

# Título

Antonio Molner Domenech

Trabajo de Fin de Grado  
Ingeniería Informática

Supervisado por:  
Alberto Guillén



**UNIVERSIDAD  
DE GRANADA**

Universidad de Granada, España  
Junio 2020

# Índice general

|  |             |
|--|-------------|
| Listado de figuras   | III         |
| Listado de tablas  | IV          |
| <b>1. Objetivos</b>  | <b>v</b>    |
| 1.1. Alcance de los objetivos . . . . .                            | VI          |
| <b>2. Introducción</b>   | <b>vii</b>  |
| 2.1. Herramientas para el análisis de rayos cósmicos . . . . .     | VII         |
| <b>3. Planificación del trabajo</b>                                | <b>viii</b> |
| <b>4. Presupuesto</b>  | <b>ix</b>   |
| <b>5. Diseño del marco de trabajo</b>                              | <b>x</b>    |
| <b>6. Diseño del autoencoder</b>                                   | <b>xi</b>   |
| <b>7. Experimentos</b>   | <b>xii</b>  |
| 7.1. Cuantificación del ahorro del tiempo de desarrollo (opcional) | XII         |
| 7.2. Resultados . . . . .  | XII         |
| <b>8. Anexo: Manual de Usuario</b>                                 | <b>xiii</b> |
| <b>9. Referencias</b>  | <b>xiv</b>  |

# Listado de figuras

|  |    |
|--|----|
| Figure 4.1 This is an example figure . . . | pp |
| Figure x.x Short title of the figure . . . | pp |

# Listado de tablas

|   |    |
|---|----|
| Table 5.1 This is an example table . . .  | pp |
| Table x.x Short title of the figure . . . | pp |

# Capítulo 1

## Objetivos

El objetivo de este proyecto es el de desarrollar un marco de trabajo para machine learning enfocado en la reproducibilidad y buenas prácticas que explicaremos más adelante. Por otro lado, como objetivo secundario tenemos la aplicación de dicho framework para resolver un problema real.

A modo de resumen, los principales objetivos son:

- Diseño e implementación de un framework de reproducibilidad: El desarrollo de una herramienta que permita instrumentalizar proyectos de Machine Learning con mínimo esfuerzo, orientada a mantener unas buenas prácticas de desarrollo y seguir una filosofía MLOps. Dentro de este objetivo, de manera secundaria, incluimos una contribución de código a uno de los proyectos de código libre que componen el módulo central de nuestra herramienta, Mlflow.
- Especificación de buenas prácticas: La creación de una lista de pautas y requisitos necesarios para hacer reproducible un proyecto. Desde la recolección de datos hasta la gestión de experimentos.
- Aplicación de la herramienta a la resolución de un problema real: Este objetivo está orientado a la experimentación, trata de la aplicación de diferentes técnicas de Machine Learning tradicional y Deep learning para la resolución de un problema común en física, la detección de

primarios. En dicha aplicación, hacemos un uso extensivo de la herramienta y valoramos los beneficios y el coste en recursos humanos y capitales de su uso para este caso concreto.

## 1.1. Alcance de los objetivos

Para el primer objetivo, el alcance incluye el desarrollo integral de una herramienta en Python que permita cumplir con la mayoría de requisitos que consideramos necesarios para que un proyecto sea reproducible fácilmente por la comunidad científica. Esta herramienta debe ser flexible y permitir integrarse con frameworks de Machine Learning o Deep learning existentes, así como con proyectos orientados al análisis de datos en lugar de al modelado.

En relación con el primer objetivo, se debe desarrollar una especificación de buenas prácticas basadas en problemas existentes, y con el objetivo de reducir aquella deuda técnica que concierne a este tipo de proyectos, tanto durante el desarrollo o experimentación, como en el momento de compartir el trabajo con otras personas.

# Capítulo 2

## Introducción

- Rayos cósmicos
- Remarcar problemas en la reproductibilidad

### 2.1. Herramientas para el análisis de rayos cósmicos

- ROOT Framework
- Corkiska
- CERN

# Capítulo 3

## Planificación del trabajo

- Planificación optimista
- Planificación real



# Capítulo 4

## Presupuesto

- Comparativa cluster propio vs AWS, Azure, GDC
- Coste de titulado superior (36€)

# Capítulo 5

## Diseño del marco de trabajo

- Herramientas (y el por qué)
- Comparativa herramientas para la reproductibilidad (Polyaxon y similares)

# Capítulo 6

## Diseño del autoencoder

- Que es una red neuronal
- Que es un autoencoder
- Autoencoder simple
- Autoencoder profundo
- Autoencoder variacional

# Capítulo 7

## Experimentos

7.1. Cuantificación del ahorro del tiempo de desarrollo  
(opcional)

7.2. Resultados

## Capítulo 8

### Anexo: Manual de Usuario

# Capítulo 9

## Referencias