

PROGETTI 2022

Titolo progetto	Big Data e Statistica ufficiale per gestire i Beni Culturali
Descrizione	<p>Il progetto si configura come un'attività sperimentale finalizzata a valutare la possibilità di integrare i dati sulla soddisfazione del turista rilevati dall'indagine ISTAT sulla domanda turistica "Viaggi e Vacanze" con i dati, opportunamente trattati, derivabili dalle recensioni pubblicate dagli utenti sui social network e sulle piattaforme digitali.</p> <p>Il tema della soddisfazione/del giudizio del turista, infatti, riveste particolare interesse per gli utenti delle statistiche del turismo (realtà territoriali, manager delle destinazioni culturali, <i>policy makers</i> locali) e si colloca pienamente in una delle dimensioni chiave per la misurazione della sostenibilità sociale del turismo. Per questo motivo, dal 2019 "Viaggi e Vacanze" ha inserito nel questionario di indagine una batteria di quesiti finalizzata a misurare il livello di soddisfazione del turista (sia in termini di aspettative che di performance) in relazione a diversi aspetti del luogo visitato (il paesaggio e l'ambiente naturale, l'offerta culturale e artistica, l'offerta di intrattenimento, la qualità della ristorazione, la qualità dei servizi locali, il costo della vita del luogo). La raccolta di tali informazioni consente di costruire, per ciascun aspetto e territorio visitato (regione), misure sintetiche medie sia sui livelli di importanza attribuiti sia su quelli di soddisfazione vissuta. Ulteriori approfondimenti richiederebbero l'inserimento di quesiti sul tema particolarmente onerosi da sottoporre ai rispondenti nell'ambito di un'intervista già molto lunga (il focus Viaggi e vacanze è attualmente inserito nell'intervista finale dell'Indagine sulle Spese delle Famiglie).</p> <p>Da qui nasce l'idea di valutare la possibilità di utilizzare, per finalità statistiche, le recensioni pubblicate su internet dagli utenti attraverso le quali raccontano in modo spesso molto dettagliato gli aspetti che ritengono importanti sulla loro esperienza di visita.</p> <p>Lo studio dei contenuti di tali testi potrebbe consentire di approfondire dal punto di vista qualitativo l'informazione sulla soddisfazione.</p>

Obiettivi

Il progetto si propone di utilizzare tecniche e metodi di *Text Mining* per estrarre conoscenza dai dati testuali pubblicati mediante l'applicazione *Google Maps*. Oggetto di analisi sono le recensioni e le opinioni pubblicate dagli utenti in relazione alle esperienze di visita presso punti di interesse, in particolare musei e monumenti presenti sul territorio italiano.

L'obiettivo è quello di sviluppare metodi e procedure a supporto della produzione statistica ufficiale per ampliare e migliorare la rilevazione della soddisfazione del turista. In particolare, la raccolta di recensioni direttamente dai social consentirebbe contemporaneamente di: 1) aumentare la tempestività dei dati raccolti; 2) ridurre il *burden* statistico a carico dei rispondenti; 3) accedere ad un patrimonio informativo notevolmente più ampio rispetto a quello raccolto dall'indagine campionaria; 4) focalizzare l'analisi su territori o destinazioni mirate.

I risultati ottenuti da questo progetto sperimentale potranno essere utilizzati dalle istituzioni per migliorare la gestione dei Beni Culturali con ricadute positive sui cittadini che fruiscono degli stessi beni.

Metodologia

In questo progetto è stata adottata la Latent Dirichlet Allocation (LDA), metodologia molto utilizzata per adattare un modello tematico.

In particolare, LDA è stata utilizzata per approfondire il livello di soddisfazione dei turisti sugli aspetti già rilevati dall'indagine "Viaggi e Vacanze": il paesaggio e l'ambiente naturale, l'offerta culturale e artistica, l'offerta di intrattenimento, la qualità della ristorazione, la qualità dei servizi locali, il costo della vita del luogo. I risultati dell'indagine possono essere integrati con i risultati della *Topic Analysis* tecnica propria del *Text Mining*.

Risultati ottenuti

INTRODUZIONE

Una delle ipotesi del progetto è che il maggior numero di recensioni sul web siano rilasciate per le principali attrazioni culturali, spesso localizzate all'interno delle grandi città. Per confrontare queste recensioni con il livello di soddisfazione espresso al ritorno del viaggio mediante i dati d'indagine è necessario operare una selezione statisticamente affidabile delle destinazioni, data la natura campionaria dell'indagine Viaggi e vacanze.

La scelta delle città è stata determinata, quindi, sulla base dell'analisi delle distribuzioni di frequenza dei viaggi per destinazione principale a livello regionale, provinciale e comunale, garantendo la rappresentatività di tutte le aree del Paese e l'affidabilità delle stime. All'interno della cinque ripartizione del territorio italiano (Nord-Ovest, Nord-Est, Centro, Sud, Isole), si è scelto il livello di dettaglio territoriale le cui stime presentano un errore relativo accettabile.

La logica seguita è stata a "a imbuto": dai flussi di viaggi distinti per ripartizione geografica, si è esaminata la consistenza dei viaggi per regione e, all'interno di ciascuna regione, per provincia e laddove necessario, si è scesi fin nel dettaglio comunale. Questo primo screening ha portato, a scegliere cinque regioni (Lombardia, Veneto, Lazio, Campania e Sicilia) e una sola provincia per regione: province di Milano, Venezia, Roma, Napoli, Palermo.

Dall'analisi degli errori relativi, l'affidabilità dei dati è risultata essere accettabile a livello regionale per tutte e cinque le regioni (con un errore inferiore o uguale al 10%), mentre a livello provinciale non lo è per Palermo (con un errore relativo di circa il 30% comporterebbe un rischio troppo elevato di fornire stime errate). A livello comunale, invece, soltanto Roma e Milano hanno stime affidabili, perché polarizzano la quasi totalità dei flussi provinciali, permettono un'analisi territoriale più fine.

Per ragioni legate soltanto al contenimento dei tempi del laboratorio, abbiamo scelto di considerare esclusivamente la città di Roma, per la quale confrontare il livello di soddisfazione espresso dai dati d'indagine con quello delle recensioni sul Colosseo, nel 2022.

RISULTATI TECNICI E PIATTAFORMA IT

Le recensioni sul Colosseo sono state raccolte mediante un software realizzato utilizzando ChatGPT-4.

La tecnologia chiamata *deep learning* ci ha permesso di addestrare un sistema di intelligenza artificiale a svolgere l'attività di raccolta dati. Il sistema che abbiamo creato utilizza a sua volta un servizio di terze parti che si collega mediante il protocollo HTTP alle API di Google Places.

Sfruttando il potere del linguaggio naturale e delle API, siamo in grado di fornire ai ricercatori un accesso rapido e conveniente alle informazioni di valutazione dei luoghi al centro delle loro ricerche.

RISULTATI TEMATICI E METODOLOGICI

Analizzando circa 1500 recensioni abbiamo osservato che emergono recensioni dettagliate e statisticamente rilevanti in relazione a tre delle dimensioni indagate dall'indagine "Viaggi e Vacanze": il costo della vita nel luogo visitato, l'offerta culturale e artistica e la qualità dei servizi offerti declinandoli in relazione alla specificità della visita.

L'analisi effettuata permette di ottenere risultati caratterizzati da maggior dettaglio informativo rispetto alla scala numerica da 0 a 10 attualmente prevista nel questionario dell'indagine.

Inoltre, nei commenti pubblicati e selezionati utilizzando LDA è possibile rinvenire consigli utili all'amministrazione pubblica e territoriale per una gestione più efficiente dei beni culturali, riferita in questo caso al Colosseo.

SVILUPPI E IMPATTO SUL PROCESSO DI INDAGINE

Riteniamo che i risultati raggiunti siano molto promettenti e si possa dare luogo a una sperimentazione su più larga scala che prenda in esame un insieme rappresentativo di beni culturali anche in altre regioni, oltre il Lazio, caratterizzate dai flussi turistici più importanti.

Potendo contare su un adeguato supporto infrastrutturale sarebbe possibile dare inizio a un progetto più ampio nell'ambito della prossima edizione del Laboratorio Innovazione da affiancare all'indagine Viaggi e Vacanze. Avendo a disposizione un maggior numero di recensioni sarebbe possibile analizzare la soddisfazione dei turisti anche rispetto agli altri aspetti non ancora considerati.

La piattaforma IT ivi realizzata in forma prototipale unitamente all'applicazione della LDA su larga scala, potrebbero consentire di affiancare sistematicamente i risultati dell'indagine Viaggi e Vacanze a quelli ottenuti con un'analisi basata sul Text Mining, fino a sostituire le informazioni sulla soddisfazione rilevate dall'indagine con quelle di fonte big data.

.