Open University of Catalonia (UOC) Master's Degree in Data Science

MASTER'S THESIS

Area: Data Science

Density-based algorithms applied to galaxy groups catalogs

-

Author: Carlos Toro Peñas

Tutor: Laura Ruíz Dern

Professor: -

Madrid, October 1, 2025

Credits/Copyright

A page with the specification of credits/copyright for the project (either application on one side and documentation on the other, or unified), as well as the use of third-party trademarks, products or services (including source code). If a person other than the author collaborated on the project, their identity and what they did must be explicitly stated.

Below is the most common case, but it can be modified for any other alternative:

FINAL PROJECT RECORD

Title of the project:	Density-based algorithms applied to galaxy groups catalogs
Author's name:	Carlos Toro Peñas
Collaborating teacher's name:	Laura Ruíz Dern
PRA's name:	First and last name
Delivery date (mm/yyyy):	MM/YYYY
Degree or program:	Master's degree in Data Sicience
Final Project area:	4
Language of the project:	English
Keywords:	Clustering, Galaxy groups, cosmology

Dedication/Quote

To my wife, to whom this work owes more than she imagines.

Abstract

This work focuses primary on the application of density-based algorithms to datasets obtained from various surveys, such as the Two-degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS). As a result of this application, a performance evaluation will be conducted to identify the strengths and weaknesses of these algorithms for galactic cluster detection. In the future, these algorithms may be applied to new surveys and other regions of the sky.

Text with a summary of the project, that is, a concise explanation of the project/problem addressed, its objectives/resolution methods, and the results and conclusions (it cannot be a list, but rather a continuous text written in a structured way). If a reference is necessary in this text, it will be noted at the bottom of the same page. In this section, a more literary and colloquial language can be used than for the rest of the document.

The Abstract will be written twice. One version must be **obligatorily in English**. The other version must be written in Catalan or Spanish. If the rest of the document is not written in English, it will be necessary to write the second version of the Abstract in the language used for the rest of the report. The word Abstract will be changed to **Resum**" or **Resumen**" in the Catalan and Spanish versions, respectively.

Recommended length: maximum 250 words.

How to write a good Abstract:

http://www.ece.cmu.edu/ koopman/essays/abstract.html

Keywords: Clustering, Galaxy groups, cosmology.

Resumen

Este trabajo tiene como tema central la aplicación de algoritmos basados en densidad a juegos de datos obtenidos en diferentes estudios como 2DFGRS y SDSS. Como resultado de esa aplicación, se hará una evaluación del desempeño encontrando fortalezas y debilidades de tales algoritmos para la detección de cúmulos galácticos. Futuramente, se podrán realizar aplicaciones de estos algoritmos a nuevos estudios y otras regiones del cielo.

Palabras clave: Clusterización, cúmulos de galaxias, cosmología.

Contents

Abstract vii

Resumen ix

Table of Contents xi

List of Figures xiii

List of Tables 1

1 Introduction . . . 1
- 1.1 Context and motivation . . . 1
- 1.2 Goals . . . 1
- 1.3 Sustainability, diversity, and ethical/social challenges . . . 2
- 1.4 Approach and methodology . . . 2
- 1.5 Schedule . . . 2
- 1.6 Summary of the outputs of the project . . . 3
- 1.7 Brief description of the remaining chapters of the report . . . 3

2 Methods and resources . . . 3

3 Results . . . 3

4 Conclusions and future work . . . 3

5 Glossary . . . 4

Bibliography 4

6 Appendices . . . 5

List of Figures

1 Sample figure. It will be indexed in the “List of Figures”. 1

List of Tables

1 Introduction

This template aims to be a guide for students. This template can be adapted to the specific needs of each project if the project supervisor agrees with the changes.

Figure 1: Sample figure. It will be indexed in the "List of Figures".

1.1 Context and motivation

Starting point of the project (Which is the problem that needs to be solved? Why is it a relevant topic? How is the problem currently being solved?) and description of the contribution (What is the intended outcome?)

1.2 Goals

List of the goals to be achieved in this project.

1.3 Sustainability, diversity, and ethical/social challenges

This section should assess the positive/negative impact of the project in the following dimensions. It is not required to reach a positive impact in any/all dimensions, but it is necessary to consider and discuss whether there is an impact or not from the beginning of the project.

Sustainability In the development of the project or during its entire lifecycle (e.g., deployment, retirement), does the output of this project have an impact on sustainability and/or ecological footprint (energy/resource consumption/savings, waste, pollution, depletion of raw materials)? Is it affected by laws or regulations on this matter? Considering another perspective, does it affect any of the Sustainable Development Goals (SDG) related to these dimensions? If it does not have any impact, either positive or negative, you should explain how you reached this conclusion and justify your answer.

Ethical behaviour and social responsibility Is the outcome of the project too technical to have any positive/negative impact in ethical/social aspects? Does it have an impact on laws/regulations (data, privacy, labour, intellectual property, personal security, ...)? Does it adhere to the deontological principles of the profession? Does it endanger/improve/worsen any job position? If it does not have any impact, either positive or negative, you should explain how you reached this conclusion and justify your answer.

Diversity, gender and human rights Is the result of this project so technical that it has no positive/negative impact in terms of gender, diversity, or human rights? And in any laws/regulations? And in terms of accessibility, disability, ergonomics and/or information security? If it does not have any impact, either positive or negative, you should explain how you reached this conclusion and justify your answer.

1.4 Approach and methodology

Describe the potential strategies to develop this project and explain the selected strategy. Discuss why the selected strategy is the most suitable one to achieve the project's goals.

1.5 Schedule

Description of the resources needed to develop the project, the tasks to be carried out and a schedule for each task using a Gantt diagram or equivalent. This planning should define the milestones that will be completed in each Continuous Assessment Test.

1.6 Summary of the outputs of the project

It is not necessary to describe each output in detail: this will be done in the remaining chapters of the project.

1.7 Brief description of the remaining chapters of the report

Brief description of the contents of each chapter and their relationship with the rest of the project.

2 Methods and resources

In these sections, it is necessary to describe:

- The most relevant aspects of the design and development of the project.
- The methodology used in the development process, describing the potential alternatives, the decisions that have been taken and the criteria used to take these decisions.
- A description of the products that have been created.

The structure of these sections may change according to the type of project being developed.

3 Results

Describe the results achieved using the previously described methodology.

4 Conclusions and future work

This section should include the following: Citando a longair [1]

- A description of the conclusions of the work.
- A critical assessment of the degree of achievement of the initial goals.
- A critical assessment of the schedule and methodology used in the project.
- Considering the sustainability, diversity and ethical-social challenges linked to the project.
- A discussion of topics for potential future work that have not been explored in this project.

5 Glossary

Definition of the most relevant terms and acronyms used in this report. Random citation [2] embeddeed in text.

Bibliography

[1] Longair S. Malcom. (1996). *Our Evolving Universe*. United Kindom, UK.

[2] Yongda Zhu.(2025) Tsutomu T. Takeuchi1, Suchetha Cooray. soptics: A modified density-based algorithm for identifying galaxy groups/clusters and brightest cluster galaxies. *ArchivX*.

6 Appendices

List of sections that are too long to be included in the body of the report and that are self-contained.