

STATE-OF-THE-ART ML-DRIVEN TOURISM FORECASTING AND INSIGHT SYSTEM WITH ACTIONABLE FOOTFALL AND REVENUE MAXIMIZATION

Project Reference No.: 49S_MSC_0235

College : St. Joseph's University, Bengaluru
Branch : Department of Computer Science
Guide(s) : Dr. B G Prasanthi
: Dr. B Nithya
Student(s) : Mr. Deemanth S
: Mr. B K Kaushik Kumar
: Mr. S Syed Zaffer
: Mr. Vishwas Madhak

Keywords:

AI-Powered Decision Intelligence, Tourism Analytics Engine, Retrieval-Augmented Generation (RAG), Aspect-Based Sentiment Intelligence, Semantic Vector Search, Real-Time Insight Generation, Smart Governance Systems

Introduction:

1. Tourism plays a crucial role in Karnataka's economic growth, but decision-making is often limited by unstructured and underutilized data such as online reviews.
2. This project presents an AI-powered Tourism Insight Assistant that transforms large-scale tourist feedback into structured, actionable insights. By leveraging Retrieval-Augmented Generation (RAG), semantic vector search, and aspect-based sentiment analysis, the system enables a deeper understanding of tourist experiences across key factors like cleanliness, safety, and accessibility.
3. The proposed solution bridges the gap between raw data and informed decision-making, offering a scalable and intelligent platform for enhancing tourism management and visitor experience.

Objectives:

1. To develop an AI-powered tourism insight engine for transforming large-scale review data into decision-ready intelligence
2. To implement a Retrieval-Augmented Generation (RAG) architecture for accurate, context-aware insight generation
3. To perform aspect-level sentiment intelligence across key tourism factors such as cleanliness, safety, and accessibility
4. To deliver actionable, data-driven insights that enable effective tourism management and service improvement

Methodology:

The proposed system follows a modular AI-driven architecture designed to transform unstructured tourism data into structured insights.

1. Tourism reviews are collected and preprocessed to remove noise and convert them into meaningful text segments. These segments are transformed into vector embeddings using advanced language models, enabling semantic understanding of the data. The embeddings are stored in a vector database (Qdrant) to support efficient and relevant information retrieval.
2. A Retrieval-Augmented Generation (RAG) pipeline is implemented, where contextually relevant data is retrieved and passed to a Large Language Model (LLM) to generate accurate and coherent insights. In parallel, an aspect-level sentiment intelligence module evaluates key tourism parameters such as cleanliness, safety, accessibility, and crowd conditions.
3. The generated insights are delivered through an interactive analytics dashboard, allowing users to query specific locations and receive structured outputs including summaries, sentiment scores, and key thematic tags. This architecture ensures scalability, efficiency, and real-time insight generation for practical deployment.

Result and Conclusion:

1. The developed system successfully demonstrates the ability to transform large-scale, unstructured tourism reviews into structured and actionable insights. By integrating Retrieval-Augmented Generation (RAG) with aspect-level sentiment intelligence, the system delivers accurate, context-aware understanding of tourist feedback across multiple dimensions.
2. The functional prototype validates the feasibility of real-world deployment, enabling users to query locations and obtain instant insights through an interactive dashboard. This approach significantly enhances the ability of tourism authorities to identify key issues, prioritize improvements, and optimize overall visitor experience.
3. In conclusion, the proposed solution establishes a scalable and intelligent framework for tourism analytics, bridging the gap between raw data and informed decision-making, with strong potential for practical implementation.

References:

- P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.
- T. Mikolov, I. Sutskever, K. Chen, et al., "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.
- Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Future Scope:

The future scope of this project includes:

1. Evolution into a real-time tourism intelligence infrastructure, enabling continuous monitoring and adaptive decision-making at scale
2. Transformation into a nationwide decision intelligence platform, extendable across multiple states, sectors, and public service domains
3. Deep integration with smart governance ecosystems, leveraging live data streams for proactive, data-driven policy and operational optimization