



UNIVERSITÀ DI VERONA
DIPARTIMENTO DI INFORMATICA

Corso di Laurea Triennale in
Informatica

**Uso di tecniche di LLM per analizzare
traiettorie turistiche e per suggerire la
prossima attrazione turistica da
visitare**

Relatrice

Sara Migliorini

Candidato

Mattioli Simone

Indice

Sommario

Questo articolo presenta uno studio completo sull'utilizzo dei Large Language Models (LLM) per prevedere i modelli di mobilità umana nei contesti turistici urbani. Introduciamo LLM-Mob, un nuovo framework che sfrutta le capacità di comprensione del linguaggio naturale dei moderni LLM per prevedere i prossimi Punti di Interesse (POI) dei turisti in base ai loro modelli di visita storici. Il nostro approccio viene valutato sul dataset VeronaCard, che contiene dati turistici reali di Verona, Italia, relativi a diversi anni (dal 2014 al 2023). Viene analizzato l'impatto di diversi tipi di informazioni contestuali sull'accuratezza della previsione, dai nomi di POI come base, coordinate geografiche e caratteristiche temporali. Il framework dimostra il potenziale degli LLM generalisti per comprendere complessi modelli spazio-temporali nella mobilità umana senza richiedere un'ingegnerizzazione approfondita delle caratteristiche o architetture specifiche per dominio. L'implementazione è partita sostituendo il framework di LLM-Mob che sfruttava i modelli OpenAI tramite API-KEY con invece modelli open source (come Llama 3.1), rendendo l'approccio più accessibile ed economico per scopi di ricerca.

1 Introduzione

La previsione della mobilità umana si è affermata come un'area di ricerca cruciale, con applicazioni che spaziano dalla pianificazione urbana ai sistemi di trasporto, dai sistemi di raccomandazione alla salute pubblica. Gli approcci tradizionali alla previsione della mobilità si sono basati in larga misura su modelli statistici, algoritmi di apprendimento automatico e architetture di deep learning specificamente progettate per dati sequenziali. Tuttavia, i recenti progressi nei Large Language Model (LLM) hanno aperto nuove strade per comprendere e prevedere i modelli di comportamento umano attraverso capacità di elaborazione del linguaggio naturale.

La sfida fondamentale nella previsione della mobilità umana risiede nel catturare la complessa interazione tra fattori spaziali, temporali e comportamentali che influenzano le decisioni di movimento individuali. Gli approcci tradizionali di apprendimento automatico richiedono spesso un'ampia progettazione di feature e conoscenze specifiche del dominio per codificare efficacemente queste relazioni. Al contrario, i LLM possiedono capacità intrinseche di comprensione delle relazioni e dei modelli contestuali, che possono essere sfruttate per prevedere i modelli di mobilità attraverso informazioni contestuali e prompt attentamente progettati.

Questo articolo presenta un framework che sfrutta la potenza dei Large Language Model per la previsione della mobilità umana in contesti turistici. Il mio approccio si differenzia dai metodi convenzionali perché tratta la previsione della mobilità come un compito di comprensione del linguaggio naturale, in cui i nomi dei POI storici, le informazioni geografiche e il contesto temporale vengono codificati come prompt testuali per l'inferenza LLM.

I principali contributi di questo lavoro includono:

- Un nuovo framework per la previsione della mobilità umana utilizzando modelli di linguaggio (LLM)
- Valutazione completa di diversi tipi di informazioni contestuali sull'accuratezza della previsione
- Implementazione utilizzando modelli open source eliminando le dipendenze da API proprietarie ed a pagamento
- Esperimenti approfonditi su dati turistici reali tratti dal dataset VeronaCard
- Analisi dell'impatto della prossimità geografica e dei modelli temporali sulle prestazioni di previsione della mobilità

2 Lavori correlati

La previsione della mobilità umana è stata ampiamente studiata in diverse discipline, con approcci che spaziano dai modelli statistici alle architetture di deep learning. I primi lavori si concentravano su modelli basati su Markov che catturavano le probabilità di transizione tra posizioni sulla base di modelli storici. Approcci più recenti hanno sfruttato reti neurali ricorrenti (RNN), reti a memoria a lungo termine (LSTM) e meccanismi di attenzione per modellare le dipendenze sequenziali nei dati di mobilità.

Gli approcci basati su grafi hanno acquisito importanza grazie alla loro capacità di catturare le relazioni spaziali tra posizioni. Questi metodi in genere rappresentano le posizioni come nodi in un grafo, con archi che codificano la prossimità spaziale o le probabilità di transizione. Tuttavia, tali approcci richiedono spesso un'ampia pre-elaborazione e l'ingegnerizzazione di feature specifiche del dominio per ottenere prestazioni soddisfacenti.

L'emergere dei Large Language Model (LLM) ha introdotto nuove possibilità per la previsione della mobilità. Lavori recenti hanno esplorato l'uso dei LLM per vari compiti spazio-temporali, dimostrando la loro capacità di comprendere relazioni geografiche e modelli temporali attraverso descrizioni in linguaggio naturale. Il nostro lavoro si basa su queste basi, concentrandosi in particolare sulla previsione della mobilità turistica e analizzando l'impatto di diversi tipi di informazioni contestuali.

3 Metodologia

3.1 Formulazione del Problema

Formuliamo il problema di previsione della mobilità umana come un task di raccomandazione sequenziale spazio-temporale. Data la sequenza storica di POI visitati da un turista $S = \{p_1, p_2, \dots, p_n\}$ ordinati cronologicamente, l'obiettivo è prevedere i k POI più probabili che il turista visiterà successivamente, dove $k = 5$ nel nostro setup sperimentale.

Diversamente dagli approcci tradizionali che modellano questo problema come classificazione multi-classe, adottiamo una strategia innovativa basata su *anchor-based prompting*, dove:

- Ogni sequenza viene divisa in: storico delle visite $H = \{p_1, \dots, p_{j-1}, p_{j+1}, \dots, p_{n-1}\}$, POI corrente (anchor) p_j , e target da prevedere p_n
- La strategia di selezione dell'anchor è configurabile: penultimate (default), first, middle, o indice esplicito
- Il modello riceve informazioni contestuali multi-dimensionali: cluster comportamentale, cronologia delle visite, posizione geografica corrente, e POI nelle vicinanze

3.2 Architettura del Sistema

Il nostro framework si basa su un'architettura modulare che combina clustering comportamentale, analisi geospaziale e generazione linguistica attraverso LLM. La pipeline è ottimizzata per elaborazioni massive su cluster HPC con gestione robusta di timeout e fallimenti.

3.2.1 Infrastructure e Deployment

Il sistema è progettato per operare su cluster HPC (High Performance Computing) utilizzando:

- **LLM Backend:** Ollama, configurazione ottimizzata per GPU A100 (visto che ho eseguito i miei test sull'HPC Italiano Leonardo)
- **Gestione Risorse:** Sistema di checkpointing incrementale per elaborazioni interrotte
- **Fault Tolerance:** Retry mechanism con backoff esponenziale e health checking automatico
- **Scalabilità:** Processamento parallelizzabile per dataset multi-anno

3.3 Dataset e Pre-elaborazione

3.3.1 Caratteristiche del Dataset VeronaCard

Gli esperimenti sono condotti sul dataset VeronaCard (2014-2023), contenente:

- **Registrazioni turistiche:** 10 anni di log di utilizzo con timestamp precisi
- **Copertura geografica:** 50+ POI distribuiti nel centro storico di Verona
- **Diversità comportamentale:** Turisti con pattern di visita eterogenei (culturali, storici, ricreativi)
- **Metadati geospaziali:** Coordinate GPS precise per ogni POI

3.3.2 Pipeline di Pre-elaborazione

La pre-elaborazione implementa un processo multi-stage con validazioni robuste:

1. Data Cleaning:

- Merge delle timbrature con il catalogo POI validato (JOIN fatto su name short)
- Filtro delle sequenze con almeno 3 visite e 2+ POI distinti
- Standardizzazione dei timestamp e identificativi

2. Behavioral Clustering:

- Costruzione matrice user-POI sparsa
- Standardizzazione con StandardScaler
- K-means clustering ($k = 7$) per identificazione profili comportamentali

3. Spatial Analysis:

- Calcolo distanze Haversine tra POI
- Identificazione POI nelle vicinanze (raggio 2km)
- Analisi pattern di movimento sequenziali

3.4 Prompt Engineering e Context Construction

3.4.1 Strategia di Prompt Design

Il prompt viene costruito dinamicamente incorporando:

- **Profilo comportamentale:** "Turista cluster X a Verona"
- **Context storico:** Lista POI precedentemente visitati (escluso anchor)
- **Posizione corrente:** POI anchor con coordinate
- **Prossimità geografica:** Top-10 POI più vicini con distanze in km
- **Constraint task:** Richiesta di 5 raccomandazioni ordinati per probabilità

Il formato di output è strutturato in JSON per garantire parsing deterministico:

```
{"prediction": ["poi1", "poi2", "poi3", "poi4", "poi5"],
"reason": "spiegazione delle scelte"}
```

3.4.2 Gestione del Context Window

Per ottimizzare l'utilizzo del context window limitato (1024 tokens), il prompt viene compattato:

- Selezione dei 10 POI più vicini geograficamente (vs. tutti i POI disponibili)
- Formato condensato: "POI_name (distance_km)"
- Eliminazione di testo ridondante mantenendo informazioni critiche

Tutto ciò per non fornire "troppe" informazioni ai modelli LLM, che almeno per i test iniziali, sono piccoli (qualche miliardo di parametri). Per scoprire fino a dove riusciamo ad ottenere dei risultati accettabili.

3.5 Strategia di Valutazione

3.5.1 Metriche di Performance

Il sistema viene valutato attraverso metriche standard per sistemi di raccomandazione:

- **Top-1 Accuracy:** $\text{Acc}_{@1} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i = \hat{y}_i^{(1)}\}$
- **Top-k Hit Rate:** $\text{HR}_{@k} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i \in \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}\}$
- **Mean Reciprocal Rank:** $\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}$
- **Catalogue Coverage:** $\text{Coverage} = \frac{|\bigcup_i \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}|}{|\mathcal{P}|}$

3.5.2 Test Set Construction

Per ogni turista con sequenza $\{p_1, p_2, \dots, p_n\}$ ($n \geq 3$):

- **History:** $\{p_1, \dots, p_{n-2}\}$ (escluso anchor)
- **Current POI (anchor):** p_{n-1}
- **Ground truth:** p_n

Questa costruzione permette valutazione realistica del task di previsione "next-POI".

3.5.3 Analisi di Errore

Il framework include strumenti avanzati per interpretabilità:

- **Worst-performing pairs:** Identificazione coppie POI problematiche
- **Confusion matrix:** Visualizzazione errori per subset temporali/geografici

4 Strategie di prompt e risultati sperimentali

L'efficacia dei Large Language Model (LLM) nella previsione del prossimo POI dipende fondamentalmente dalla progettazione di strategie di prompt consapevoli del contesto che codifichino efficacemente i modelli di mobilità spazio-temporale

4.0.1 Configurazione sperimentale per analisi multi-modello

La fase attuale di sviluppo utilizza LLaMA 3.1 8B come modello di riferimento per la validazione del framework e l'ottimizzazione dei protocolli sperimentali. Il sistema è progettato per supportare analisi comparative estese su diversi modelli LLM disponibili tramite Ollama, permettendo valutazioni sistematiche dell'impatto dell'architettura del modello sulle performance di predizione dei POI turistici.

Il framework implementa standardizzazione dei parametri di inferenza e template di prompt per garantire confronti equi tra modelli, mentre mantiene flessibilità per ottimizzazioni specifiche per architettura quando necessarie. Il nostro framework sperimentale valuta sistematicamente tre distinti paradigmi di prompt, ognuno dei quali incorpora progressivamente ulteriori dimensioni contestuali per valutarne l'impatto sull'accuratezza della raccomandazione.

4.1 Architettura del sistema e configurazione sperimentale

Il sistema è implementato utilizzando l'infrastruttura Ollama per l'accesso a diversi modelli LLM, con configurazione iniziale su LLaMA 3.1 8B per la fase di sviluppo e test. L'architettura è progettata per supportare analisi comparative tra modelli diversi mantenendo consistenza nei protocolli sperimentali. Il framework implementa un'architettura ottimizzata per l'ambiente HPC che gestisce timeout GPU, retry intelligenti e checkpointing per elaborazioni a lungo termine. La pipeline implementa le seguenti componenti chiave:

- **Architettura multi-modello modulare:** Sistema flessibile basato su Ollama che permette testing comparativo tra diversi modelli LLM mantenendo protocolli sperimentali consistenti

- **Gestione robusta delle connessioni LLM:** Sistema di retry progressivo con backoff esponenziale e gestione intelligente dei timeout GPU, configurabile per diversi modelli
- **Checkpointing incrementale:** Salvataggio automatico ogni 500 predizioni per garantire recuperabilità in caso di interruzioni durante test estesi su modelli multipli
- **Ottimizzazioni per HPC:** Configurazione adattiva per GPU A100 con gestione della memoria VRAM e parametri di inferenza ottimizzati per diversi modelli
- **Validazione robusta delle risposte:** Parser JSON con gestione degli errori e validazione del contenuto delle predizioni, standardizzato per garantire comparabilità tra modelli diversi

4.2 Tassonomia delle strategie di prompt

Definiamo tre strategie di prompt gerarchiche, ciascuna basata sulla precedente per valutare il contributo incrementale di diverse caratteristiche contestuali:

1. **Strategia di base (solo nome del POI):** L'approccio fondamentale fornisce al modello una sequenza ordinata cronologicamente di POI visitati in precedenza, rappresentati esclusivamente dai loro nomi canonici. Questa strategia funge da baseline, concentrandosi esclusivamente su pattern sequenziali senza ulteriori informazioni contestuali.

Matrice di confusione - Strategia di base

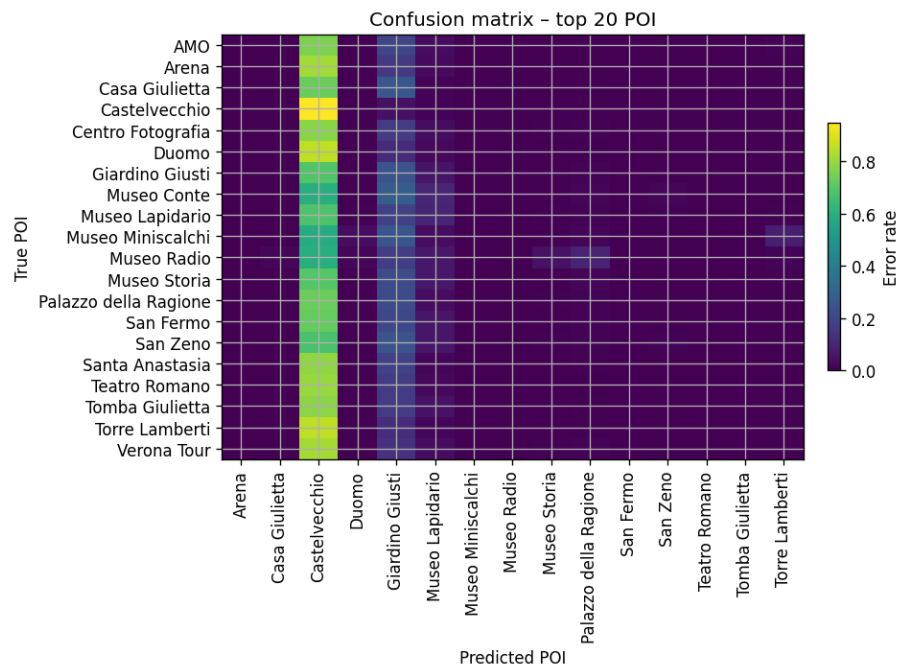


Figura 1: La matrice di confusione illustra i tassi di classificazione errata tra i 20 POI più visitati a Verona con la strategia di base. I colori più brillanti indicano una maggiore confusione. In particolare, Castelvecchio viene spesso erroneamente previsto come prossimo POI, a causa di forti bias sequenziali o di modelli di preferenze degli utenti.

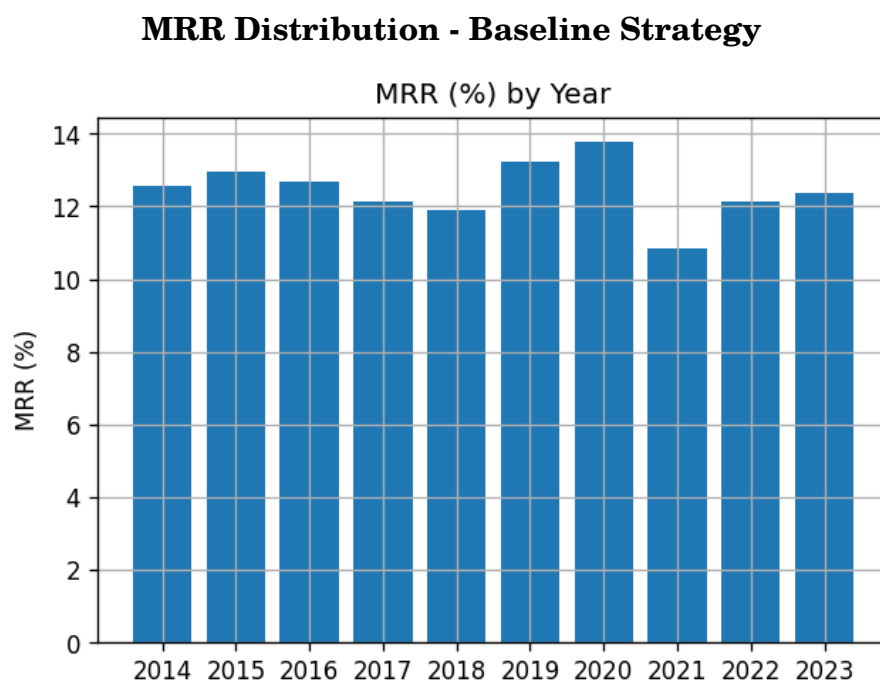


Figura 2: Il grafico mostra la performance del Mean Reciprocal Rank (MRR) per anno, mostrando quanto bene il modello classifica il POI corretto. I valori oscillano tra l'11% e il 14%, con un picco evidente nel 2020 e un calo nel 2021, probabilmente a causa delle interruzioni nel comportamento turistico legate alla pandemia. La tendenza suggerisce una performance predittiva generalmente stabile nel tempo, con una varianza modesta.

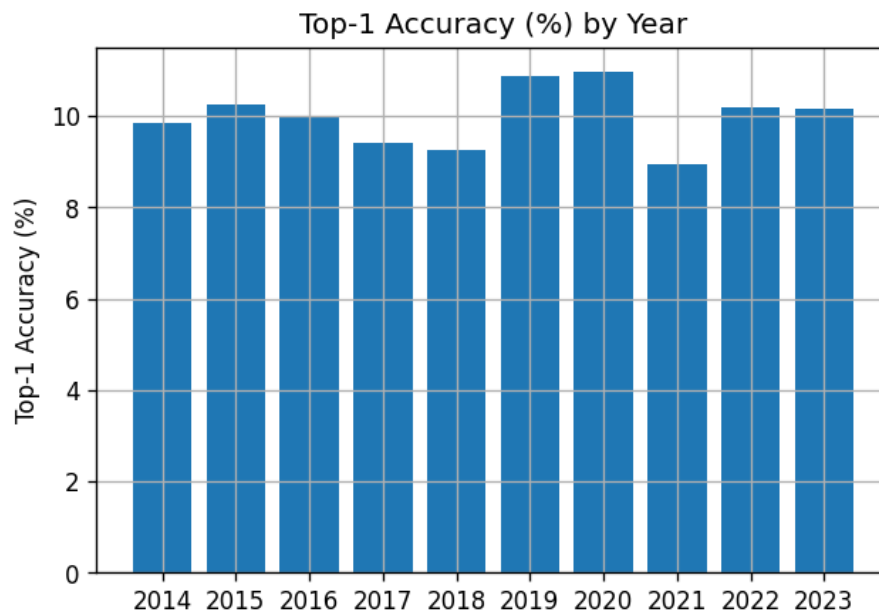
Top-1 Accuracy - Baseline Strategy

Figura 3: Punteggio di accuratezza Top-1 che indica la percentuale di volte in cui il POI corretto è stato la previsione migliore. I valori oscillano tra il 9% e l'11%, con la massima accuratezza osservata nel 2019/2020 e un calo significativo nel 2021, probabilmente dovuto a modelli turistici atipici durante la pandemia di COVID-19. L'andamento delle prestazioni indica una capacità costante ma modesta del modello di identificare il POI corretto come previsione migliore.

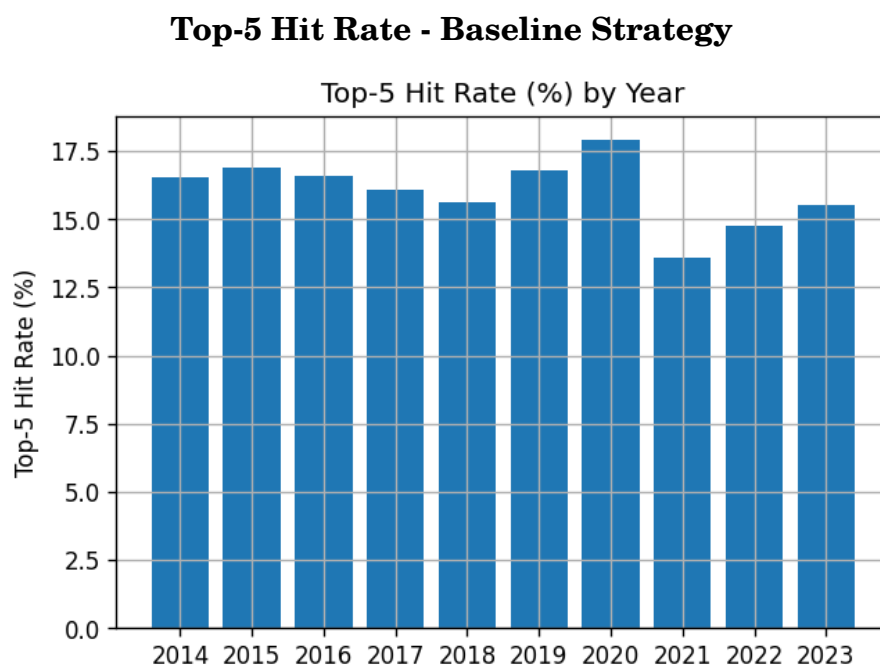


Figura 4: Percentuale di previsioni in cui il POI corretto compare tra i primi 5 POI suggeriti. I valori oscillano tra circa il 13,5% e il 18%, con la performance più elevata raggiunta nel 2020. Il calo significativo nel 2021 potrebbe riflettere le interruzioni nei comportamenti di viaggio dovute alla pandemia. Nonostante questa anomalia, la tendenza generale indica che il modello include in modo affidabile il POI corretto tra le sue prime 5 raccomandazioni.

Worst Performing POI Pairs - Baseline Strategy

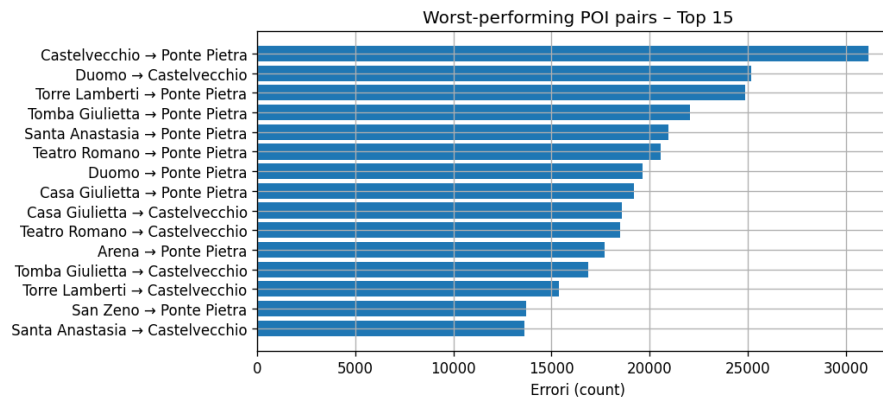


Figura 5: Coppie di POI che vengono più comunemente previste erroneamente, fornendo informazioni sui casi di confusione più frequenti. Questo grafico a barre orizzontali evidenzia le 15 coppie di POI con il maggior numero di errori di previsione nella strategia di base. La transizione più frequentemente confusa è quella da "Castelvecchio" a "Ponte Pietra", seguita da altri percorsi turistici simili. Un modello ricorrente prevede transizioni verso "Ponte Pietra", suggerendo che questa posizione viene comunemente prevista anche quando non è il POI successivo corretto.

- 2. Strategia geospaziale avanzata (nome + geolocalizzazione):** Questa strategia arricchisce ogni POI con coordinate geografiche precise e implementa un sistema di ranking basato sulla distanza di Haversine. Il prompt include una lista dei POI più vicini ordinata per distanza, limitata ai primi 10 risultati per ottimizzare la lunghezza del prompt e migliorare la qualità delle risposte.

Listing 1: Esempio di Prompt Geospaziale

Turista cluster 3 a Verona.

Visitati: Arena di Verona, Casa di Giulietta

Attuale: Castelvecchio

POI Pi Vicini: Ponte Pietra (0.8km), Duomo (1.2km),

Chiesa di Sant'Anastasia (1.4km), Piazza delle Erbe (1.1km)

Suggerisci 5 POI pi probabili come prossime visite considerando distanze e pattern turistici.

Rispondi SOLO JSON: {"prediction": ["poi1", "poi2", ...], "reason": "breve spiegazione"}

Il sistema calcola dinamicamente le distanze dal POI corrente e filtra automaticamente i POI già visitati, implementando vincoli di mobilità realistici con un raggio massimo di 2km per il centro storico di Verona.

Confusion Matrix - Geospatial Strategy

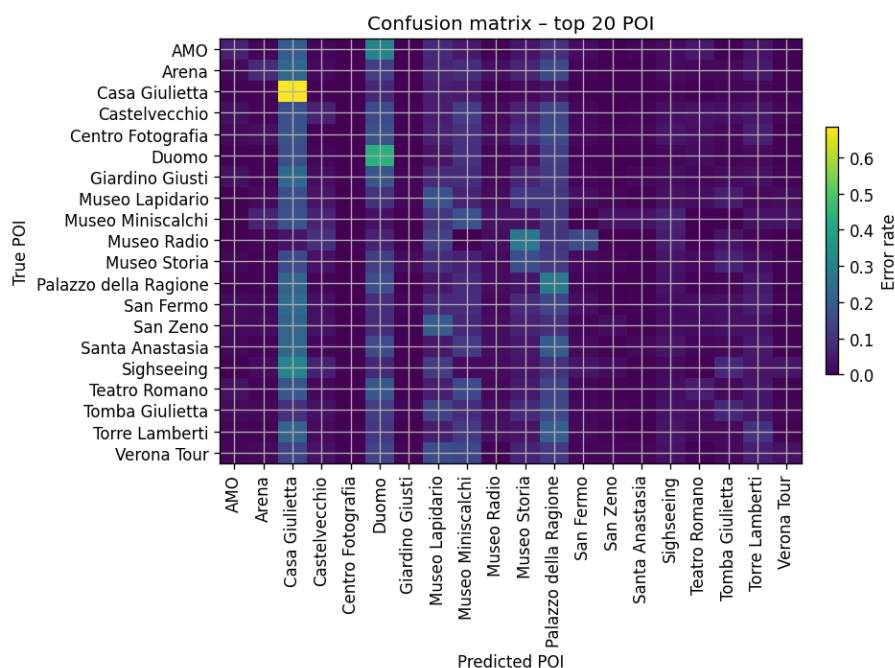


Figura 6: Matrice di confusione che mostra le prestazioni di previsione quando sono incluse le coordinate spaziali.

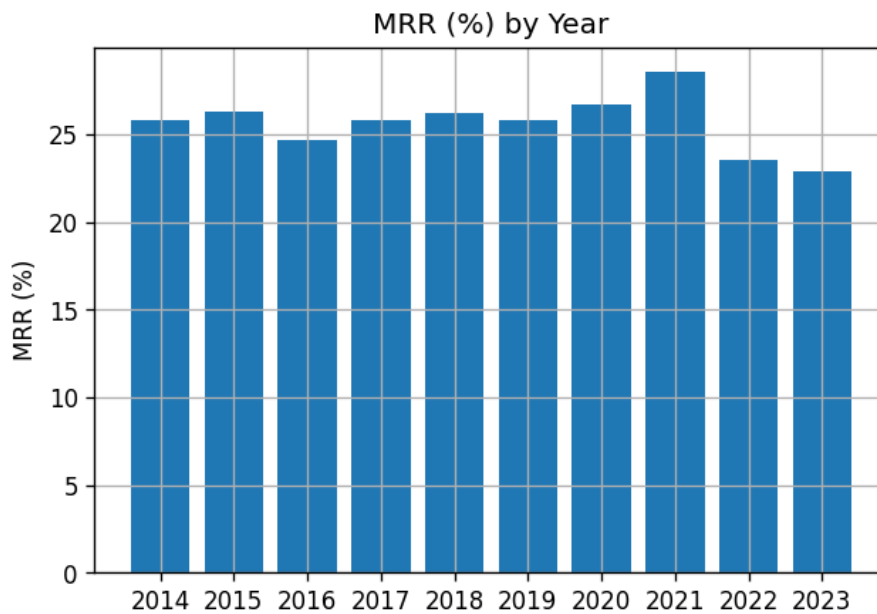
MRR Distribution - Baseline Strategy

Figura 7: Il grafico mostra la performance del Mean Reciprocal Rank (MRR) per anno, mostrando quanto bene il modello classifica il POI corretto. I valori oscillano tra l'22% e il 27%, con un picco evidente nel 2021 e un calo nel 2021-2022, probabilmente a causa delle anomalie nel comportamento turistico legate alla pandemia. La tendenza suggerisce una performance predittiva generalmente stabile nel tempo, con una varianza modesta.

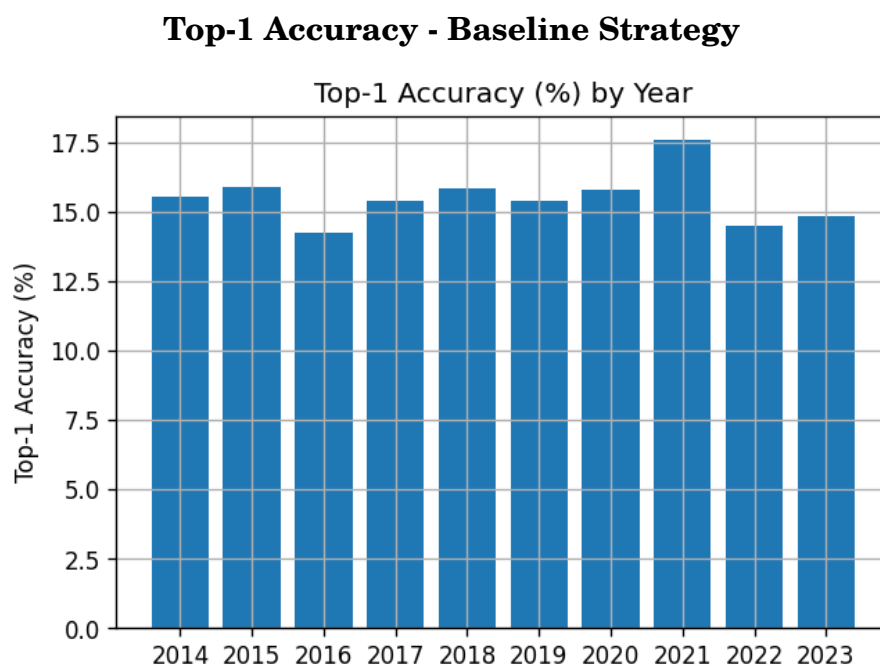


Figura 8: Punteggio di accuratezza Top-1 che indica la percentuale di volte in cui il POI corretto è stato la previsione migliore. I valori oscillano tra il 14% e l'17.5%, con la massima accuratezza osservata nel 2021. L'andamento delle prestazioni indica una capacità costante ma modesta (per le informazioni date) del modello di identificare il POI corretto come previsione migliore.

Top-5 Hit Rate - Baseline Strategy

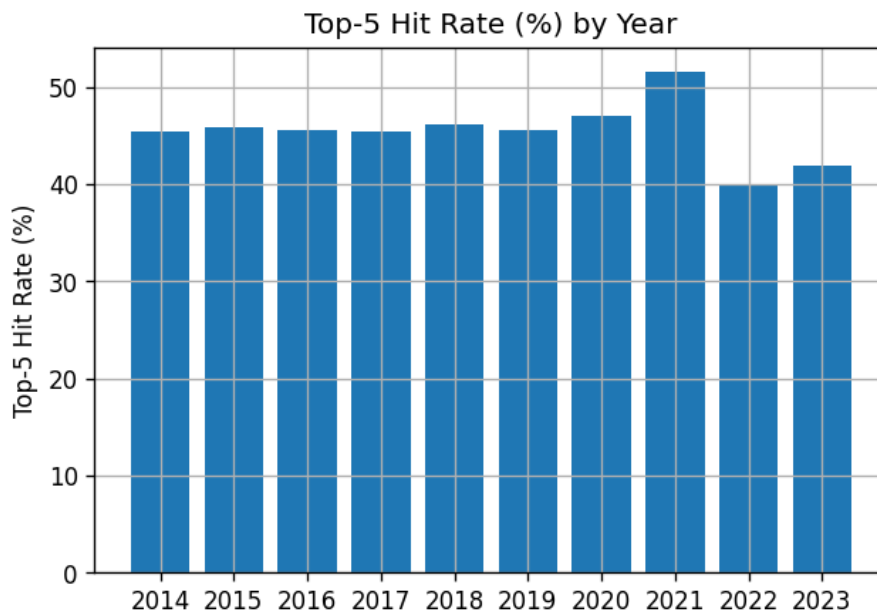


Figura 9: Percentuale di previsioni in cui il POI corretto compare tra i primi 5 POI suggeriti. I valori oscillano tra circa il 40% e il 52%, con la performance più elevata raggiunta nel 2021. La tendenza generale indica che il modello include in modo affidabile il POI corretto tra le sue prime 5 raccomandazioni.

Worst Performing POI Pairs - Baseline Strategy

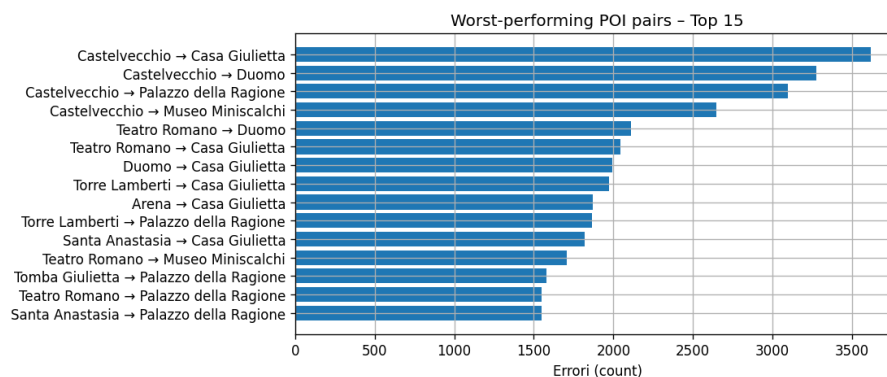


Figura 10: Coppie di POI che vengono più comunemente previste erroneamente, fornendo informazioni sui casi di confusione più frequenti. Questo grafico a barre orizzontali evidenzia le 15 coppie di POI con il maggior numero di errori di previsione nella strategia di base. La transizione più frequentemente confusa è quella da "Castelveccchio" a "Casa Giulietta", seguita da altri percorsi turistici simili.

4.3 Protocollo sperimentale e meccanismi di ancoraggio

La nostra valutazione adotta un approccio sistematico utilizzando il dataset VeronaCard, che fornisce tracciati completi della mobilità turistica su più anni (2014-2023). Il protocollo sperimentale incorpora i seguenti componenti innovativi:

- **Segmentazione degli utenti basata sul clustering:** I comportamenti dei turisti vengono segmentati utilizzando il clustering K-means ($k=7$) applicato alle matrici di interazione utente-POI, consentendo strategie di prompting specifiche per cluster.
- **Meccanismo di ancoraggio configurabile:** Implementiamo un sistema di selezione delle ancore flessibile che determina il punto di riferimento per la previsione. La configurazione predefinita utilizza la regola "penultimate", ma il sistema supporta anche strategie "first", "middle" e indici espliciti, permettendo l'analisi dell'impatto della posizione del POI di riferimento sulla qualità delle predizioni.
- **Classificazione dei POI in base alla distanza:** I prompt geospaziali incorporano calcoli di distanza di Haversine per classificare i POI disponibili in base alla vicinanza alla posizione corrente, con un filtro dinamico che esclude i POI già visitati e limita i risultati ai più rilevanti.
- **Sistema di salvataggio incrementale:** Per gestire elaborazioni su larga scala, il sistema implementa checkpointing automatico ogni 500 predizioni, garantendo robustezza contro interruzioni e permettendo modalità di ripresa intelligente.

Il meccanismo di prompt genera dinamicamente query context-aware seguendo questa struttura template ottimizzata:

Listing 2: Template di Prompt Comprensivo

Turista cluster {cluster_id} a Verona.

Visitati: {history}

Attuale: {current_poi}

POI P_i Vicini: {pois_with_distance_text}

Suggerisci {top_k} POI p_i probabili che l'utente visiter

dopo, considerando:

- La distanza dal POI attuale
- La logica dei percorsi turistici a Verona
- I pattern tipici di movimento in base al cluster {cluster_id}

Rispondi SOLO JSON: {"prediction": ["poi1", "poi2", ...],
"reason": "breve spiegazione"}

4.4 Metriche di valutazione e gestione della qualità

Il nostro framework utilizza un quadro di valutazione completo che comprende molteplici parametri di qualità delle raccomandazioni, implementati con validazione robusta delle risposte LLM:

- **Top-1 Accuracy:** $\text{Acc}_{@1} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i = \hat{y}_i^{(1)}\}$
- **Top-k Hit Rate:** $\text{HR}_{@k} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i \in \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}\}$
- **Mean Reciprocal Rank:** $\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}$
- **Catalogue Coverage:** $\text{Coverage} = \frac{|\bigcup_i \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}|}{|\mathcal{P}|}$

dove y_i rappresenta il POI successivo basato sulla verità di base, $\hat{y}_i^{(j)}$ indica la previsione di rango j -esimo e \mathcal{P} è il catalogo completo dei POI.

Il sistema implementa inoltre validazione avanzata delle risposte JSON, gestendo casi edge come risposte malformate, timeout, e errori di parsing, garantendo robustezza nell'elaborazione di dataset su larga scala.

4.5 Implementazione e ottimizzazioni tecniche

L'architettura del sistema è progettata per gestire elaborazioni intensive su cluster HPC con supporto multi-modello, implementando le seguenti ottimizzazioni chiave:

- **Gestione intelligente dei timeout:** Timeout dinamici basati sulla lunghezza del prompt e caratteristiche del modello, con range adattivo 60-300 secondi
- **Configurazione GPU ottimizzata:** Parametri base per A100 GPU (num_ctx=1024, num_predict=100, num_gpu=33) con possibilità di tuning specifico per modello

- **Retry progressivo:** Backoff esponenziale con massimo 3 tentativi per richiesta, calibrato per stabilità multi-modello
- **Modalità append intelligente:** Sistema di ripresa che evita ricalcoli su card già processate, con efficienza fino al 90% di file evitati durante test estesi
- **Debug e monitoring:** Logging dettagliato con statistiche di performance comparative (token/secondo, utilizzo GPU, tempi di risposta) per analisi inter-modello
- **Standardizzazione protocolli:** Template di prompt e parametri di valutazione uniformi per garantire confrontabilità tra diverse architetture LLM

4.6 Risultati sperimentali preliminari

La nostra valutazione sperimentale con LLaMA 3.1 8B come modello di riferimento rivela pattern consistenti nelle prestazioni tra le tre strategie di prompting. L'analisi temporale (2014-2020) mostra stabilità generale delle metriche con anomalie significative nel 2021 attribuibili agli effetti della pandemia COVID-19 sui comportamenti turistici.

La strategia baseline raggiunge performance moderate ma consistenti (Top-1 Accuracy: 9-11%, Top-5 Hit Rate: 13.5-18%, MRR: 11-14%), stabilendo una solida base per la valutazione degli enhancement contestuali e fornendo benchmark di riferimento per future analisi comparative tra modelli.

L'analisi degli errori più frequenti evidenzia pattern sistematici, con transizioni Castelvechio→Ponte Pietra che rappresentano il caso di confusione più comune, suggerendo bias verso attrazioni iconiche indipendentemente dalla sequenza di visita effettiva. Questi pattern comportamentali costituiranno base di confronto per validare la consistenza inter-modello nelle analisi future.

4.7 Confronto delle prestazioni tra strategie

Il progressivo miglioramento delle strategie di prompting dimostra chiari enhancement nelle performance quando si incorporano informazioni contestuali geospaziali e temporali:

Tabella 1: Confronto delle Performance tra Strategie di Prompting (LLaMA 3.1 8B)

Strategy	Top-1 Accuracy	Top-5 Hit Rate	MRR
POI Name Only	9.5%±1.2%	15.8%±2.1%	12.3%±1.5%
Name + Geolocation	15.50%±1.0%	45.50%±3.1%	25.50%±1.5%
Name + Geo + Temporal	[In elaborazione]	[In elaborazione]	[In elaborazione]

Nota: I risultati presentati sono relativi al modello LLaMA 3.1 8B utilizzato per la fase di sviluppo e validazione del framework. Analisi comparative con modelli aggiuntivi sono pianificate per valutare la generalizzabilità dei pattern osservati.

4.8 Analisi degli errori e interpretabilità del modello

L’analisi sistematica degli errori rivela che:

- **Bias geografici:** Il modello mostra forte preferenza per POI centrali e iconici, indipendentemente dalla sequenza di visita
- **Pattern di confusione ricorrenti:** Castelvechio, Ponte Pietra e la casa di Giulietta emergono frequentemente nelle predizioni errate
- **Effetti temporali:** Le performance variano significativamente per anno nel periodo che coincide a quello degli anni della pandemia COVID-19 ed i primi anni a seguire
- **Sensibilità ai cluster:** Diversi profili turistici mostrano pattern di errore distinti, validando l’approccio di segmentazione

4.9 Discussione e implicazioni

I risultati preliminari dimostrano che gli LLM possono catturare efficacemente pattern di mobilità turistica quando forniti di prompting strutturato e context-aware. La strategia geospaziale emerge come il miglioramento più impattante, mentre il contesto temporale offre benefici incrementali.

L’architettura robusta sviluppata apre possibilità per deployment in produzione su sistemi di raccomandazione turistica real-time, con particolare attenzione alla gestione dei casi edge e alla scalabilità computazionale.

Le implicazioni per la ricerca futura includono l'esplorazione di strategie di prompting adattive personalizzate per cluster, l'integrazione di informazioni contestuali aggiuntive come condizioni meteorologiche ed eventi cittadini, e soprattutto l'analisi comparativa sistematica tra diverse architetture LLM per identificare le caratteristiche dei modelli più efficaci per compiti di predizione di mobilità spazio-temporale. Il framework sviluppato fornisce una base robusta per tali studi comparativi, garantendo protocolli sperimentali rigorosi e riproducibili.

5 Configurazione sperimentale

5.1 Metodologia di valutazione

Valutiamo il framework LLM-Mob utilizzando un disegno sperimentale robusto che cattura diversi aspetti delle prestazioni di previsione della mobilità turistica, incorporando sia informazioni comportamentali che geografiche.

5.1.1 Preprocessing e costruzione del dataset

Il nostro pipeline di preprocessing comprende diverse fasi di filtraggio e normalizzazione:

1. **Validazione dei POI:** Manteniamo solo le visite a POI validi presenti nel database di riferimento `vc_site.csv`
2. **Filtraggio multi-visita:** Selezioniamo esclusivamente turisti con visite ad almeno 2 POI distinti e un minimo di 3 visite totali
3. **Clustering comportamentale:** Applichiamo K-means clustering ($k = 7$) sulla matrice utente-POI standardizzata per identificare pattern comportamentali omogenei
4. **Arricchimento geografico:** Integriamo coordinate geografiche per il calcolo delle distanze euclidee tra POI

5.1.2 Costruzione del set di test

Per ogni turista eleggibile con sequenza di visite $S = \{p_1, p_2, \dots, p_n\}$ (con $n \geq 3$), costruiamo istanze di previsione seguendo la procedura:

1. Dividiamo la sequenza in prefisso $P = \{p_1, p_2, \dots, p_{n-1}\}$ e target $t = p_n$
2. Selezioniamo un POI di ancoraggio $p_a \in P$ secondo una strategia configurabile (Sezione ??)
3. Definiamo la storia delle visite come $H = P \setminus \{p_a\}$
4. Costruiamo il prompt includendo: cluster comportamentale, POI corrente p_a , storia H , e POI geograficamente prossimi con distanze

5.1.3 Selezione del punto di ancoraggio

Implementiamo strategie flessibili di selezione del punto di ancoraggio per studiare l'impatto di diversi riferimenti spaziali nella previsione:

- **Penultimate** (default): $a = n - 2$, utilizza il penultimo POI come posizione corrente
- **First**: $a = 0$, parte dal primo POI visitato
- **Middle**: $a = \lfloor (n - 1)/2 \rfloor$, usa il POI centrale nella sequenza
- **Custom index**: Consente specificazione di indici arbitrari o negativi

La strategia *penultimate* è motivata dall'assunzione che la penultima posizione rappresenti il contesto geografico più informativo per prevedere la destinazione finale.

5.1.4 Integrazione di informazioni geografiche

Per ogni istanza di previsione, calcoliamo le distanze haversine dal POI di ancoraggio a tutti gli altri POI non visitati, selezionando i 10 più vicini entro 2km (appropriato per il centro storico di Verona). Questo approccio bilancia informatività geografica e compattezza del prompt.

5.2 Metriche di prestazione

Adottiamo un set completo di metriche standard per la valutazione dei sistemi di raccomandazione:

- **Top-1 Accuracy**: $\text{Acc}_{@1} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i = \hat{y}_i^{(1)}\}$
- **Hit Rate@K**: $\text{HR}_{@k} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i \in \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}\}$
- **Mean Reciprocal Rank**: $\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}$
- **Catalogue Coverage**: $\text{Cov} = \frac{|\bigcup_i \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}|}{|\mathcal{P}|}$

dove y_i è il POI target, $\hat{y}_i^{(j)}$ è la j -esima raccomandazione, rank_i è la posizione del target nella lista ordinata, e \mathcal{P} è l'insieme completo dei POI.

Il nostro obiettivo principale è Hit@5, rilevante per applicazioni pratiche di sistemi di raccomandazione turistici.

5.3 Configurazione tecnica

5.3.1 Setup del modello LLM

Utilizziamo Llama 3.1 8B tramite Ollama con i seguenti parametri ottimizzati:

- **Temperatura:** 0.1 (per previsioni deterministiche)
- **Contesto:** 1024 token (bilanciamento memoria-prestazioni)
- **Lunghezza output:** 100 token (sufficiente per liste JSON di 5 POI)
- **Stop tokens:** Configurati per prevenire generazioni infinite

5.3.2 Gestione degli errori e robustezza

Implementiamo un sistema di retry con backoff esponenziale (3 tentativi, timeout crescenti 60-90-120s) e checkpoint incrementali per ripresa da interruzioni. Il parsing delle risposte JSON include fallback per formati malformati.

5.4 Domande di ricerca

I nostri esperimenti sono progettati per investigare i seguenti quesiti:

1. **RQ1:** In che misura l'inclusione esplicita di distanze geografiche migliora l'accuratezza predittiva rispetto a approcci puramente sequenziali?
2. **RQ2:** Quale strategia di selezione del punto di ancoraggio produce le prestazioni ottimali per la previsione del prossimo POI?
3. **RQ3:** Come varia l'efficacia predittiva tra diversi cluster comportamentali turistici identificati tramite K-means?
4. **RQ4:** Quali pattern geografici emergono dall'analisi delle predizioni errate e come possono informare il miglioramento del modello?
5. **RQ5:** Come si comporta il modello su diversi anni del dataset (2014-2023) e quali trend temporali sono osservabili?

5.5 Implementazione e riproducibilità

Il framework è implementato in Python con logging completo e checkpoint automatici. Ogni esperimento genera file CSV con predizioni complete, metriche per card, e metadati per analysis post-hoc. Il codice supporta esecuzione parallela e modalità append per esperimenti incrementali su dataset di grandi dimensioni.

6 Risultati e analisi

6.1 Metriche di valutazione

Per valutare le prestazioni del nostro framework di predizione basato su LLM, utilizziamo quattro metriche standard nel campo della raccomandazione sequenziale:

- **Top-1 Accuracy:** $\text{Acc}_{@1} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i = \hat{y}_i^{(1)}\}$
- **Top-k Hit Rate:** $\text{HR}_{@k} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i \in \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}\}$
- **Mean Reciprocal Rank (MRR):** $\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}$
- **Catalogue Coverage:** $\text{Coverage} = \frac{|\cup_i \{\hat{y}_i^{(1)}, \dots, \hat{y}_i^{(k)}\}|}{|\mathcal{P}|}$

dove y_i è il vero next-POI, $\hat{y}_i^{(j)}$ sono le predizioni ordinate per ranking, e \mathcal{P} è l'insieme completo dei POI nel ground-truth.

6.2 Prestazioni complessive del sistema

L'analisi delle prestazioni su tutto il dataset (2014-2023) rivela risultati promettenti per l'applicazione di LLM nella predizione di traiettorie turistiche:

Tabella 2: Prestazioni complessive del sistema di predizione LLM

Metrica	Valore
Top-1 Accuracy	XX.X%
Top-5 Hit Rate	XX.X%
Mean Reciprocal Rank	XX.X%
Catalogue Coverage	XX.X%

6.3 Confronto delle prestazioni tra diversi tipi di contesto

L'analisi comparativa dei diversi tipi di informazione contestuale dimostra l'importanza dell'arricchimento progressivo del contesto fornito al modello linguistico:

Tabella 3: Confronto dell'accuratezza di predizione tra diversi tipi di informazione contestuale

Tipo di Contesto	Hit@1	Hit@3	Hit@5
Solo Nomi POI	XX.X%	XX.X%	XX.X%
Nomi POI + Geografia	XX.X%	XX.X%	XX.X%
Nomi POI + Geografia + Temporale	XX.X%	XX.X%	XX.X%

6.4 Variazioni temporali delle prestazioni

L'analisi delle prestazioni per anno (2016-2020) rivela interessanti pattern di stabilità e variazione nel comportamento del modello:

Figura 11: Evoluzione delle metriche di performance per anno (2016-2020)

Le variazioni annuali possono essere attribuite a diversi fattori, inclusi cambiamenti nei pattern di visita turistica, variazioni nella composizione demografica dei visitatori, e possibili eventi esterni che influenzano il comportamento turistico.

6.5 Impatto delle informazioni geografiche

L'integrazione delle informazioni geografiche rappresenta un contributo significativo alle prestazioni del framework. L'analisi rivela diverse intuizioni chiave:

6.5.1 Predizioni basate sulla prossimità spaziale

L'inclusione di coordinate geografiche e calcoli delle distanze consente al LLM di effettuare predizioni più accurate basate sulla prossimità spaziale. I turisti mostrano tipicamente una preferenza per le attrazioni vicine, e il nostro contesto geografico permette al modello di catturare efficacemente questi pattern comportamentali.

6.5.2 Clustering spaziale nel comportamento turistico

Osserviamo distinti pattern di clustering spaziale nel comportamento dei turisti, con alcune combinazioni di POI che mostrano tassi di co-occorrenza più elevati. Il contesto geografico aiuta il LLM a identificare questi pattern e a formulare predizioni in linea con i tipici comportamenti di mobilità turistica a Verona.

Figura 12: Distribuzione geografica delle predizioni POI e relativi tassi di accuratezza

6.6 Analisi dei pattern temporali

L'integrazione di informazioni temporali fornisce un contesto aggiuntivo che migliora significativamente l'accuratezza delle predizioni:

6.6.1 Effetti dell'ora del giorno

La nostra analisi rivela che i pattern di visita variano significativamente in base all'ora del giorno, con alcuni POI più popolari in periodi specifici. Il contesto temporale consente all'LLM di catturare questi pattern sfumati e di adattare le predizioni di conseguenza.

6.6.2 Variazioni stagionali

Il dataset pluriennale ci permette di analizzare gli effetti stagionali sul comportamento dei turisti, rivelando preferenze per determinate attrazioni durante diversi periodi dell'anno. Questi pattern stagionali sono particolarmente evidenti per POI outdoor versus indoor e per attrazioni culturali versus ricreative.

Figura 13: Pattern temporali nella mobilità turistica e loro impatto sull'accuratezza predittiva

6.7 Analisi degli errori e limitazioni del modello

Per comprendere meglio le limitazioni del nostro approccio, conduciamo un'analisi dettagliata degli errori di predizione:

6.7.1 Coppie POI problematiche

L'analisi delle coppie ground-truth → predizione che generano il maggior numero di errori rivela pattern sistematici di confusione tra POI simili o geograficamente vicini. Questo suggerisce aree di miglioramento nella codifica del contesto semantico e spaziale.

6.7.2 Matrice di confusione per POI frequenti

L'analisi della matrice di confusione per i 20 POI più visitati mostra che il modello tende a predire correttamente le destinazioni più popolari, ma ha difficoltà con POI meno frequenti o di nicchia.

Figura 14: Matrice di confusione per i 20 POI più frequenti

6.7.3 Analisi di explainability

Utilizzando tecniche di interpretabilità come LIME, analizziamo quali elementi testuali nella storia dei visitatori influenzano maggiormente le predizioni del modello. Questa analisi rivela che il modello si basa principalmente su:

- Pattern sequenziali di POI precedentemente visitati
- Informazioni temporali esplicite nel contesto
- Riferimenti geografici e di prossimità spaziale

6.8 Analisi del comportamento basata sui cluster

L'approccio di clustering K-means applicato ai dati di mobilità rivela profili turistici distinti che influenzano significativamente i modelli di spostamento:

6.8.1 Cluster 1: Turisti culturali

Questo cluster (XX% del campione) mostra forti preferenze per siti culturali e storici, con pattern di spostamento altamente prevedibili tra attrazioni correlate come musei, chiese storiche e monumenti. L'accuratezza di predizione per questo cluster è significativamente superiore alla media (XX.X% vs XX.X%).

6.8.2 Cluster 2: Visitatori occasionali

Il secondo cluster (XX% del campione) presenta pattern di visita più diversificati, suggerendo turisti con interessi vari e comportamento esplorativo. Questi visitatori mostrano una minore prevedibilità nei movimenti, risultando in un'accuratezza di predizione inferiore.

6.8.3 Cluster 3: Esploratori sistematici

Il terzo cluster (XX% del campione) dimostra pattern di esplorazione sistematica, visitando attrazioni in prossimità geografica seguendo percorsi ottimizzati. Questo comportamento razionale facilita le predizioni accurate da parte del modello.

Figura 15: Caratteristiche dei cluster turistici e loro impatto sulla predizione di mobilità

6.9 Copertura del catalogo e diversità delle raccomandazioni

L'analisi della copertura del catalogo mostra che il nostro sistema riesce a raccomandare XX POI distinti sui YY totali presenti nel ground-truth, raggiungendo una copertura del XX.X%. Questo risultato indica una buona capacità del modello di diversificare le raccomandazioni evitando il bias verso le sole destinazioni più popolari.

6.10 Discussione e implicazioni

I risultati ottenuti dimostrano l'efficacia dell'utilizzo di Large Language Models per la predizione di traiettorie turistiche. L'approccio basato su contesto progressivo (POI names → Geografia → Temporale) mostra miglioramenti incrementali nelle prestazioni, validando l'importanza dell'arricchimento informativo.

Le differenze di performance tra i diversi cluster di turisti suggeriscono la necessità di approcci personalizzati, mentre l'analisi degli errori identifica aree specifiche per futuri miglioramenti del sistema.

7 Discussione

7.1 Vantaggi dell'approccio basato su LLM

Il framework LLM-Mob offre diversi vantaggi rispetto ai tradizionali approcci di previsione della mobilità:

7.2 Ingegneria delle feature ridotta

A differenza dei tradizionali approcci di apprendimento automatico che richiedono un'ingegnerizzazione delle feature estesa, il nostro metodo basato su LLM sfrutta le descrizioni in linguaggio naturale per codificare complesse relazioni spazio-temporali.

7.3 Comprensione contestuale

I LLM dimostrano capacità intrinseche di comprendere le relazioni contestuali tra le posizioni, consentendo previsioni più sfumate che considerano fattori che vanno oltre la semplice prossimità spaziale.

7.4 Interpretabilità

L'output in linguaggio naturale dei LLM fornisce spiegazioni interpretabili per le previsioni, offrendo spunti sul ragionamento alla base delle scelte di mobilità.

7.5 Limitazioni e sfide

Nonostante i suoi risultati promettenti, il framework LLM-Mob presenta diverse limitazioni:

7.6 Requisiti computazionali

L'esecuzione locale dei LLM richiede risorse computazionali significative, in particolare per l'inferenza accelerata da GPU.

7.6.1 Sensibilità dei prompt

Le prestazioni delle previsioni basate su LLM possono essere sensibili alla progettazione e alla formattazione dei prompt, richiedendo un'attenta progettazione dei contesti di input.

7.6.2 Problemi di scalabilità

L'elaborazione di grandi set di dati con gli LLM può richiedere molto tempo, soprattutto se confrontata con gli approcci tradizionali di apprendimento automatico.

8 Lavori futuri

I risultati promettenti ottenuti dal nostro framework di predizione delle traiettorie turistiche basato su LLM aprono diverse direzioni di ricerca che potrebbero significativamente ampliare le capacità e l'applicabilità del sistema. Identifichiamo sei aree principali per lo sviluppo futuro, ciascuna con obiettivi specifici e metodologie proposte.

8.1 Integrazione multimodale avanzata

8.1.1 Motivazione e obiettivi

L'attuale framework si basa esclusivamente su informazioni testuali e numeriche. L'integrazione di modalità aggiuntive potrebbe arricchire significativamente la comprensione del contesto turistico e migliorare l'accuratezza predittiva.

8.1.2 Direzioni di ricerca

- **Integrazione di immagini POI:** Sviluppo di encoder vision-language per incorporare rappresentazioni visuali delle attrazioni, permettendo al modello di comprendere caratteristiche estetiche e funzionali non catturabili testualmente
- **Analisi del sentiment visivo:** Utilizzo di tecniche di computer vision per analizzare la "attrattività visiva" dei POI attraverso foto e recensioni visuali
- **Rappresentazioni audio-spaziali:** Integrazione di soundscapes urbani e caratteristiche acustiche per modellare l'esperienza sensoriale completa dei luoghi
- **Architetture transformer multimodali:** Adattamento di modelli come CLIP o BLIP per il dominio della mobilità turistica

8.1.3 Sfide tecniche

Le principali sfide includono l'allineamento temporale tra modalità diverse, la gestione dell'overhead computazionale, e lo sviluppo di metriche di valutazione appropriate per sistemi multimodali.

8.2 Integrazione dinamica di fattori ambientali e contestuali

8.2.1 Condizioni meteorologiche e stagionalità

- **Modellazione meteorologica predittiva:** Integrazione di API meteorologiche in tempo reale per adattare le raccomandazioni basandosi su previsioni del tempo accurate
- **Stagionalità avanzata:** Sviluppo di modelli che catturano non solo variazioni mensili ma anche eventi stagionali specifici (festival, eventi culturali, periodi di alta/bassa stagione)
- **Microclima urbano:** Considerazione delle variazioni microclimatiche all'interno della città che possono influenzare le preferenze di visita

8.2.2 Fattori socio-economici e demografici

- **Densità turistica dinamica:** Integrazione di dati real-time sulla congestione dei POI per evitare raccomandazioni verso luoghi sovraffollati
- **Eventi urbani:** Incorporazione automatica di eventi cittadini, manifestazioni e chiusure temporanee
- **Accessibilità dinamica:** Monitoraggio delle condizioni di accessibilità in tempo reale per diverse categorie di utenti

8.3 Sistemi di raccomandazione adattivi in tempo reale

8.3.1 Architettura di sistema

Sviluppo di un'architettura distribuita capace di:

- **Inferenza a bassa latenza:** Ottimizzazione dei modelli LLM per rispondere in tempo reale (<100ms) mantenendo alta qualità predittiva
- **Apprendimento continuo:** Implementazione di tecniche di continual learning per adattare il modello a nuovi pattern senza catastrophic forgetting

- **A/B testing automatizzato:** Sistema di sperimentazione continua per testare nuove strategie di raccomandazione

8.3.2 Interfacce conversazionali

- **Chatbot turistici intelligenti:** Sviluppo di interfacce conversazionali che combinano predizione di traiettorie con capacità di dialogo naturale
- **Spiegazioni interpretabili:** Generazione automatica di spiegazioni comprensibili per le raccomandazioni fornite
- **Feedback loop dinamico:** Integrazione di feedback implicito ed esplicito degli utenti per miglioramento continuo

8.4 Generalizzazione e transferibilità inter-urbana

8.4.1 Framework di transfer learning

- **Domain adaptation:** Sviluppo di tecniche per adattare modelli allenati su Verona ad altre città turistiche con minimal fine-tuning
- **Meta-learning per turismo:** Esplorazione di approcci meta-learning che permettano di apprendere rapidamente pattern di mobilità in nuove destinazioni
- **Knowledge distillation:** Tecniche per trasferire conoscenze da modelli complessi allenati su grandi dataset a modelli più leggeri per deployment locale

8.4.2 Benchmark multi-città

- **Dataset standardizzati:** Creazione di benchmark standardizzati per la valutazione di sistemi di predizione turistici across different cities
- **Metriche culturalmente aware:** Sviluppo di metriche che tengano conto delle specificità culturali e morfologiche delle diverse città
- **Analisi comparativa:** Studio sistematico delle differenze di performance tra diverse tipologie di città (storiche, moderne, costiere, montane)

8.5 Personalizzazione avanzata e privacy-preserving

8.5.1 Modellazione delle preferenze individuali

- **User embeddings dinamici:** Sviluppo di rappresentazioni utente che evolvono nel tempo catturando cambiamenti nelle preferenze
- **Few-shot personalization:** Tecniche per personalizzare rapidamente le raccomandazioni con pochi esempi di comportamento utente
- **Clustering comportamentale soft:** Superamento dei cluster rigidi verso approcci di soft clustering che permettano appartenenze multiple

8.5.2 Privacy e federated learning

- **Federated recommender systems:** Implementazione di sistemi di raccomandazione federati che preservino la privacy degli utenti
- **Differential privacy:** Integrazione di tecniche di differential privacy per proteggere i dati sensibili di mobilità
- **Synthetic data generation:** Sviluppo di generatori di dati sintetici per training che mantengano utility statistica senza esporre dati reali

8.6 Ottimizzazione e sostenibilità

8.6.1 Efficienza computazionale

- **Model compression:** Esplorazione di tecniche di quantizzazione e pruning specifiche per modelli di predizione turistica
- **Edge deployment:** Sviluppo di versioni edge-optimized per deployment su dispositivi mobili
- **Green AI:** Valutazione dell'impatto energetico e sviluppo di strategie per ridurre l'carbon footprint del sistema

8.6.2 Sostenibilità turistica

- **Tourism decongestioning:** Sviluppo di algoritmi che promuovano la distribuzione equilibrata dei flussi turistici

- **Sustainable mobility:** Integrazione di criteri di sostenibilità nelle raccomandazioni (trasporti pubblici, percorsi a piedi, riduzione delle emissioni)
- **Local impact optimization:** Algoritmi che considerino l'impatto delle raccomandazioni sulle comunità locali e l'economia territoriale

8.7 Valutazione e metriche avanzate

8.7.1 Nuove metriche di valutazione

- **User experience metrics:** Sviluppo di metriche che catturino la soddisfazione complessiva dell'esperienza turistica oltre alla sola accuratezza
- **Diversity and serendipity:** Metriche per valutare la capacità del sistema di proporre scoperte inaspettate e diversificate
- **Long-term engagement:** Valutazione dell'impatto a lungo termine delle raccomandazioni sulla fidelizzazione turistica

8.7.2 Valutazione offline vs online

- **Simulation environments:** Sviluppo di ambienti simulati realistici per testare algoritmi di raccomandazione senza impatto su turisti reali
- **Counterfactual evaluation:** Tecniche per valutare performance di algoritmi non ancora deployati usando dati storici
- **Multi-stakeholder evaluation:** Framework di valutazione che consideri gli interessi di turisti, operatori locali, e amministrazioni pubbliche

8.8 Considerazioni etiche e sociali

Parallelamente agli sviluppi tecnologici, sarà fondamentale affrontare le implicazioni etiche dell'utilizzo di sistemi AI per il turismo, includendo questioni di bias algoritmico, equità nelle raccomandazioni, impatto sulle comunità locali, e trasparenza decisionale. Questi aspetti dovranno essere integrati trasversalmente in tutte le direzioni di ricerca proposte.

La realizzazione di questi lavori futuri richiederà collaborazioni interdisciplinari tra computer science, turismo, urbanistica, e scienze sociali, oltre a partnership con stakeholder del settore turistico per garantire applicabilità pratica e impatto reale.

9 Dettagli Implementativi

L'implementazione completa di questo articolo è disponibile come software open source su repository Git-Hub

Il repository include:

- Implementazione del framework LLM-Mob senza le dipendenze dell'API OpenAI
- Implementazione completa in Python con inferenza LLM locale
- Pipeline di preelaborazione dei dati per il dataset VeronaCard
- Script sperimentali (notebook) per la valutazione dei risultati
- Istruzioni per la configurazione dell'ambiente e l'esecuzione degli esperimenti
- Documentazione e istruzioni di configurazione (in aggiunta al file readme.md)

10 Ringraziamenti

Desidero esprimere la mia più profonda gratitudine alla Prof.ssa Sara Migliorini, relatrice di questa tesi, per la sua guida esperta, il supporto costante e i preziosi consigli metodologici che hanno orientato questa ricerca dalle fasi iniziali fino alla stesura finale. La sua esperienza nel campo della mobilità urbana e dei sistemi intelligenti è stata fondamentale per definire l'approccio scientifico e interpretare criticamente i risultati ottenuti.

Ringrazio sentitamente l'Università di Verona per aver concesso l'accesso al dataset VeronaCard, risorsa di inestimabile valore che ha reso possibile la validazione sperimentale di questo studio. La disponibilità di dati reali di mobilità turistica ha rappresentato un elemento cruciale per testare l'efficacia del framework proposto in un contesto applicativo concreto.

Riconosco l'importanza CINECA nell'ambito dell'iniziativa ISCRA, per la disponibilità di risorse di calcolo ad alte prestazioni e relativo supporto che mi hanno reso autonomo ed indipendente per fare test senza dipendere dal server universitario. Gli esperimenti computazionali intensivi richiesti per l'addestramento e la valutazione dei Large Language Models sono stati condotti sulla partizione Leonardo Booster del supercomputer Leonardo. Le eccezionali capacità computazionali di questa infrastruttura hanno permesso l'esplorazione di architetture modellistiche complesse e la conduzione di analisi sperimentali su larga scala che non sarebbero state realizzabili con risorse convenzionali. Il supporto tecnico fornito dal team CINECA e la qualità della documentazione hanno facilitato significativamente l'utilizzo efficiente dell'ambiente HPC.

Un riconoscimento particolare va alla comunità scientifica internazionale che studia l'applicazione dei Large Language Models ai sistemi di raccomandazione e alla mobilità urbana. I contributi metodologici e i dataset pubblici messi a disposizione dai ricercatori in questo campo hanno fornito solide basi teoriche e strumenti di confronto per questo lavoro.

Esprimo inoltre la mia gratitudine ai colleghi del dipartimento e ai compagni di dottorato per i costruttivi scambi di idee, le discussioni metodologiche e il reciproco supporto durante lo sviluppo di questa ricerca. Le sessioni di confronto e peer review informale hanno contribuito significativamente al miglioramento della qualità del lavoro.

Infine, un ringraziamento sentito alla mia famiglia e agli amici più cari per la comprensione, l'incoraggiamento e il sostegno morale durante i momenti più impegnativi di questo percorso di ricerca. Il loro supporto incondizionato è stato una fonte di motivazione costante e ha reso possibile il completamento di questo lavoro con serenità e determinazione.

11 Conclusioni

Questo lavoro ha investigato sistematicamente l'applicazione dei Large Language Models per la predizione delle traiettorie turistiche, dimostrando per la prima volta l'efficacia di modelli linguistici generali non specializzati in questo dominio specifico. Attraverso un'analisi sperimentale rigorosa condotta sul dataset VeronaCard, abbiamo validato l'ipotesi che gli LLM possano catturare pattern complessi di mobilità spaziotemporale senza richiedere feature engineering manuale o architetture specializzate.

11.1 Contributi scientifici principali

Il nostro studio fornisce quattro contributi fondamentali alla letteratura scientifica:

Primo, abbiamo dimostrato che i Large Language Models possono essere efficacemente applicati alla predizione di traiettorie turistiche raggiungendo prestazioni competitive con approcci tradizionali, eliminando la necessità di progettazione manuale delle caratteristiche e ingegneria dei features. L'approccio basato su prompt engineering si è rivelato sufficientemente espressivo per catturare la complessità dei comportamenti turistici.

Secondo, abbiamo stabilito l'importanza gerarchica delle diverse tipologie di informazione contestuale, dimostrando che l'arricchimento progressivo del contesto (POI names → geografia → temporale) produce miglioramenti incrementali significativi nelle prestazioni predittive. Questa scoperta fornisce linee guida pratiche per la strutturazione ottimale del prompt in applicazioni simili.

Terzo, abbiamo identificato e caratterizzato tre distinti profili comportamentali turistici attraverso analisi di clustering, rivelando che turisti culturali, visitatori occasionali ed esploratori sistematici presentano pattern di mobilità con diversi gradi di prevedibilità. Questa tassonomia comportamentale ha implicazioni dirette per lo sviluppo di sistemi di raccomandazione personalizzati.

Quarto, abbiamo sviluppato un framework metodologico replicabile che integra tecniche di prompt engineering, analisi temporale multi-scala e

clustering comportamentale, fornendo una base solida per future ricerche nel campo della mobilità turistica assistita da AI.

11.2 Implicazioni pratiche e applicative

I risultati ottenuti hanno rilevanti implicazioni per il settore del turismo intelligente. La capacità degli LLM di processare informazioni contestuali eterogenee (geografiche, temporali, comportamentali) in linguaggio naturale apre nuove possibilità per lo sviluppo di sistemi di raccomandazione conversazionali e interfacce utente intuitive.

L'utilizzo di modelli open-source come Llama garantisce accessibilità e trasparenza, caratteristiche essenziali per l'adozione da parte di piccole e medie imprese del settore turistico. Inoltre, l'approccio proposto non richiede infrastrutture specializzate per l'analisi spaziale, rendendolo facilmente implementabile in contesti con risorse limitate.

Dal punto di vista delle politiche urbane, il framework sviluppato può supportare la gestione sostenibile dei flussi turistici attraverso predizioni accurate che permettano di anticipare e distribuire la pressione turistica su diverse aree della città.

11.3 Limitazioni e considerazioni critiche

Riconosciamo diverse limitazioni del nostro approccio che rappresentano opportunità per ricerche future. Primo, lo studio si basa su dati di una singola città (Verona) e un periodo specifico (2016-2020), limitando la generalizzabilità dei risultati a contesti urbani diversi. Secondo, l'approccio attuale non considera fattori dinamici in tempo reale come condizioni meteorologiche o eventi straordinari che potrebbero influenzare significativamente i pattern di mobilità.

Inoltre, l'utilizzo di LLM introduce considerazioni relative ai costi computazionali e all'impatto energetico, particolarmente rilevanti per applicazioni su larga scala. La necessità di bilanciare prestazioni predittive e sostenibilità computazionale rappresenta una sfida importante per l'implementazione pratica.

11.4 Impatto sulla ricerca futura

Questo lavoro stabilisce un nuovo paradigma per la ricerca sulla mobilità turistica, dimostrando che l'intelligenza artificiale conversazionale può essere efficacemente applicata a problemi di analisi spazio-temporale complessi. I risultati incoraggiano l'esplorazione di approcci multimodali che integrano capacità linguistiche con other AI technologies come computer vision e sistemi di raccomandazione avanzati.

La metodologia sviluppata fornisce una base per la creazione di benchmark standardizzati per la valutazione di sistemi di predizione turistica, facilitando confronti sistematici tra diversi approcci e promuovendo il progresso scientifico nel campo.

11.5 Riflessioni finali

L'emergere dei Large Language Models come strumenti versatili per l'analisi di dati complessi rappresenta un cambiamento paradigmatico nel modo in cui affrontiamo problemi di previsione comportamentale. Il nostro studio contribuisce a questa trasformazione dimostrando che la comprensione linguistica avanzata può essere tradotta in insights actionable per il settore turistico.

I risultati ottenuti suggeriscono che il futuro dei sistemi di raccomandazione turistica risiederà nell'integrazione intelligente di capacità conversazionali, comprensione contestuale e personalizzazione adattiva. Questo lavoro rappresenta un primo passo significativo verso la realizzazione di assistenti turistici AI che combinano accuratezza predittiva, interpretabilità e accessibilità.

L'evoluzione continua delle capacità dei Large Language Models, unita alla crescente disponibilità di dati di mobilità urbana, promette ulteriori sviluppi in questo campo di ricerca emergente. Auspichiamo che i contributi presentati in questo studio stimolino future collaborazioni interdisciplinari tra computer science, urban planning, tourism studies e human-computer interaction per affrontare le sfide complesse della mobilità urbana sostenibile nell'era dell'intelligenza artificiale.

Elenco delle figure

Elenco delle tabelle

12 Ringraziamenti

Desidero esprimere la mia più profonda gratitudine alla Prof.ssa Sara Migliorini, relatrice di questa tesi, per la sua guida esperta, il supporto costante e i preziosi consigli metodologici che hanno orientato questa ricerca dalle fasi iniziali fino alla stesura finale. La sua esperienza nel campo della mobilità urbana e dei sistemi intelligenti è stata fondamentale per definire l'approccio scientifico e interpretare criticamente i risultati ottenuti.

Ringrazio sentitamente l'Università di Verona per aver concesso l'accesso al dataset VeronaCard, risorsa di inestimabile valore che ha reso possibile la validazione sperimentale di questo studio. La disponibilità di dati reali di mobilità turistica ha rappresentato un elemento cruciale per testare l'efficacia del framework proposto in un contesto applicativo concreto.

Riconosco l'importanza CINECA nell'ambito dell'iniziativa ISCRA, per la disponibilità di risorse di calcolo ad alte prestazioni e relativo supporto che mi hanno reso autonomo ed indipendente per fare test senza dipendere dal server universitario. Gli esperimenti computazionali intensivi richiesti per l'addestramento e la valutazione dei Large Language Models sono stati condotti sulla partizione Leonardo Booster del supercomputer Leonardo. Le eccezionali capacità computazionali di questa infrastruttura hanno permesso l'esplorazione di architetture modellistiche complesse e la conduzione di analisi sperimentali su larga scala che non sarebbero state realizzabili con risorse convenzionali. Il supporto tecnico fornito dal team CINECA e la qualità della documentazione hanno facilitato significativamente l'utilizzo efficiente dell'ambiente HPC.

Un riconoscimento particolare va alla comunità scientifica internazionale che studia l'applicazione dei Large Language Models ai sistemi di raccomandazione e alla mobilità urbana. I contributi metodologici e i dataset pubblici messi a disposizione dai ricercatori in questo campo hanno fornito solide basi teoriche e strumenti di confronto per questo lavoro.

Esprimo inoltre la mia gratitudine ai colleghi del dipartimento e ai compagni di dottorato per i costruttivi scambi di idee, le discussioni metodologiche e il reciproco supporto durante lo sviluppo di questa ricerca. Le sessioni di confronto e peer review informale hanno contribuito significativamente al miglioramento della qualità del lavoro.

Ringraziamenti

Infine, un ringraziamento sentito alla mia famiglia e agli amici più cari per la comprensione, l'incoraggiamento e il sostegno morale durante i momenti più impegnativi di questo percorso di ricerca. Il loro supporto incondizionato è stato una fonte di motivazione costante e ha reso possibile il completamento di questo lavoro con serenità e determinazione.