

Camellini Agnese

Master Universitario di I livello
Analisi Dati per la Business Intelligence

Abstract della Tesi

Titolo: Quale eco per Torino? Proposta di studio sugli eventi dell'autunno torinese.

La tesi prende forma alla fine del percorso di stage previsto per il Master di **Analisi Dati per la Business Intelligence** e che è stato svolto presso Città di Torino con riferimento al Dottor Stefano Benedetto in qualità di tutori aziendale ed in collaborazione con CSI-Piemonte con riferimento alla Dott.ssa Giuliana Bonello in qualità di tutor accademico ed il Dipartimento di Informatica dell'Università degli Studi di Torino con riferimento alla Prof.ssa Rosa Meo.

Il progetto ha l'obiettivo di analizzare la relazione tra gli eventi più importanti dell'autunno torinese ed il loro impatto sulla città in termini di fruizione. Esso si sviluppa in due fasi distinte: la prima di raccolta dati, che comprende la strutturazione di massima delle analisi che si andranno a svolgere ed il rapporto con i soggetti coinvolti per la fornitura dei dati, e la seconda di elaborazione dei dati raccolti e di esposizione dei risultati ottenuti.

La fase di raccolta dati ha occupato quasi i due terzi del periodo di stage ed ha coinvolto numerose organizzazioni, fondazioni, associazioni ed enti in relazione alla fornitura dei dati. Sono stati selezionati quattro eventi di tipologie diverse, **Salone del Gusto** e **Terra Madre** (23 – 27 ottobre 2014), l'internazionale d'arte contemporanea **Artissima** (6 – 9 novembre 2014), il festival di musica elettronica **Club to Club** (6 – 9 novembre 2014), la gara podistica **Turin Marathon** (16 novembre 2014). I dati sono stati raccolti contestualmente al loro svolgimento per il periodo compreso tra il 1° ottobre ed il 30 novembre 2014 e si articolano in tre punti:

- I. *Dati sui flussi turistici nei giorni degli eventi*: riguardano i dati sugli ingressi ai musei (Fonti: Fondazione Fitzcarraldo e Fondazione Torino Musei), i dati sulle presenze turistiche nelle strutture ricettive dell'area del Comune di Torino (Fonte: Sviluppo Piemonte Turismo) ed i dati sulle presenze delle utenze telefoniche (Fonte: Telecom Italia);
- II. *Dati sulle affluenze agli eventi*: riguardano i dati sul numero di ingressi ad ogni singolo evento con un dettaglio giornaliero e sono stati richiesti direttamente alle organizzazioni che ne hanno gestito lo svolgimento;
- III. *Dati che parlano degli eventi*: riguardano i dati testuali provenienti dai quotidiani, è stata selezionata la rassegna stampa relativa agli eventi, e quelli estratti dai principali canali di social networking quali Twitter e Facebook; purtroppo questa ultima tipologia di dati non è stata analizzata per carenza dei dati stessi e per mancanza di tempo.

Sui dati così strutturati sono stati predisposti due blocchi di elaborazioni, distinti e successivi.

Strumenti utilizzati: **Knime**, **R**, **QGIS**.

1. *Elaborazioni preliminari*

Sono state evidenziate delle serie temporali di valori relative ad ogni tipologia di dato e sono state analizzate individuando le varie componenti e commentandone gli andamenti; testando la randomicità dei valori osservati della serie e della componente casuale; valutando la significatività statistica, tramite il calcolo del *p-value*, di outliers individuati dalla visualizzazione della distribuzione di frequenza.

L'obiettivo di questa prima parte di elaborazioni è stato valutare se all'interno delle serie storiche individuate ci fossero dei valori significativi di presenze durante i giorni di svolgimento degli eventi selezionati.

2. *Elaborazioni incrociate*

La seconda parte di elaborazioni ha l'obiettivo di incrociare le varie tipologie di dati attraverso l'applicazione di un **algoritmo di clustering** che assegna ad ogni area d'interesse selezionata, aggregazione alla quale sono stati trattati i dati, un cluster di appartenenza. Sulla base del risultato possiamo ipotizzare che le presenze relative alle aree dello stesso cluster abbiano avuto un andamento simile per il periodo di osservazione scelto.