MASTER di I LIVELLO in

ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE

A.A. 2024/2025

TITOLO TESI

Modelli statistici per la generazione di indagini virtuali

MARRAS RUGGERO

Abstract

Negli ultimi decenni la crescita esponenziale dei dati digitali ha reso sempre più centrale l'uso della Data Science e della Business Intelligence per supportare decisioni strategiche. Tuttavia, solo una minima parte dei dati prodotti viene effettivamente analizzata, a causa della loro eterogeneità e della presenza di rumore. In questo contesto, i questionari e le indagini campionarie rimangono una fonte informativa di primaria importanza, soprattutto per l'analisi dei comportamenti sociali e delle intenzioni individuali.

Il presente lavoro si inserisce nell'ambito di una collaborazione con l'azienda Metis-Ricerche, finalizzata all'analisi e al monitoraggio delle intenzioni di viaggio dei cittadini italiani, con particolare riferimento al Piemonte. L'obiettivo principale è stato sviluppare un modello statistico predittivo basato sulla regressione logistica in grado di simulare le risposte ai questionari sulle intenzioni di vacanza, riducendo costi e tempi di raccolta dei dati.

Il progetto ha previsto una fase preliminare di pulizia e armonizzazione dei dati, la costruzione di una Analytical Base Table (ABT) coerente e la definizione di set di training e validation. La modellazione è stata sviluppata secondo una pipeline sequenziale, in cui le risposte alle domande precedenti vengono utilizzate come predittori per le domande successive, consentendo di simulare l'intero percorso del questionario. Sono stati confrontati modelli basati su variabili osservate, su variabili predette e su selezioni mirate di feature.

I risultati mostrano che i modelli basati su variabili osservate garantiscono una maggiore capacità discriminante, mentre l'approccio con variabili predette, pur meno accurato, consente di esplorare scenari realistici anche in assenza di dati completi. Il framework sviluppato si configura quindi come uno strumento flessibile per la generazione di indagini virtuali, con applicazioni potenziali nella pianificazione turistica e nelle attività di business intelligence.

Sviluppi futuri includono l'introduzione di tecniche di machine learning avanzato, l'automatizzazione della pulizia dei dati e l'integrazione dei modelli in strumenti di supporto decisionale.