November 2023

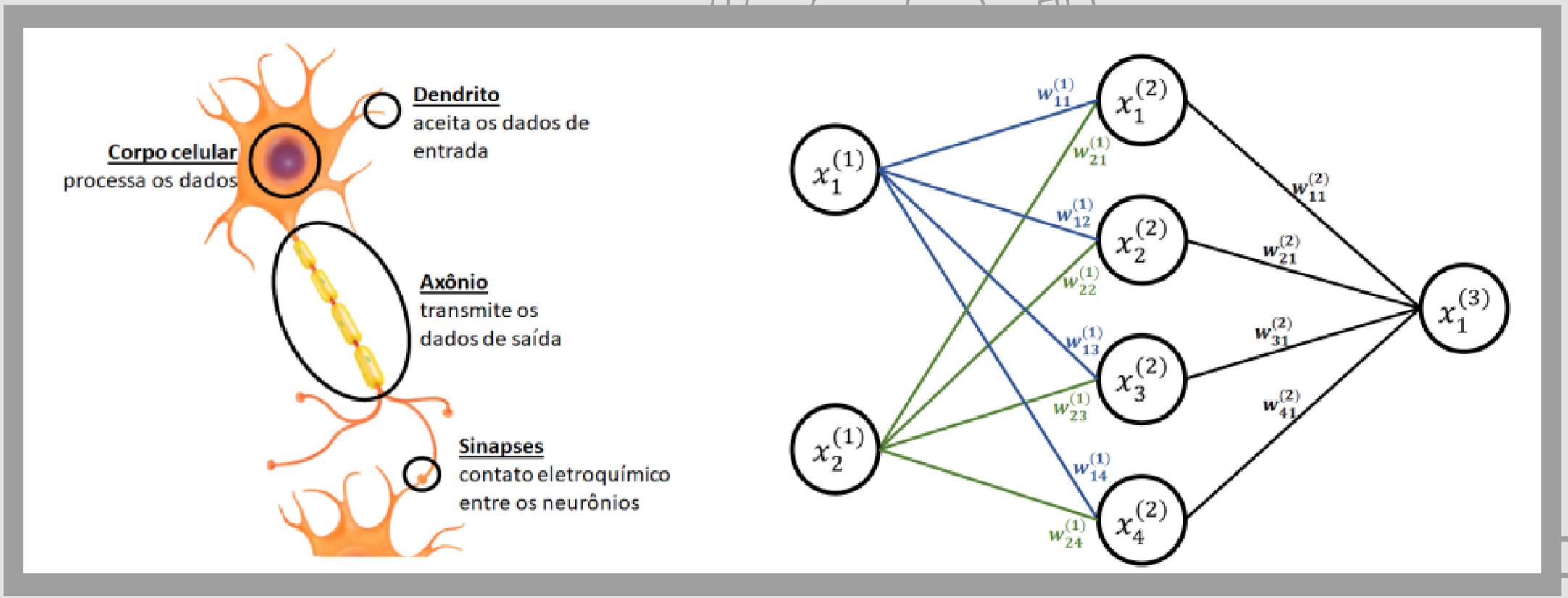
Analysis of photometric data obtained through machine learning and K-d Tree

Masters dissertation - Amanda Farias

Alex Sander do Carmo - Physics undergraduate student
Raphael Pereira Rolim e Silva - Physics undergraduate student
Prof. Dr. Elcio Abdalla

Summary

1.0 - Machine Learning



- Learning model: Supervised machine
- Training set
- Test Set
- Validation set

2.0 - Data Preprocessing

VIPERS

- z ~ 0,5 to 1,2
- 47646 galaxies
- Converted pixels to align with DES

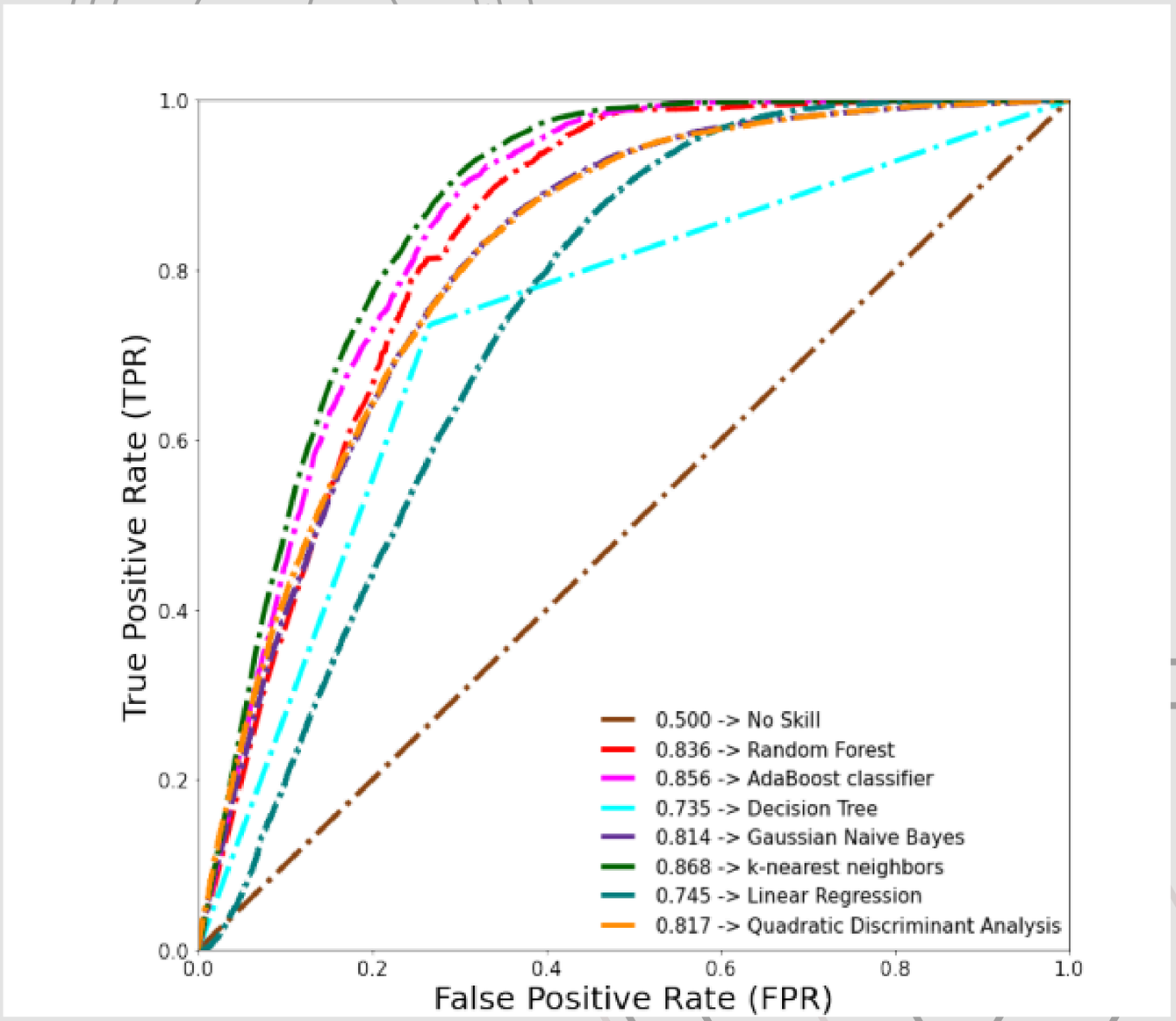
DES

- Galaxies on VIPERS (1) or not (0)
- RA and DEC to refine data
- 0.19% of DES catalogue downloaded

- Equal quantity of galaxies 1 and 0

3.0 - K-d Tree code and results

- K = 4 (colors) and K = 5 (colors - magnitude)
- 4096 galaxy group - 12 layers
- Probability of galaxies being photoelectrically similar
- Training set: 65% of the galaxies
- Test set: 25% of the galaxies
- Validation set: 10% of the galaxies
- Best model - k-nearest neighbors



4.0 - GPz and ANNz

GPz

- Phot. redshift from gaussian processes
- Input: colors and magnitude errors
- GPz - Magnitude and GPz - Color
- Less errors with GPz - Color

ANNz

- Phot. redshift from artificial neural network
- Smaller mean quadratic error

$$\mathrm{MQE}(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y)^2$$

- Better performance at $0.5 \leq z \leq 1.0$

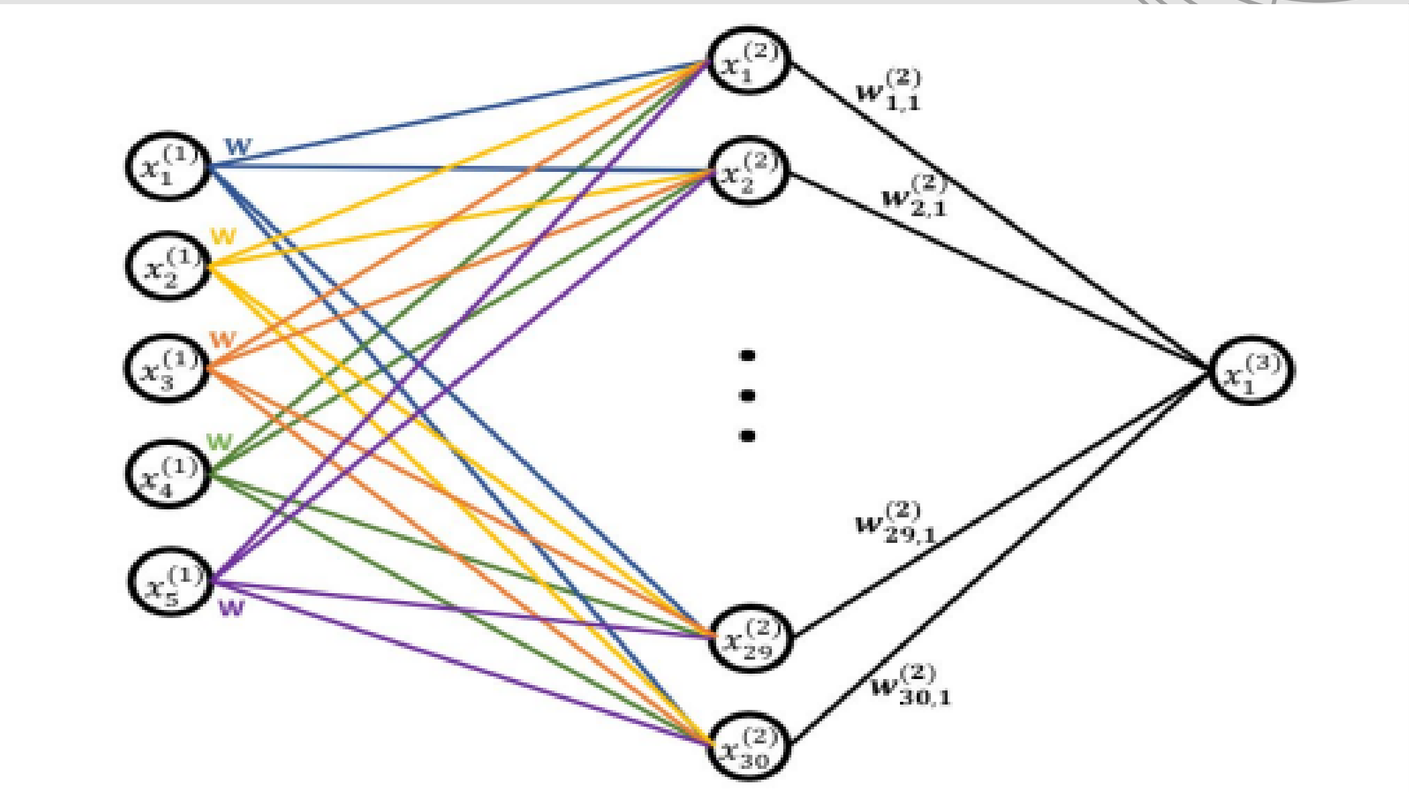
4.0 - Keras

Node Test

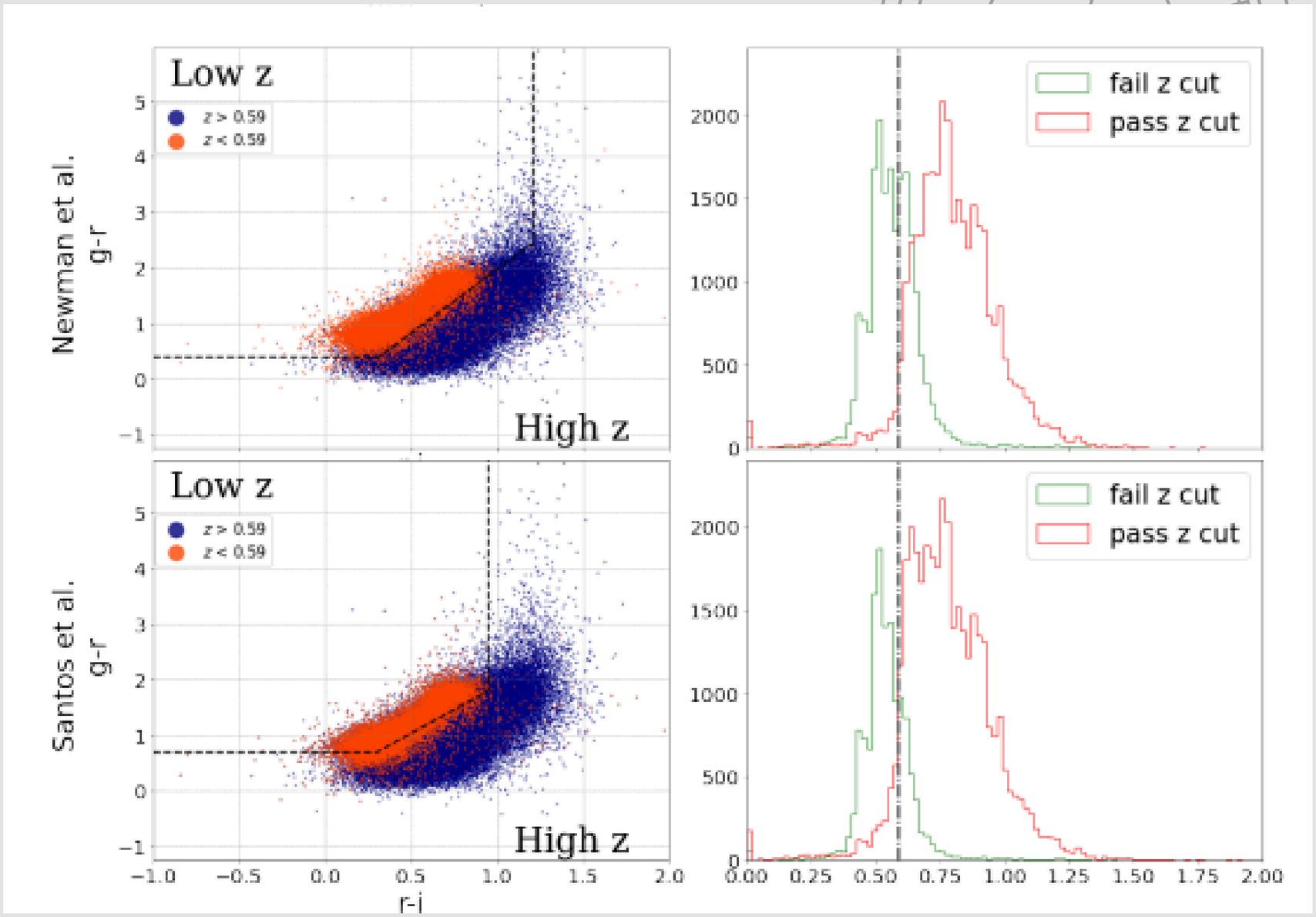
- Error-based evolution (MQE)
- An inner layer with 30 nodes

RESULTS

- 10 different machines
- Final result: Average value of the machines



5.0 - Comparision of Results



Redshift classification acurracy:

- Newman et al. 82,39%
- Santos et al. 87,83%

Redshift cut analysis

- z > 45 (higher redshift)
- z < 45 (lower redshift)

6.0 - Conclusions

- All models performed well;
- All analyzes returned similars values;
- The λ parameter, related to probability, was well adjusted;
- Keras is a machine learning as good as the existing ones.

November 2023

THANK YOU