

*Studentessa: Flavia Stella*

*Master di I livello in Data Analysis per la Business Intelligence e la Data Science*

*Tesi: "Ottimizzazione dei controlli sui fondi pubblici: un approccio data-driven tramite tecniche di Machine Learning"*

#### ABSTRACT

La presente tesi si pone l'obiettivo di analizzare i dati di un'azienda finanziaria pubblica che, in seguito ad alcune variazioni normative a livello comunitario, ha deciso di adottare un approccio data driven alla propria metodologia di lavoro. L'adozione di tale approccio comporta sia un cambiamento organizzativo, sia l'acquisizione di nuove competenze ed in particolare un maggiore investimento in attività di tipo statistico.

I diversi database aziendali vengono quindi analizzati per individuare informazioni utili alla valutazione dei rischi e indispensabili per ottimizzare i costi.

Dopo una prima esplorazione descrittiva dei dati, si procede mediante l'applicazione di alcuni modelli di Machine Learning con lo scopo di verificare se, a partire dai dati storici presenti in azienda sulle rendicontazioni dei progetti agevolati, sia possibile prevedere l'esito delle future rendicontazioni che saranno presentate dai beneficiari.

Per pervenire ad un dataset coerente ai fini dell'applicazione dei modelli, si è utilizzato il linguaggio di programmazione open source Python; successivamente i dati sono stati organizzati in un database dedicato costruito su PostgreSQL; sono quindi state applicate alcune query (mediante il linguaggio SQL) per ottenere tutti i dati necessari, adeguatamente preparati.

Il piano dell'opera prevede quattro capitoli oltre all'introduzione.

Il secondo capitolo descrive l'azienda coinvolta, le esigenze specifiche da cui è nata la necessità di adottare un approccio statistico, le caratteristiche e l'organizzazione dei dati.

Il terzo capitolo contiene l'analisi descrittiva del dataset, discutendo le peculiarità delle variabili qualitative e quantitative utilizzate.

Il quarto capitolo espone la metodologia di applicazione dei modelli di Machine Learning, con particolare attenzione agli iperparametri.

Il quinto capitolo mostra l'applicazione dei modelli al dataset aziendale costruito e illustra i risultati ottenuti, delineando altresì le conclusioni.