



**UNIVERSITÀ  
DEGLI STUDI  
DI GENOVA**

Università degli Studi di Genova  
Dipartimento di Scienze Matematiche, Fisiche e Naturali del DIBRIS

# **Implementazione di una Pipeline di Machine Learning per l'Addestramento di Modelli di Object Detection basata su Faster R-CNN**

Corso di Laurea in Informatica

**Candidato:**  
Loris Prataiolo

**Relatore:**  
Prof.ssa Giovanna Guerrini

**Tutor Aziendale:**  
Dott. Luca Bianconi

Anno Accademico 2024/2025



## **Abstract**

La presente tesi si propone di mostrare la progettazione di una pipeline per l'addestramento di modelli di intelligenza artificiale su Kubeflow, utilizzando come modello di base: Faster R-CNN. l'obbiettivo di questo lavoro, è la realizzazione di un sistema, robusto e adattabile a diversi contesti, in grado di poter gestire in autonomia l'intero ciclo di addestramento di un modello di AI. Il progetto è stato sviluppato su Google Cloud Platform (GCP), utilizzando come orchestratore il sistema Google Kubernetes Engine (GKE), e Kubeflow per la realizzazione effettiva della pipeline. Il risultato è un sistema versatile, in grado di supportare diversi modelli di AI, e diverse tipologie di dataset; mostrando l'efficacia di un'architettura MLOps nel semplificare il ciclo di training di un modello, prima della fase di testing e in fine del suo deploy.



# Contents

<b>1</b>	<b>Introduzione</b>	<b>4</b>
1.1	Contesto e obiettivi del progetto . . . . .	4
<b>2</b>	<b>Tecnologie e Modelli Object Detection</b>	<b>5</b>



# Chapter 1

## Introduzione

La presente tesi è frutto del lavoro svolto durante il tirocinio curricolare presso STAM Tech S.r.l., azienda Genovese operante nel settore IT, e attiva nello sviluppo di tecnologie innovative a livello nazionale e internazionale.

### 1.1 Contesto e obiettivi del progetto

Il progetto è stato pensato per esplorare e implementare soluzioni innovative per l'integrazione di strumenti di pipelining AI alternativi ad Apache Airflow, con l'obiettivo di migliorare l'efficienza e la scalabilità dei flussi di lavoro di machine learning. Il cuore del progetto risiede nell'utilizzo di **KubeFlow**, e del suo modulo **Pipelines** per l'esecuzione e monitoraggio del processo di training dei modelli.

Gli obiettivi principali del progetto includono:

- Studio e comprensione delle funzionalità di KubeFlow e della sua integrazione con Kubernetes.
- Sviluppo di pipeline di machine learning utilizzando KubeFlow Pipelines, prendendo come base il modello Faster R-CNN.
- Studio dello strumento **MLFlow** come strumento di raccolta e monitoraggio delle metriche di training dei modelli.
- Valutazione dell'efficacia di KubeFlow rispetto ad Apache Airflow in termini di scalabilità, facilità d'uso e di come l'integrazione di MLFlow possa migliorare il monitoraggio dei modelli in fase di training.

Il lavoro svolto ha richiesto uno studio approfondito delle tecnologie coinvolte, nonché una fase di sperimentazione pratica su come allenare un modello di object detection, e uno studio di come monitorare le metriche di training in modo efficace.

## Chapter 2

# Tecnologie e Modelli Object Detection