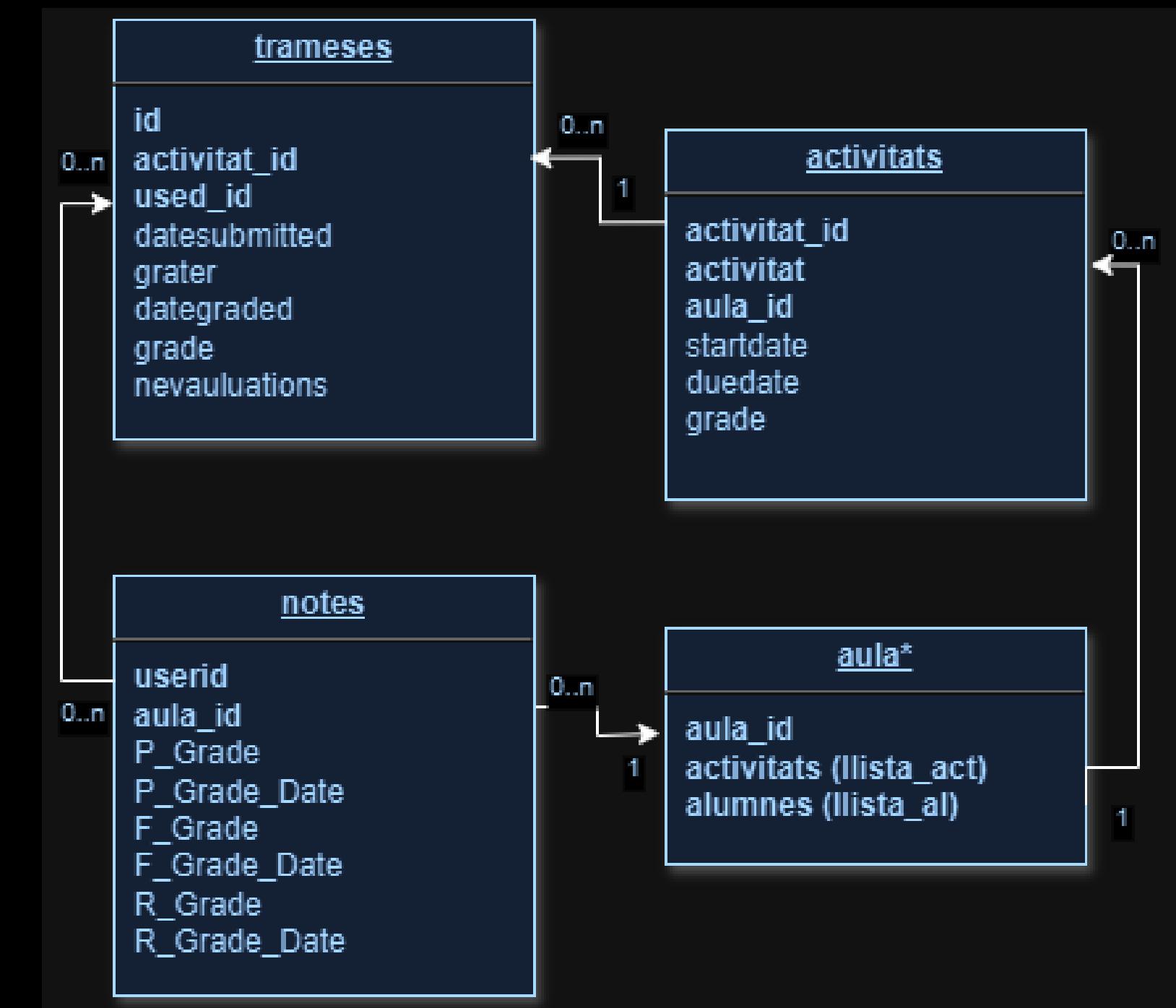
- Entendre els datasets
- Data Processing
- Predicció i Resultats
- Prototip web
- Conclusions i millores

ENTENDRE ELS DATASETS

● Detectar variables útils

- datesubmitted
- P_Grat
- F_Grate
- grade
- activitat_id
- label*

● Relacionar les taules



DATA PROCESSING

● Variables eliminades

- F_Grade_Date (NOTES)
- P_Grade_Date (NOTES)
- grader (TRAMESES)
- dategraded (TRAMESES)

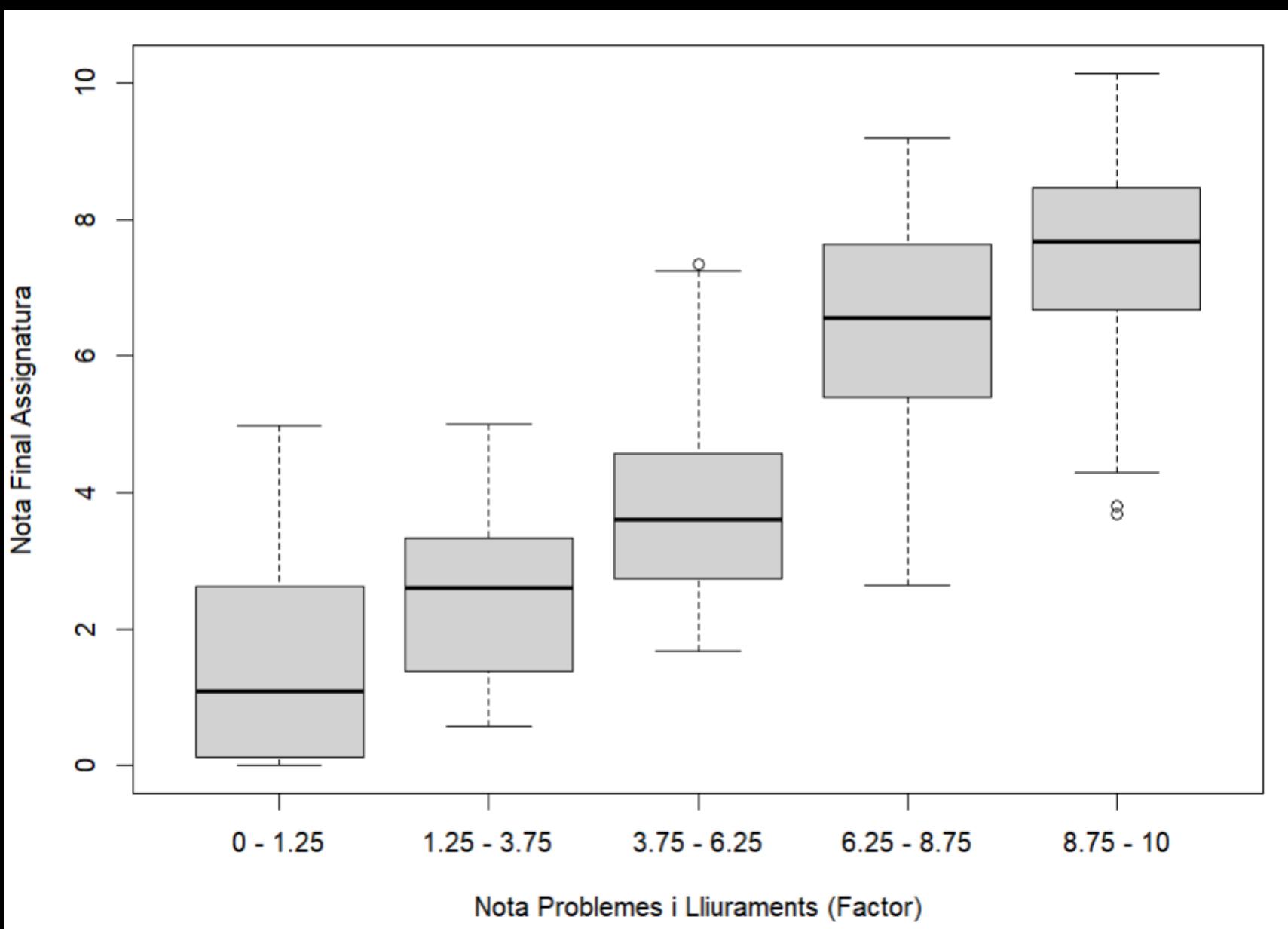
● Variables afegides

- Nota final
- Nota problemes
- Nota exàmens
- Shared ID per a les diferents activitats
- Labels

● Files eliminades

- Aula no apareix a la taula NOTES
- “grade” == NULL & “nevaluation” != 0 (a la taula TRAMESES)
- Les files duplicades a la taula TRAMESES
- Les entrades de la taula TRAMESES per a les que “userid” no apareix a la taula NOTES
- Les entrades de la taula NOTES per a les que P_Grade == NULL & F_Grade == NULL

ANÀLISI DESCRIPTIVA



x: nota problemes i lliuraments
y: nota final assignatura

$$\text{Corr}(x, y) = 0.7085$$

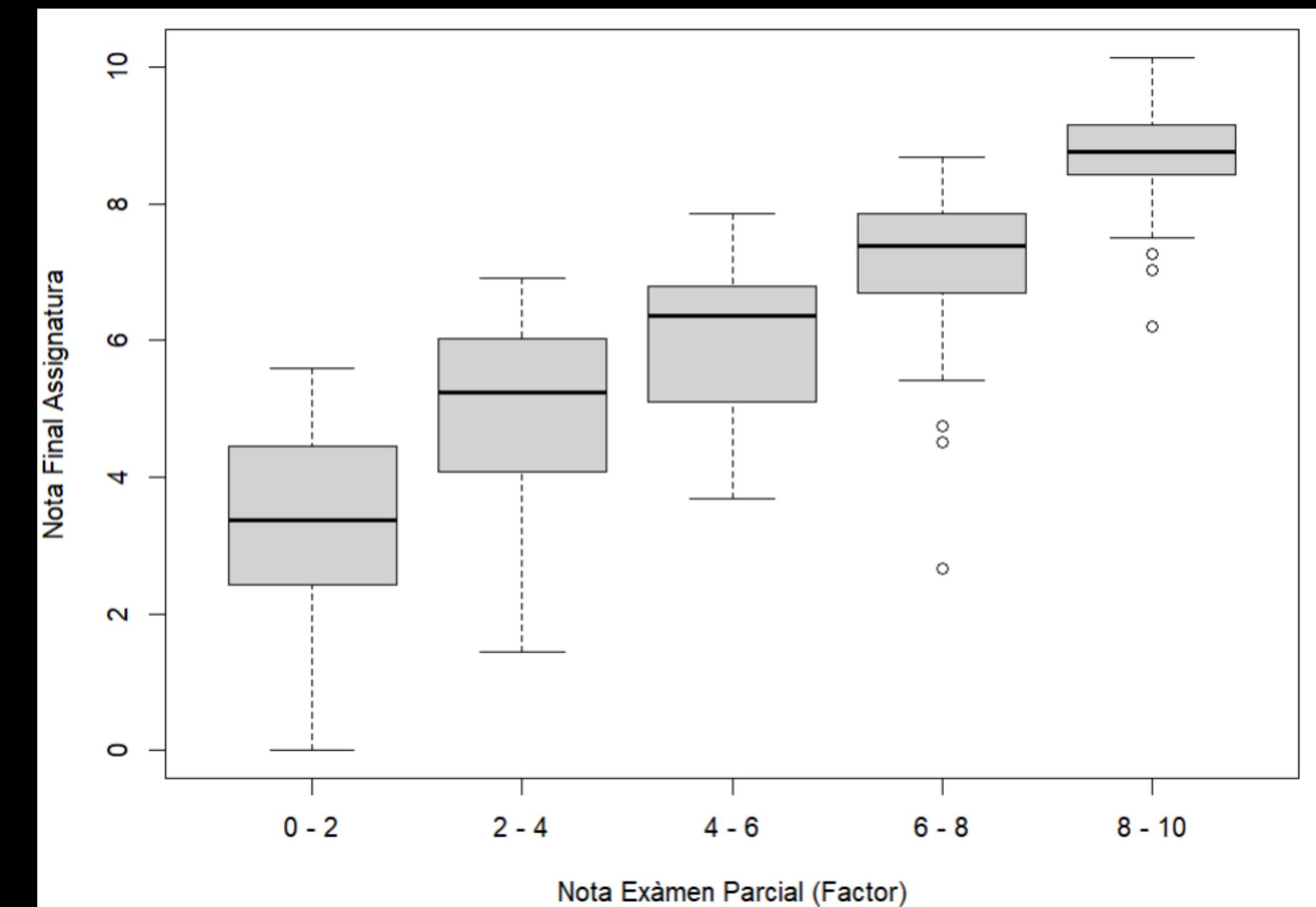
```
data: notes$notaFinal and notes$FactorNotaProblemes
  1   2   3   4
  2 0.01071 -
  3 4.7e-09 0.00017 -
  4 < 2e-16 < 2e-16 < 2e-16 -
  5 < 2e-16 < 2e-16 < 2e-16 7.7e-15
P value adjustment method: holm
```

ANÀLISI DESCRIPTIVA

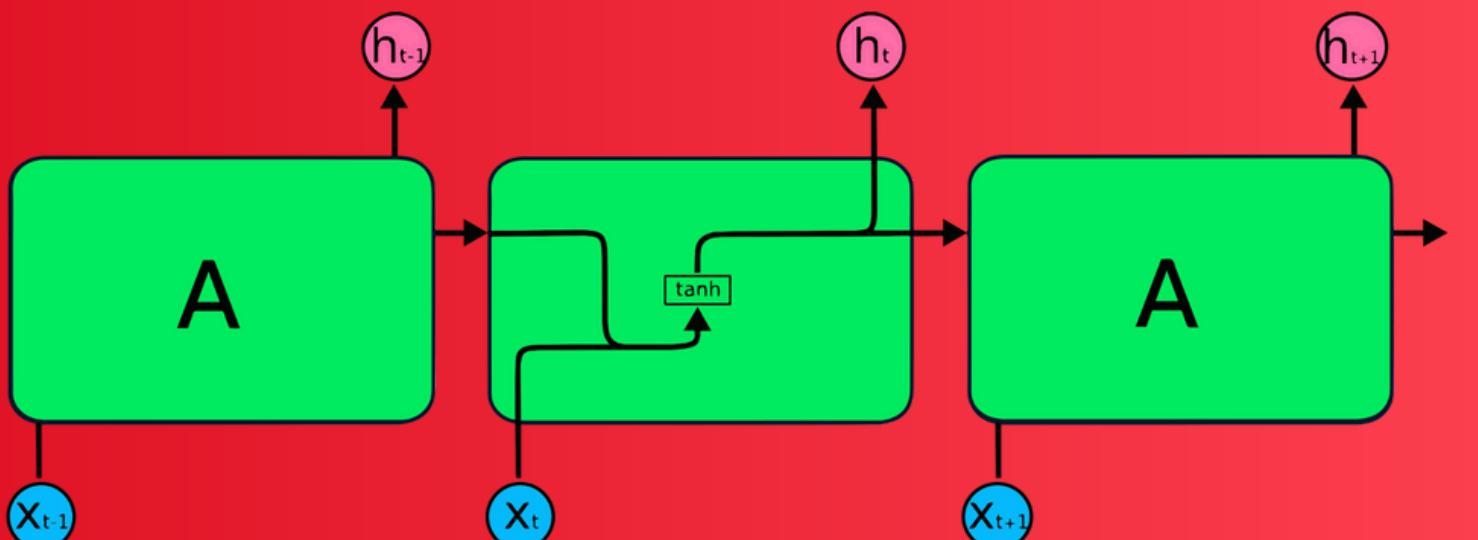
x: nota exàmen parcial
y: nota final assignatura

$$\text{Corr}(x, y) = 0.8603$$

```
data: notesHiHaParcial$notaFinal and notesHiHaPa
      0 - 2   2 - 4   4 - 6   6 - 8
2 - 4  5.8e-09  -     -     -
4 - 6  < 2e-16  2.8e-05  -     -
6 - 8  < 2e-16  1.5e-15  2.8e-05  -
8 - 10 < 2e-16  < 2e-16  < 2e-16  5.5e-11
P value adjustment method: holm
```



PREDICCIÓ

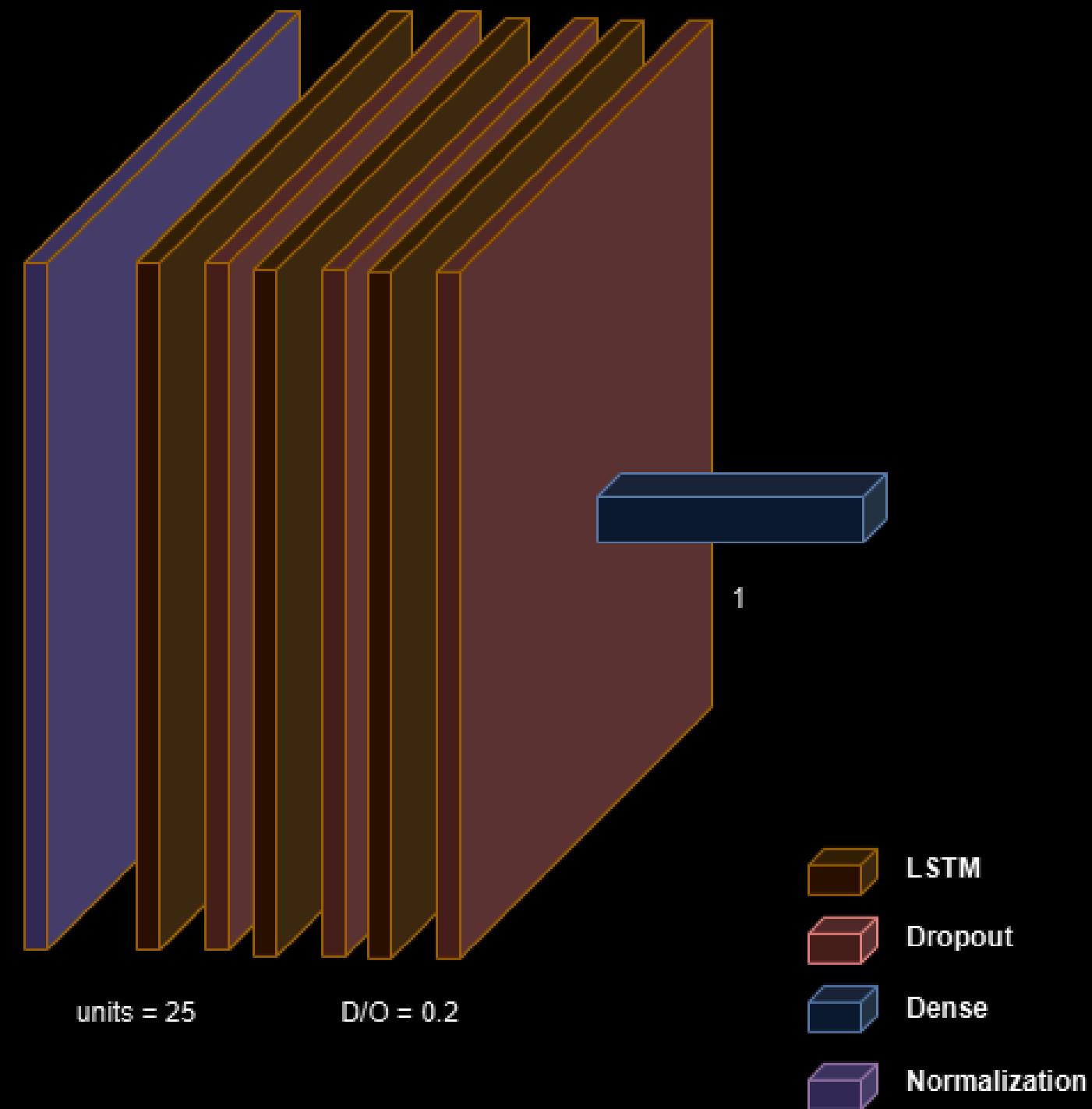


RNN

PREDICCIÓ

● Explicació model

1. Train, test
2. Padding
3. Sequential*
4. Optimizer i loss
5. Fit



RESULTATS

Epochs

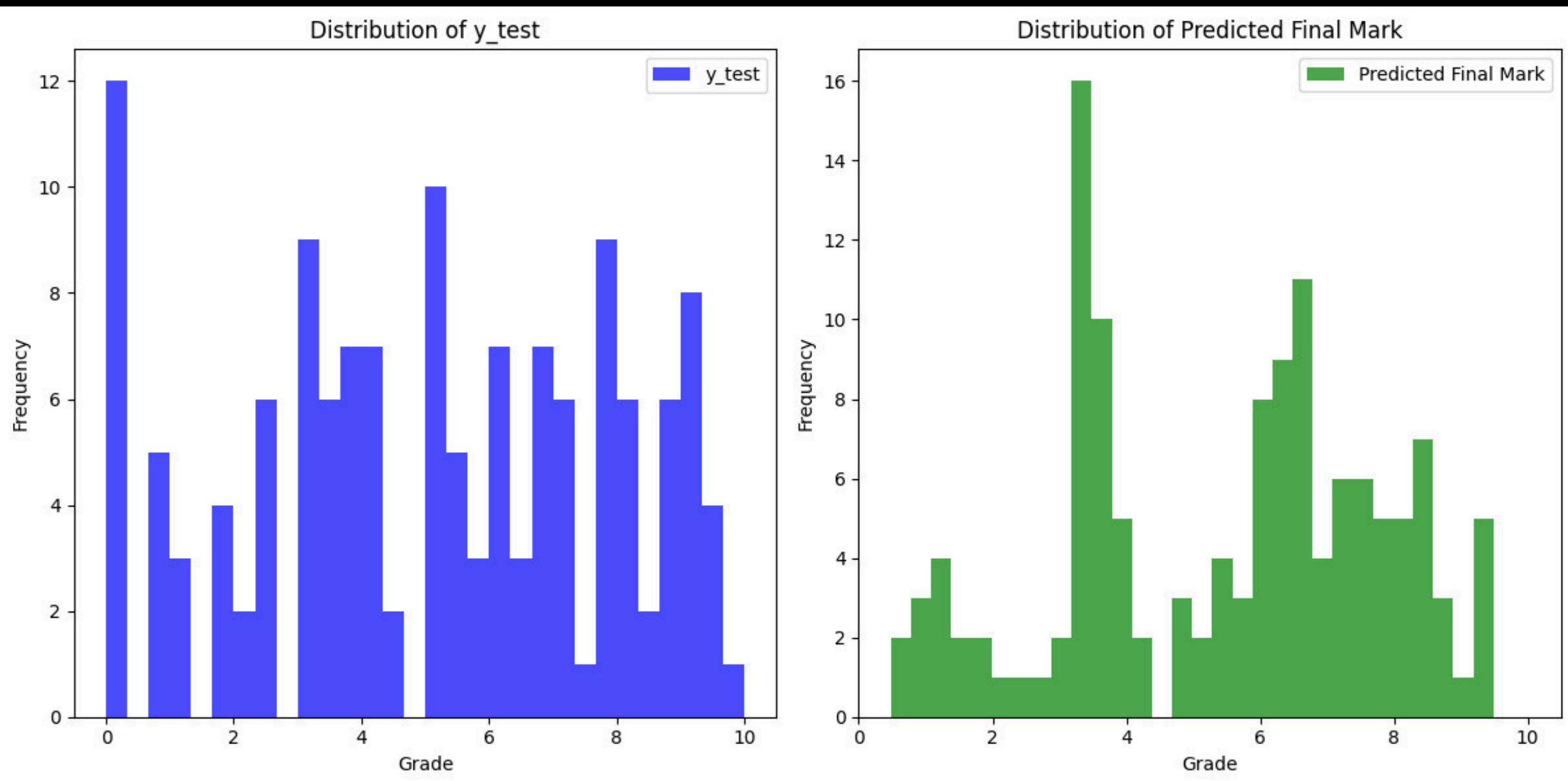
Millor de 1000

Loss

RMSE = 2.225

Learning rate

lr = 0.001



PROTOTIP WEB



MILLORES

- Importància d'una bona base de dades
- Probar models sense historial -> KNN
- SSM Mamba / Regresió Lineal
- Reduir Overfitting
- Data augmentation

RESUM

Comprendre els datasets
(Disseny BD)



Procesar les dades
(Cleaning)



Predictió (RNN)



Anàlisi Descriptiva
(Correlació)

CARONTE CHALLENGE

Datathon 2024

Arnau Busquets, David Franch, Marc Ugas, Jordi Ventosa