

# SÓN LES XARXES NEURONALS MÉS EFICIENTS QUE ELS ARBRES DE DECISIÓ?

Anàlisi sobre un problema d'Astronomia

Albert Ribé Costa

Desembre 2016

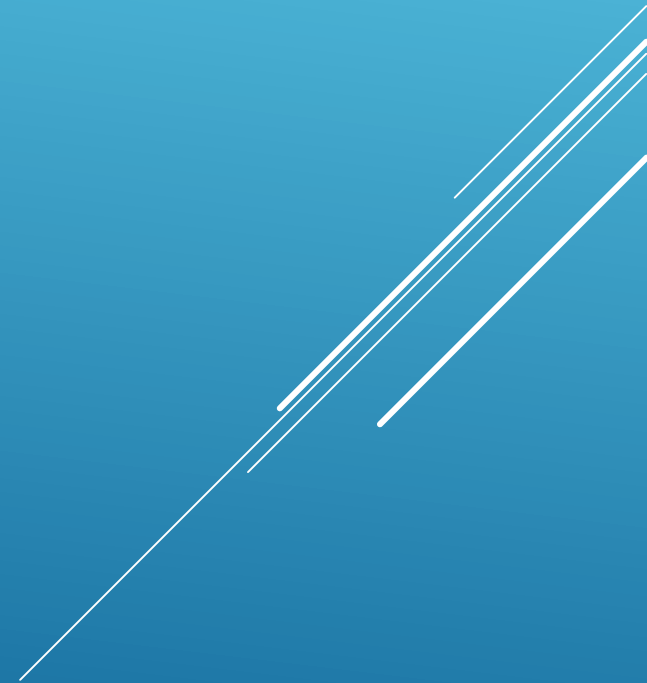
Objectius

Projecte SDSS


Solució tècnica

Anàlisis de resultats

Conclusió



## ► Objectius

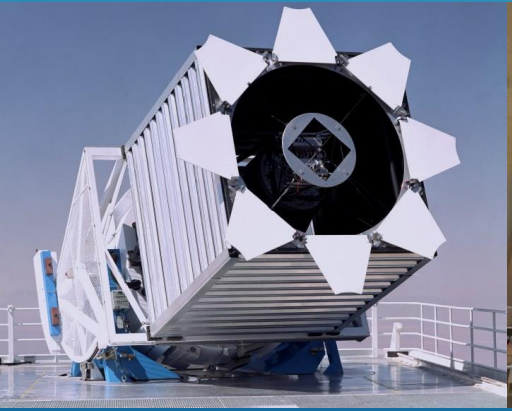
- Selecció de dades
  - Desenvolupar un DT i una FNN
  - Analitzar el funcionament d'ambdós
  - Determinar la solució idònia per al problema
  - Realitzar una formació en Deep Learning i posar en pràctica els conceptes apresos al llarg del Màster
- 

► Projecte SDSS



“The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects. Learn and explore all phases and surveys—past, present, and future—of the SDSS.”

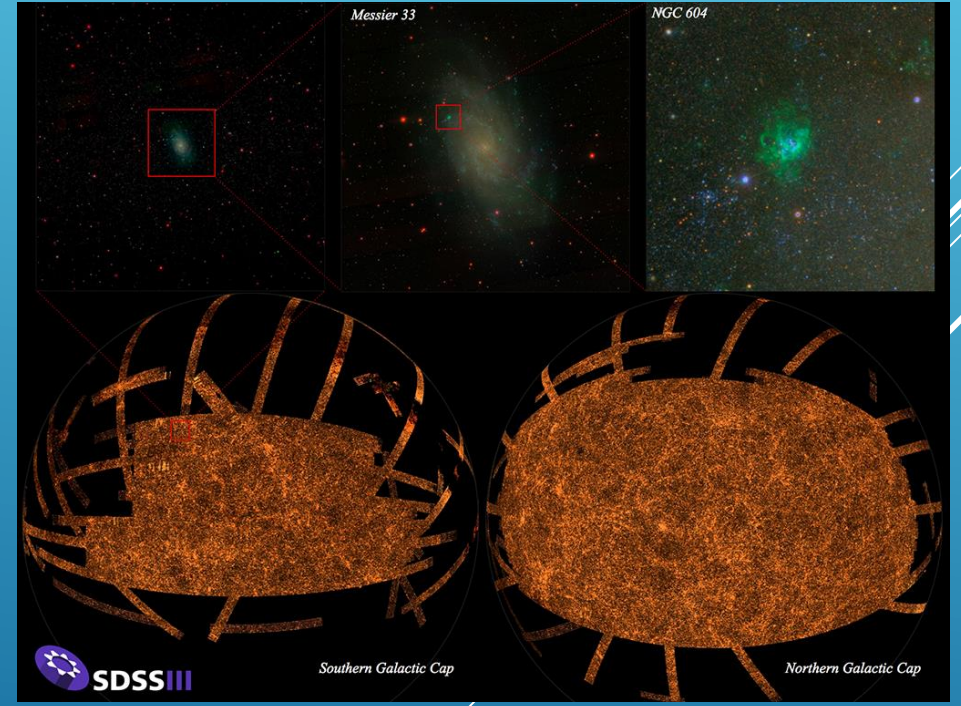
SDSS utilitza càmeres CCD de diferents observatoris, principalment el de Apache Point a nou Mèxic i l'observatori Las Campanas a Xile



Apache Point , Nou Mèxic

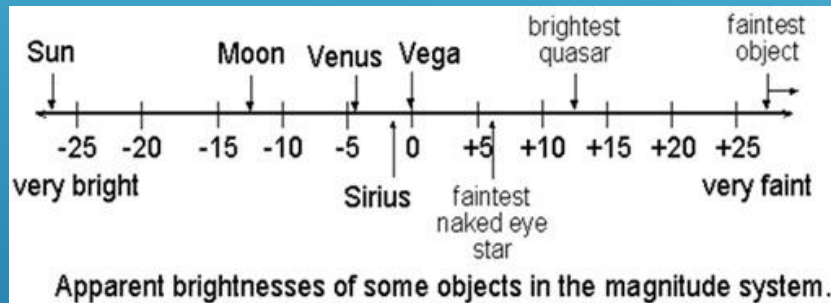


Las Campanas,Xile

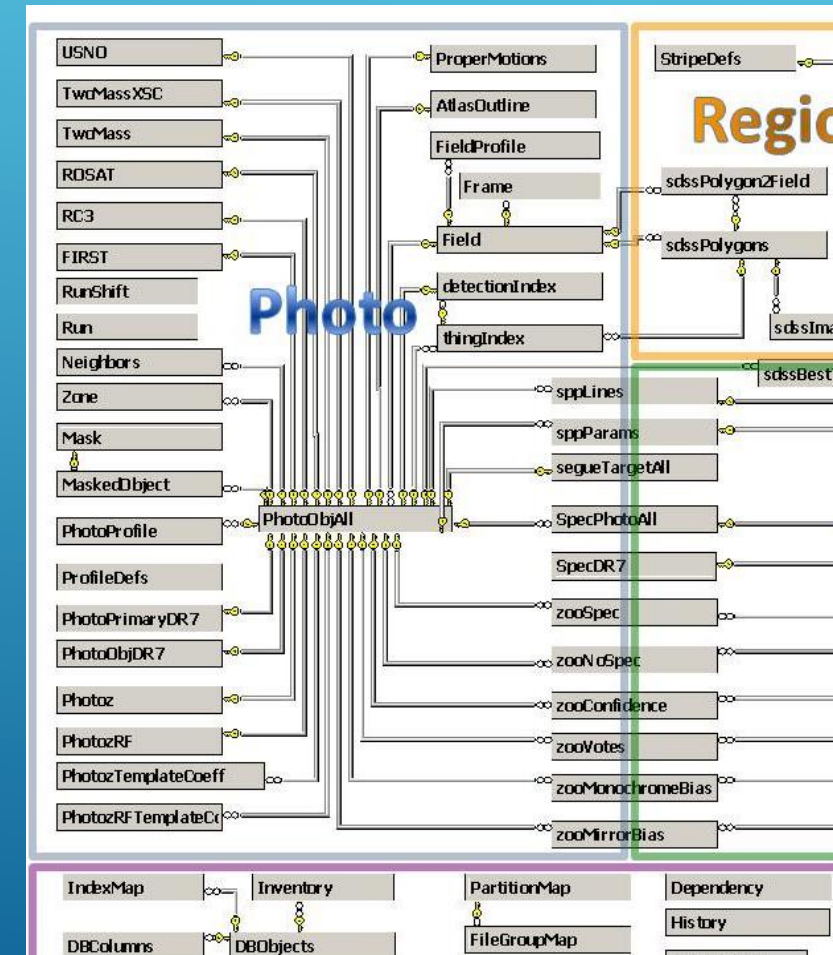


## ► Catàleg fotomètric

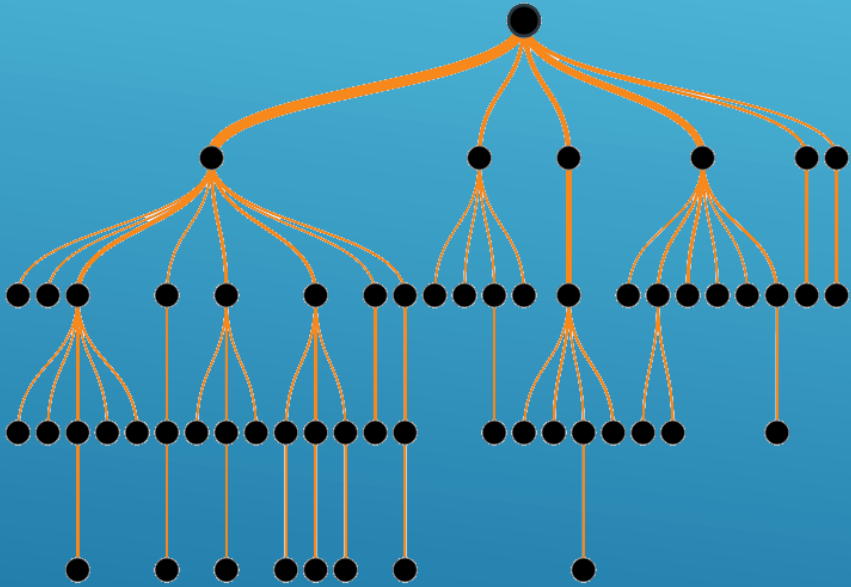
- La fotometria mesura la brillantor dels objectes celestes
- Hiparc de Nicea (190 a.C. – 120 a.C.) va establir les primeres escales de brillantors



- El catàleg fotomètric d'SDSS té a la seva taula principal 1231 milions d'elements
- La taula PhotoPrimary en té 470 milions

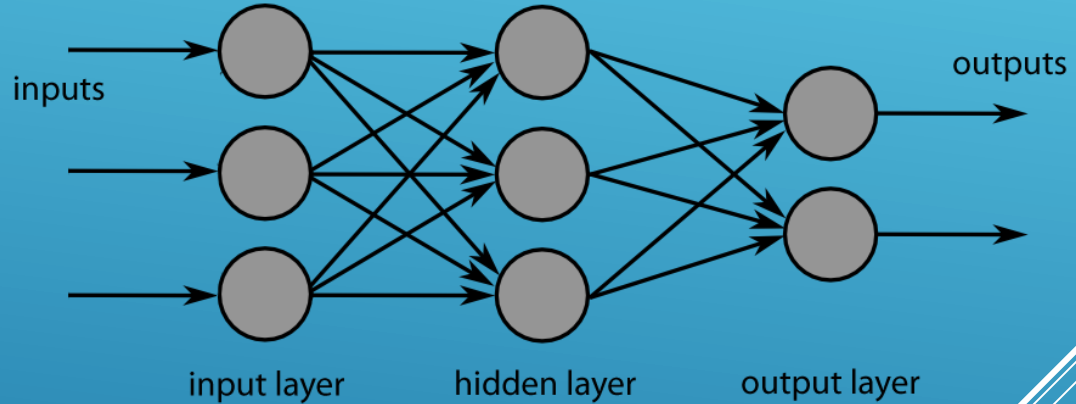


# Arbre de decisió



- De fàcil comprensió
- Poden treballar amb valors discrets i continus
- És molt flexible, es poden augmentar els nodes per afrontar decisions mes complexes

# Feed Forward Neural Network

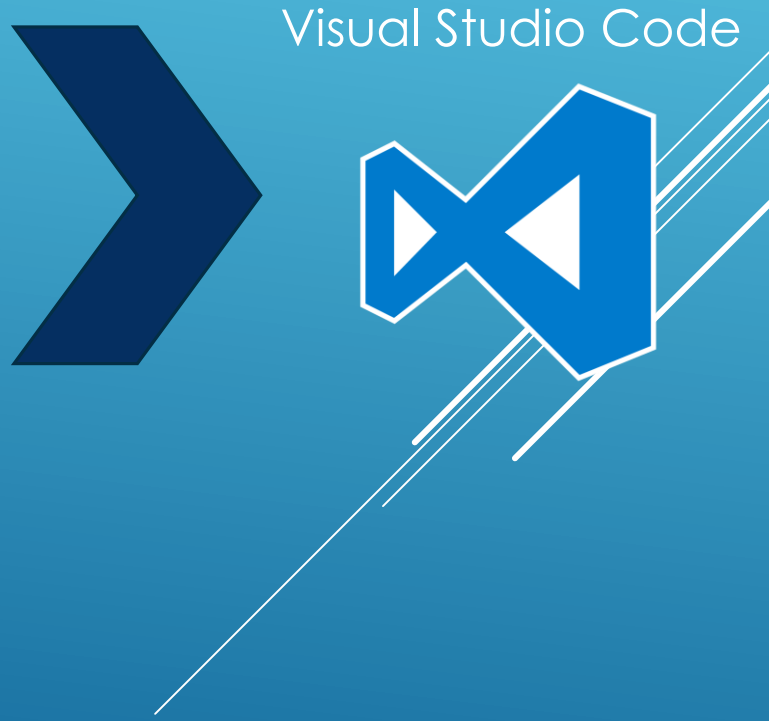


- Intenta imitar el funcionament del cervell
- Les neurones o perceptrons, s'organitzen en capes
- Les capes entre l'entrada i la sortida no tenen connexió amb l'exterior
- No hi ha connexió entre perceptrons de la mateixa capa

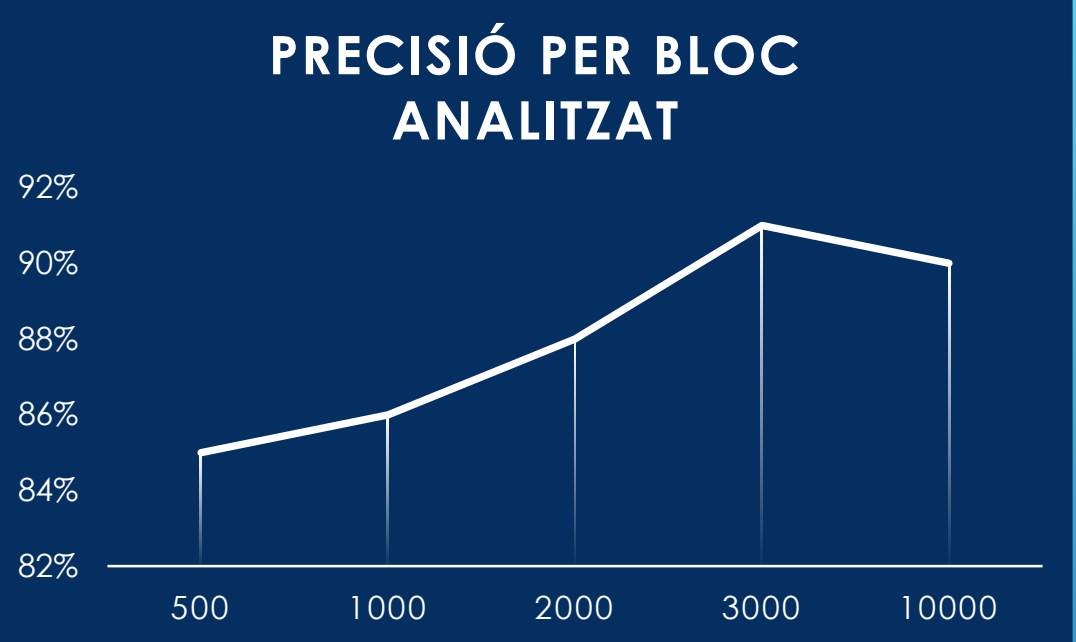


► Solució tècnica

Tecnologia		On s'utilitza
Python		DT / FNN
Anaconda		DT / FNN
Theano		FNN
Lasagne		FNN
Astropy		DT / FNN

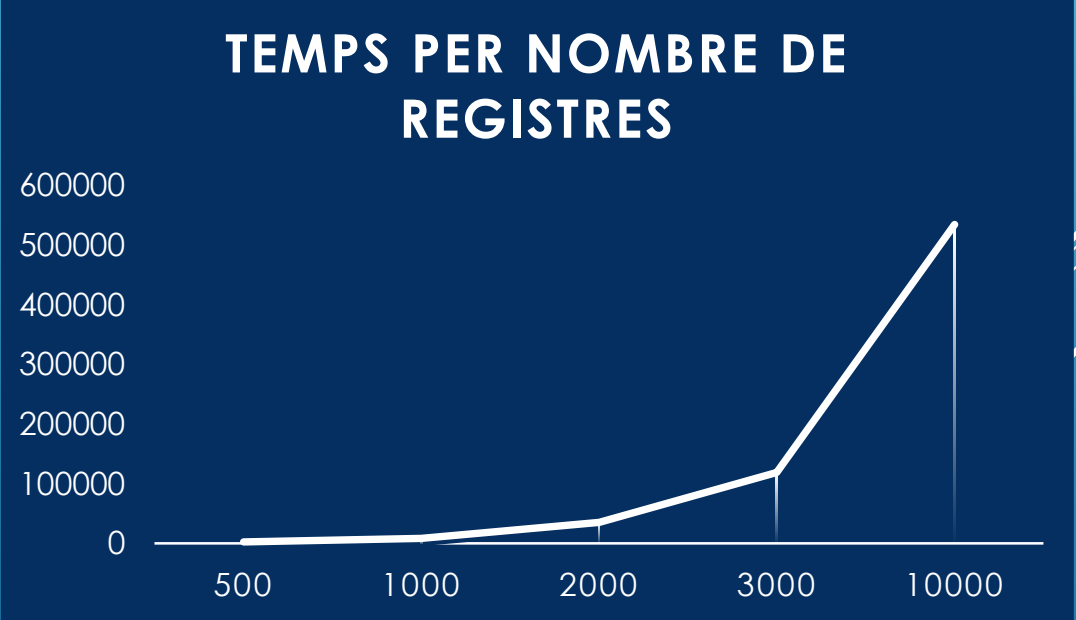


► Anàlisi de resultats



- 3 capes ocultes
- 10.000 registres
- Més de 6 dies d'entrenament
- Precisió obtinguda: 90%

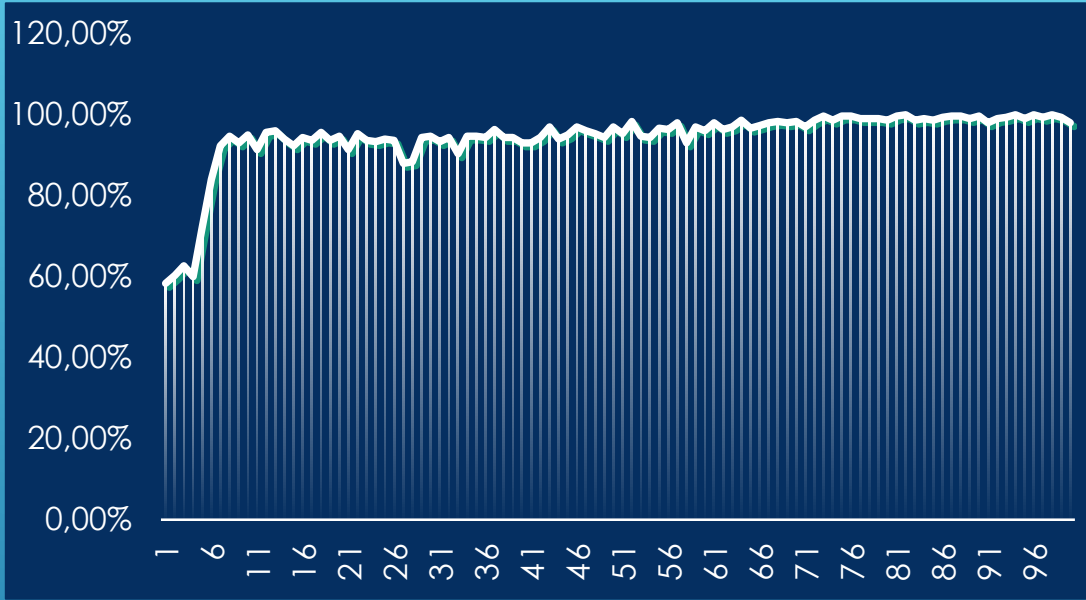
- 3 capes ocultes
- 3.000 registres
- 33 hores de temps d'entrenament
- Precisió obtinguda: 91%



# DECISION TREE

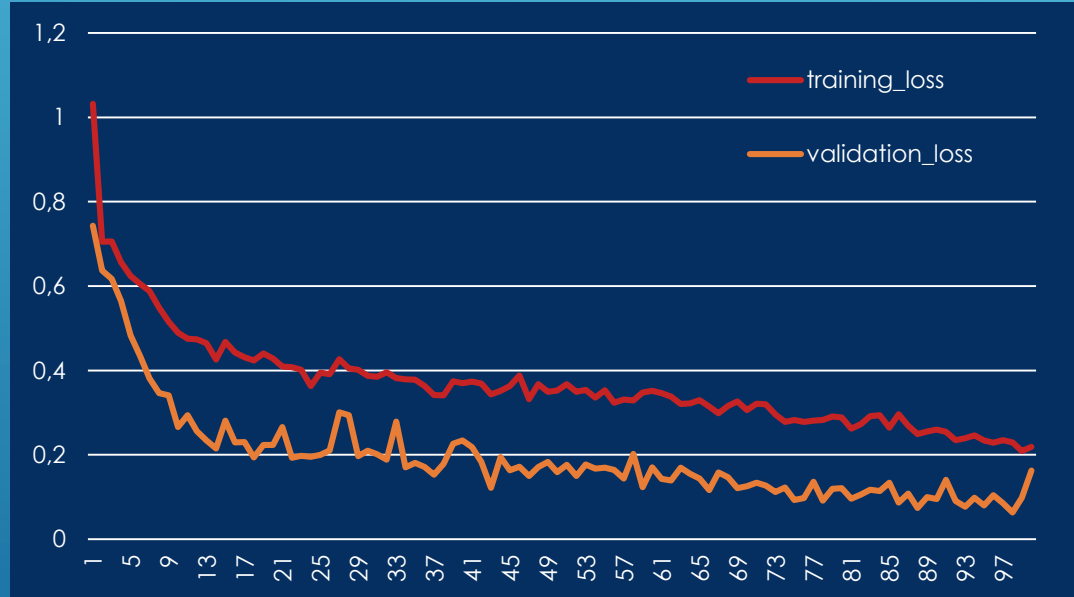


## ► Anàlisi de resultats



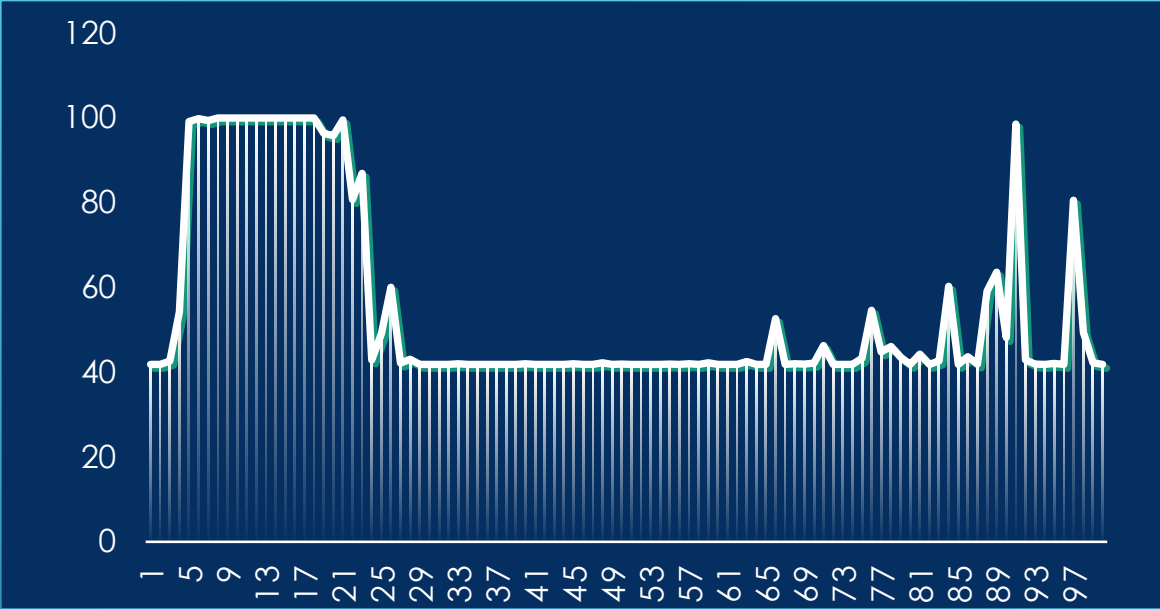
- L'Overfitting a partir de la iteració 220
- Precisió màxima quasi no augmenta

- 3.000 registres
- 100 iteracions
- Xarxa configurada per maximitzar la precisió
- Màxima precisió per iteració 100%
- Precisió en test 97,33%
- 2 hores de temps d'entrenament



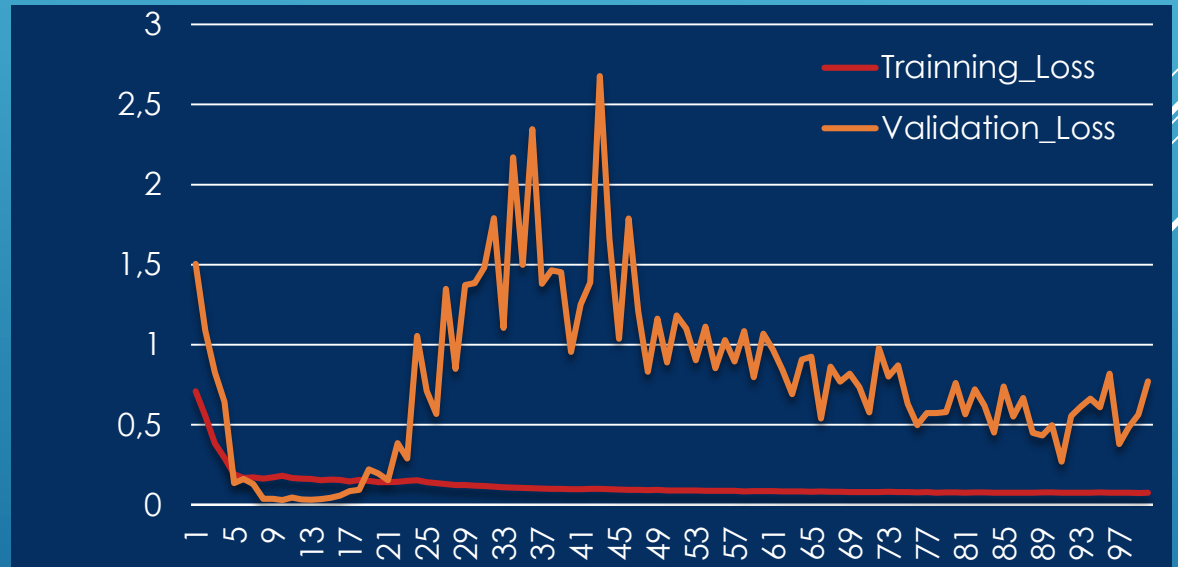
# FEED FORWARD NEURAL NETWORK

## ► Anàlisi de resultats



- Amb 10 iteracions s'obté una precisió de 99.99% en el test
- 1,5 hores de temps d'entrenament

- 200.000 registres
- 100 iteracions
- Xarxa configurada per maximitzar la precisió
- Màxima precisió per iteració 99,99%
- Precisió en test 40%
- Overfitting
- 17 hores de temps d'entrenament



## ▶ Conclusions

### ▶ La FNN aconseguix millor precisió en menys temps

- ▶ Elevada quantitat de valors de configuració de la FNN
- ▶ Lasagne tot just esta a la versió 0.1, encara hi ha poca documentació i comunitat, els errors són poc clars
- ▶ Els tipus Galàxia i Planeta suposen el 99.9 % de les dades sobre el data set complet

## ▶ Treballs futurs

- ▶ Anàlisi de FNN amb GPU
- ▶ Proves amb el Dataset complet, ara s'ha treballat amb el 6%.
- ▶ Intentar millorar la precisió de l'arbre realitzant poda
- ▶ Obtenir l'anàlisi expert de les dades per poder eliminar de l'estudi algunes de les 500 columnes o afegir alguna taula relacionada