

Master Universitario in: “ANALISI DATI PER LA BUSINESS INTELLIGENCE” A.A.
2020/2021

Titolo della tesi: “Analisi statistica e clusterizzazione di misure di parametri ambientali da sensori co-locati.”

Autore: Manuele Soncini

ABSTRACT

Nell’ambito della ricerca in climatologia vengono utilizzate strumentazioni sofisticate in grado di percepire, raccogliere e registrare accuratamente le misure di svariati parametri ambientali. Nel presente progetto sono stati elaborati dati, raccolti tra novembre 2016 e gennaio 2017 e provenienti da una serie di sensori co-locati, installati cioè presso un unico sito di misura posto in una zona pianeggiante di Castelnuovo Don Bosco, a 40 km da Torino. I dati comprendono misure di temperatura (rilevata da 7 sensori identici), radiazione solare, umidità relativa, velocità e direzione del vento (1 sensore per ciascuna grandezza). L’incertezza della strumentazione e la sperimentazione in ambiente naturale non controllato impongono un’attenta valutazione della qualità delle misurazioni: anche quando gli strumenti sono accurati e attendibili, il reale valore del parametro che viene misurato (misurando) ricade sempre all’interno di un intervallo di possibili valori, ragionevolmente attribuibili al misurando, che viene definito come “incertezza di misura”; inoltre, la variabilità delle misure *in-field* è sempre assai maggiore di quella riscontrata in ambiente controllato (laboratorio) per via delle cosiddette “variabili ambientali d’influenza”, contribuendo in maniera sostanziale all’incertezza totale da assegnare alle misure dei parametri ambientali.

Il progetto ha come obiettivo specifico la valutazione delle influenze, dinamiche e meccanismi per i quali le differenze di temperature (ΔT) misurate dai sensori co-locati risultano più o

meno ampie. In particolare, si è proceduto all'individuazione di sottopopolazioni (corrispondenti a specifici sotto-intervalli delle variabili ambientali) che possano aiutare a spiegare la dipendenza dei delta T dalle altre variabili ambientali e le corrispondenti incertezze.

I fenomeni naturali di interesse, di cui verrà presentata una panoramica tramite un'analisi esplorativa, sono stati analizzati statisticamente in maniera approfondita per raggiungere lo scopo dell'elaborato. Sono state analizzate le influenze delle variabili ambientali sui delta T attraverso modelli di regressione lineare multipla; sono state selezionate opportune sottopopolazioni per valutare, al loro interno, le dinamiche delle variabili; infine, è stato utilizzato un algoritmo di *machine learning* con apprendimento non supervisionato per clusterizzare le osservazioni. L'utilizzo dell'algoritmo K-Means ha aiutato a rilevare alcune dinamiche sottese molto interessanti all'interno del *dataset*. Ad esempio, i sensori di temperatura e i relativi delta T hanno mostrato comportamenti analoghi in due situazioni ambientali specifiche e molto diverse tra loro: in condizioni di luce solare assente o di alta velocità del vento. All'interno dell'elaborato verranno descritte con maggiori dettagli tutte le analisi effettuate.