# 

# Progetto di Sistemi Distribuiti e Cloud Computing Corso di Laurea Triennale in Ingegneria Informatica - a.a. 2024/25

Studente: \_\_\_\_\_\_\_\_\_\_Martina Giacobbe\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Matricola: \_\_\_\_\_\_\_263662\_\_\_\_\_\_\_\_\_\_\_\_

Data di assegnazione del progetto: \_\_\_\_16/05/2025\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Realizzare una pipeline automatizzata per il Machine Learning, basata su architettura serverless e tecnologie FaaS. Il flusso deve iniziare con il caricamento di un file contenente dati del seguente dataset pubblico su un servizio di storage sul cloud:

* [Adult Income Dataset](https://archive.ics.uci.edu/dataset/2/adult)

**FASE 1:** Il caricamento del dataset sullo storage deve fungere da evento trigger e attivare automaticamente una funzione FaaS responsabile della fase successiva.

**FASE 2:** In questa fase, i dati vengono letti, analizzati, trasformati per renderli idonei per l’analisi. Si effettuano, ad esempio, operazioni come la rimozione di valori nulli, la normalizzazione delle feature, la codifica di variabili categoriche, e altre trasformazioni necessarie per la qualità del dataset. Al termine di questa fase, i dati puliti e trasformati vengono salvati in una nuova posizione all'interno dello storage

**FASE 3:** a questo punto, una seconda funzione serverless viene attivata (automaticamente o tramite schedulazione). Questa funzione ha il compito di addestrare un modello di Machine Learning utilizzando i dati processati. Il tipo di modello può essere scelto tra algoritmi comunemente usati per quel dataset (es. regressione logistica, decision tree, random forest, support vector machine, ecc.) e viene addestrato utilizzando una libreria standard per il machine learning (es. scikit-learn). Una volta completato l’addestramento, il modello viene serializzato e salvato in una posizione dedicata all’interno dello storage.

**FASE 4:** questa fase implementa il servizio di inferenza e può essere implementata tramite un’ulteriore funzione FaaS o VM che espone un’interfaccia HTTP. Tale endpoint consente di inviare nuove istanze di dati per ottenere predizioni in tempo reale.

Per la realizzazione dell’applicazione, lo studente dovrà includere un’analisi dei requisiti funzionali e non funzionali del sistema da realizzare. Lo studente non ha vincoli sul linguaggio di programmazione da utilizzare. Il sistema realizzato, tuttavia, dovrà utilizzare le soluzioni di calcolo, storage e virtualizzazione messe a disposizione da:

* Amazon AWS

Per il completamento del progetto, lo studente dovrà presentare una demo funzionante dell’applicazione sopra descritta.

## Consegna e valutazione del progetto

La consegna del progetto consiste in:

* consegna mediante CD-ROM o link a spazio di Cloud storage contenente tutti i sorgenti necessari per il funzionamento del sistema e una relazione in formato pdf;
* copia cartacea della relazione da presentare il giorno della demo.

I principali criteri di valutazione del progetto saranno:

1. rispondenza ai requisiti;
2. originalità;
3. organizzazione del codice (es., leggibilità e modularità);
4. organizzazione e completezza della relazione.

Il progetto deve essere consegnato due giorni prima della data dell’esame.