

MASTER DI I LIVELLO IN
ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE

Davide Piatteletti

DALLE RECENSIONI ALLE DESTINAZIONI TURISTICHE:
CLUSTERING DI AREE TURISTICHE UTILIZZANDO DATI SOCIAL

L'obiettivo è di individuare aree turistiche partendo dalle recensioni postate sul web. Sono stati confrontate due differenti metodologie: la prima applica un clustering gerarchico con differenti funzioni legame, la seconda applica un clustering basato sui grafi. Il clustering gerarchico con funzione legame singolo, medio e completo non porta a risultati soddisfacenti in quanto individua gruppi di municipalità che hanno una distribuzione dei collegamenti con tutte le altre città simili ma le municipalità, interne al cluster, non sono collegate tra di loro. Si cambia, quindi, prospettiva e si considerano i dati come dei grafi, reti i cui nodi sono le città e i collegamenti sono il numero di utenti che hanno recensito entrambe le città. Si applica il clustering sul grafo e si ottengono delle aree composte da comuni contigui e quest'ultimi hanno un buon numero di collegamenti tra di loro. Futuri sviluppi possono riguardare: la definizione della tipologia dei turisti, la clusterizzazione delle traiettorie delle attività recensite e la creazione di indici, basati sul text mining, riguardanti la soddisfazione dell'utente per la località recensita.