Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 4, 146-159
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i4.12575

Intelligent tourist attractions recommender system with hybrid AI

Viknesh Kumar[1], Boon Chin Yeo[1*], Way Soong Lim[1]
[1]Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia; bcyeo@mmu.edu.my (B.C.Y.).

Abstract: The annual rise in tourism increases demand for a robust tourism recommendation system. Big data processing has been adopted as a key component of adaptive recommendation systems, which provide services designed to suit individual users' needs. Big data processing can be applied to tourism recommendations to offer better options for tourists and reduce the distance traveled by users to visit a location of their preference. In this system, Hybrid AI, Llama, and Logistic Regression are used to categorize big data and recommend tourism types that match the user's personality, based on sentiment analysis, distance, and cost of visit. Hybrid AI combines machine learning, which uses statistical models to analyze data, with Llama AI to provide insights and meanings related to locations presented to the user. The system is scalable to handle the complexity of the data collected. It informs users of locations with the lowest Mean Squared Error through cumulative values from different tourism criteria. The current best cumulative RMSE, MAE, and MSE achieved by the system are 0.909, 0.732, and 2.773, respectively, for a training dataset of 438 entries. The MSE is expected to improve with larger datasets for training.
Keywords: *AI, Big data, Deep learning, Recommendation system, Sentiment analysis.*

1. Introduction

Global tourism is rising to become a trillion-dollar industry, contributing to 7% of global exports and significantly impacting the global gross domestic product (GDP). International arrivals and tourism receipts are increasing by 3–5% annually, outpacing the growth of international trade. In 2016, these figures exceeded 1 billion arrivals and US$1.2 trillion in receipts [1]. A tourism recommendation system provides travel destinations, activities, accommodations, and other related services to users based on their preferences, past behaviors, and other factors. These systems are designed to offer personalized and relevant recommendations, considering users' unique interests and requirements. In this study, we will develop a tourism recommendation system that uses big data to offer better options for tourists with shorter travel distances. Due to the rapid development of technologies such as artificial intelligence and the mobile Internet, the tourism industry is becoming more integrated with the Information and Communication Technology (ICT) industry. User data is gathered, processed, and managed as big data to provide tailored services [2, 3]. Recommendation systems process data such as a user's browsing history or interests based on various sets of algorithms. In this proposal, a recommendation system that uses big data and analyzes the data using a Deep Learning classification method will be devised. Deep Learning is a branch of machine learning and a classification method to analyze big data of diverse data types [4-6]. The proposed system will be multi-layered and will provide an adaptive system capable of handling diverse datasets. The system will focus on its efficiency and scalability. Deep Learning will be integrated as a tool to help overcome rigid and singular data processing methods of the past [7-9]. In recent years, many scholars have applied Deep Learning to the tourism industry. One such scholar used Deep Learning to propose a spatially-aware hierarchical

History: Received: 7 January 2026; Revised: 16 February 2026; Accepted: 19 February 2026; Published: 31 March 2026
* Correspondence: bcyeo@mmu.edu.my

collaborative deep learning model [10]. The model employs Deep Learning to analyze personalized information from the perspective of heterogeneous features to layer spatially-aware personal preferences of users. This effectively eliminated the problem of data infrequency in travel service recommendation [10]. This study aims to use predictive models similar to those being researched by other scholars [11, 12] to facilitate global tourism. The further improvement of this technology could help tourists save time needed to browse information regarding points of interest by scavenging tourism blogs, forums, and websites [13]. A highly dependable recommendation system can overcome the overflow of information that occurs on the internet due to a lack of focus on isolating it on a single platform [13]. The goal of this research is to propose a suitable tourism recommendation method to provide personalized tourism information to its users while reducing the distance traveled by users by analyzing sentiments and travel preferences.

2. Problem Statement and Objective

2.1. Problem Statement

Long-distance travel presents significant challenges to a traveler's health and well-being. Traveling across multiple time zones takes a toll on the body, with the most common issue being jet lag, a medically recognized sleep disorder caused by the disruption of the body's internal clock, or circadian rhythm. This can lead to debilitating fatigue, insomnia, and digestive problems. Furthermore, sitting for long hours in a cramped airplane cabin increases the risk of serious health conditions like deep vein thrombosis (DVT), causes dehydration, and can heighten exposure to germs. The mental stress of navigating unfamiliar airports, facing potential delays, and overcoming language barriers can also cause significant anxiety and exhaustion. Beyond the physical effects, the logistical complexity of long-distance travel requires meticulous preparation, where a single oversight can derail a trip. The process goes far beyond just booking a flight; it often involves navigating complicated bureaucratic hurdles to obtain necessary visas, which can be both time-consuming and uncertain. Extensive planning is also required to secure suitable accommodation, arrange local transportation, and pack appropriately for different climates and cultural norms. Essential tasks like purchasing comprehensive travel insurance, managing foreign currency, and researching local customs and safety precautions are also crucial for a smooth journey. Finally, the financial and environmental costs of long-distance travel are substantial. While airfare is a major expense, the total cost quickly accumulates with accommodation, food, and activities. This high financial burden is directly linked to a steep environmental one, as air travel is one of the most carbon-intensive activities a person can undertake. The environmental impact is also worsening; a key study found that transport-related emissions from tourism surged by more than 60% between 2005 and 2016 alone. This creates a modern dilemma for travelers, who must weigh their desire to explore the world against the journey's considerable ecological footprint [14].

2.2. Objectives

This research work is guided by three key objectives. The primary goal is to develop a hybrid recommender system that integrates Llama and Logistic Regression to suggest the nearest tourism locations based on user interests, thereby reducing the need for long-distance travel. Concurrently, a sequential recommender system using only Logistic Regression will also be developed. Finally, the research will thoroughly investigate and compare the performance of these systems using various evaluation metrics.

3. Literature Review

Multiple systems were analyzed to understand the software architecture of tourism recommenders. Below are the findings based on different research

ISSN: 2576-8484
Vol. 10, No. 4: 146-159, 2026
DOI: 10.55214/2576-8484.v10i4.12575

3.1. Hotel Recommendation System

This system uses deep learning and sentiment analysis. This study recommends hotels based on users' explicit and implicit preferences, which increases its prediction accuracy. In this study, a combination of sentiment analysis with Collaborative Filtering (CF) based on deep learning is used [15]. The system also uses Natural Language Processing (NLP) and supervised classification methods to analyze sentiments and extract implicit features. The system takes advantage of Singular Value Decomposition (SVD) to improve scalability [15]. The hybrid model adopted in this study is advantageous because it overcomes the sparsity issue carried by CF, but it also has a drawback where the system is now more complex and increases the difficulty in implementation [15].

3.2. CF-based Recommendation System

A comparison is made with a Memory-Based Collaborative Filtering system in this study [16]. The system is easy to implement, and adding data is simpler. New data can be added incrementally, and it does not consider the content of the items recommended [16]. The disadvantages, however, outweigh the pros of this system. Memory-based systems heavily depend on human-guided output. The system constantly needs supervision for accurate output. Another issue with Memory-Based CF is sparsity in the data matrix [16]. This further reduces accuracy. Finally, the system lacks scalability for different types of dataset input.

3.3. Model-Based Recommendation System

A study compares a Model-Based Collaborative Filtering system, which better addresses scalability and sparsity issues. It was observed to have better accuracy than memory-based systems. The drawback is the need to trade off between scalability and accuracy [17]. Inability to balance these factors contributed to poor performance. Loss of information was also observed during the Singular Value Decomposition (SVD) reduction technique [17]. The overall accuracy of the hybrid CF system with Linear Support Vector Classification is 86.5%. The precision is 91.7%, and the recall rate is 89.3% [17]. The same system with a Multinomial Naïve Bayes yields an accuracy of 85.9%, precision of 91.1%, and a recall rate of 87.3% [17].

3.4. Deep Learning Based Sentiment Analysis

Another research, a Deep Learning system to predict sentiments in tourism analysis, uses deep-learning techniques to classify comments that tourists publish online to help new tourists decide how best to plan their trips. In this research, multiple classification methods are used to determine the sentiments reflected on the http://booking.com and http://tripadvisor.com platforms for the service received in hotels [18]. The study analyzes various classifiers such as Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM). These classifiers were trained and validated with data from hotels located on the island of Tenerife. The research finds LSTM to be the most accurate and robust [18]. The research describes the optimum result for CNN to be attained with a single convolutional layer and 64 channels. The CNN classifier is susceptible to overfitting if more layers or channels are added. The LSTM neural networks yield higher accuracies at a vector length of 300 for the internal state, resulting in an accuracy just over 89% [18]. The research also depicts that lower or higher vector lengths than 300 for LSTM result in lower accuracy [18].

3.5. RNN and CNN Comparative Study for Recommendation

The analysis in this research reiterates that deep learning outweighs other machine learning classifiers when conducting sentiment analysis. The study explains that the high accuracy of the Recurrent Neural Network (RNN) was contributed by the Long Short-Term Memory (LSTM) model, which helped surpass the performance of traditional RNNs [19]. A problem with traditional RNNs, which was overcome, is that the gradients from the objective function can vanish or explode after a few iterations of multiplying the weights of the network [19]. LSTMs are described as better suited to this task due to the presence of input gates, forget gates, and output gates, which control the flow of information through the network. This aided the RNN to marginally surpass the complex CNN classification method in this research.

3.6. Contextual Technology's Role in Sustainable Tourism

The paper argues that by leveraging diverse data sources, these contextual systems can provide tourists with personalized recommendations that align with the core dimensions of sustainability: environmental, economic, and socio-cultural. The study identifies four key criteria for successful implementation: enhancing community well-being, protecting the natural and cultural environment, maintaining high-quality tourist experiences, and enabling effective management and monitoring [20]. A similar methodology involving cost, distance, and user sentiment was devised for this research.

4. Methodology

The recommendation system was constructed using the KNIME platform. The components used for data analytics are based on the Apache Spark library. Three criteria were selected for location recommendations: distance, price, and rating. Below is the flowchart of the current algorithm. The sample data for training and prediction were obtained from the Kaggle website. The methodology involved creating two separate systems: a baseline model (Model 1) and an enhanced model (Model 2), to clearly quantify performance improvements.

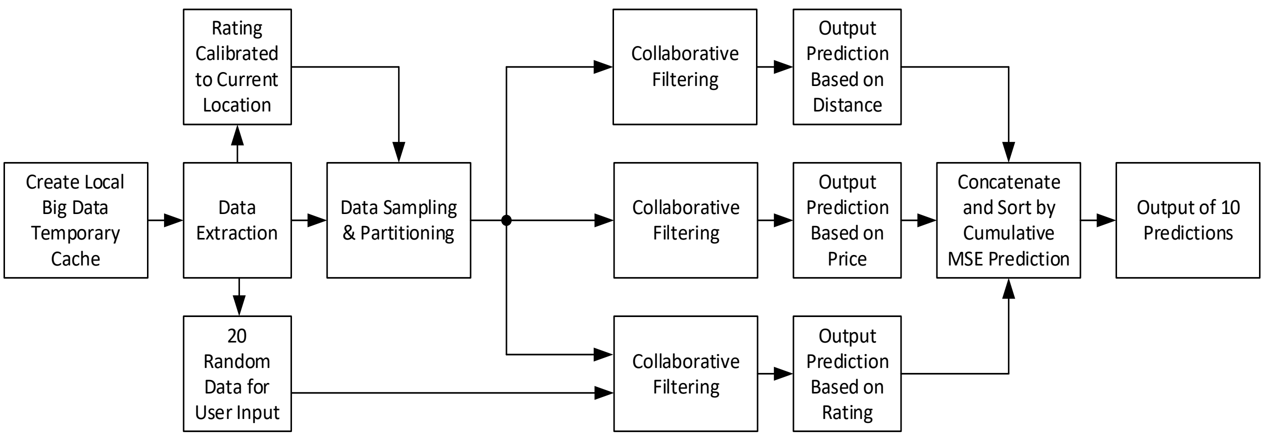

Figure 1.
Logistic Regression with Collaborative Filtering (Model 1).

The first system, as depicted in the flowchart in Figure 1, is a simplistic feedforward system that prompts 10 location outputs.

ISSN: 2576-8484
Vol. 10, No. 4: 146-159, 2026
DOI: 10.55214/2576-8484.v10i4.12575

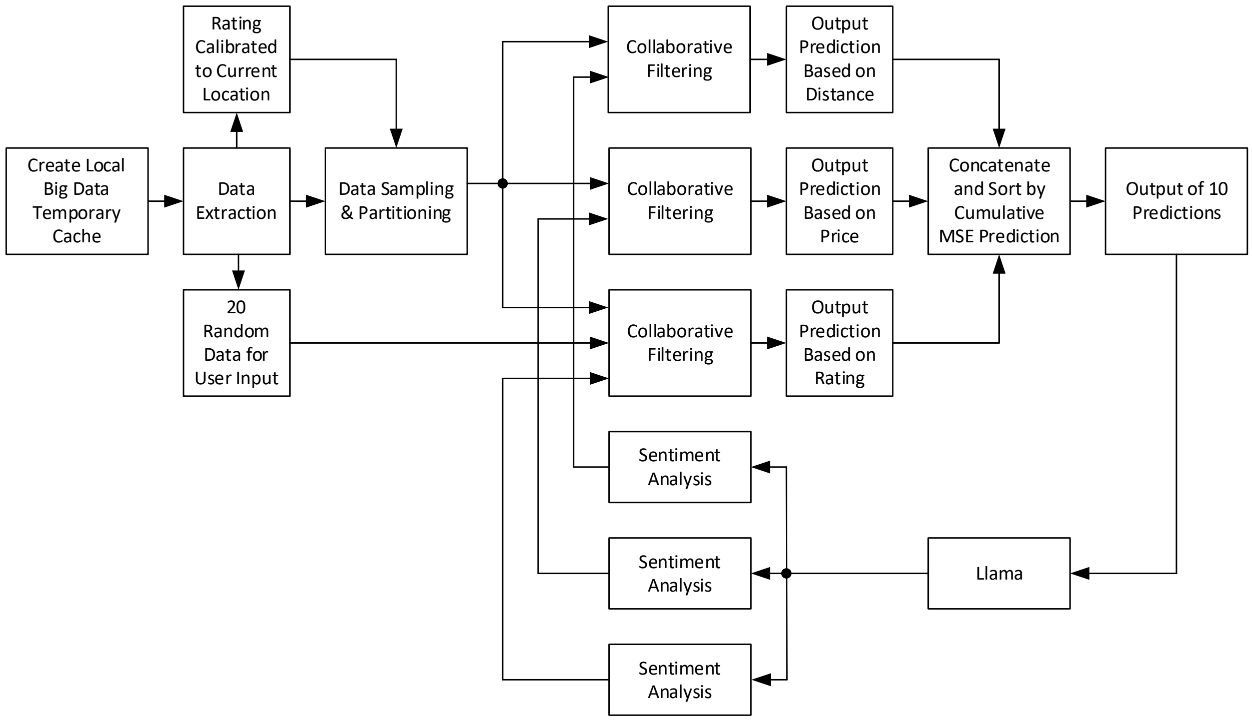

Figure 2.
Hybrid AI (Model 2).

Figure 2 depicts the model post-enhancement. This hybrid AI system is more robust. Model 2 describes the enhancements made.

4.1. Model 1: Logistic Regression with Collaborative Filtering Prediction

Data is read from an Excel sheet for training. The data is partitioned into 80% for training and 20% for testing. Samples are randomly drawn for training. Once training is complete, classification occurs through Spark Prediction. Twenty predictions are narrowed down, and the recommended sites are forwarded to the next phase. This concept applies to Rating, Distance, and Price prediction [21].

On user rating, however, an additional component, which is the training, can be influenced by the current user feedback. 20 random locations are chosen for users to provide feedback. The highest rating is 5, and the lowest is 0. If a user inputs -1, the system recognizes it as a place that has not been visited prior to the survey. Similar to Price and Distance classification, classification is done through Spark Prediction. 20 predictions are narrowed, and the recommended sites are forwarded to the next phase.

The data of Rating, Distance, and Price prediction is concatenated, and the similar predictions are grouped and sorted. The MSE is averaged, and the final recommendation is based on the lowest MSE for each location prediction. Ten outputs are sorted in ascending order of MSE. A simplified version of the entire system is explained through the software architecture diagram.

The classification method used in this project is logistic regression. In the context of deep learning, the output of multiple logistic regressions can be stacked to form a feedforward neural network. The mathematical formulation of logistic regression can be integrated into the broader context of neural networks. Though unconventional, mean squared error (MSE) is chosen over cross-entropy loss to study the feedback accuracy of the system while tuning. The mathematical depiction is provided by the equations below:

- m: Number of instances.
- n: Number of features.

- N: Total number of instances.
- X: Feature matrix ($N \times n$).
- y: True labels ($N \times c$), where c is the number of classes.
- $W^{[h]}$, $b^{[h]}$: Weights and bias for the hidden layer.
- $W^{[o]}$, $b^{[o]}$: Weights and bias for the output layer.
- $\sigma(z)$: Sigmoid activation function.

4.1.1. Hidden Layer Computation

- $Z^{[h]}=XW^{[h]}+b^{[h]}$
- $A^{[h]}=\sigma(Z^{[h]})$

4.1.2. Output Layer Computation

- $Z^{[o]}=A^{[h]}W^{[o]}+b^{[o]}$
- $A^{[o]}=\sigma(Z^{[o]})$

4.1.3. Computation Cumulative Mean Squared Error (MSE)

$$MSE_{cum} = \frac{1}{N \cdot c}\sum_{i=1}^{N} \sum_{j=1}^{c} (y_{i,j} - A_{i,j}^{[o]})^2$$

MSE is chosen for assessing the accuracy of a predictive model and fine-tuning, particularly in regression tasks. It quantifies the average squared difference between the predicted values and the actual values. The lower the MSE, the better the model's predictions align with the true values. With this concept, the system is devised to be trained to achieve the lowest Cumulative MSE to predict 10 out of 438 location data points fed to the system. The Cumulative MSE comprises results from predicting affordable price points, shortest distances, and user ratings.

Model 2: Adding qualitative insights with Llama and sentiment analysis over logistic regression with collaborative filtering

Once the top 10 places are identified, the Llama model provides additional insights for each. This step enables the system to offer qualitative descriptions of each recommendation, highlighting unique aspects and advantages of visiting each place.

- Prompt Engineering for Llama: Design prompts that feed structured data and context (e.g., “Based on distance, price, and user rating, why should a user visit this place?”) into Llama. The model returns concise, appealing summaries or advantages for each place.
- Fine-tuning Llama for Contextual Relevance: To ensure relevant responses, fine-tune Llama with training data that includes location descriptions and user feedback. This enhances the model’s ability to generate meaningful descriptions that align with the user’s interests.

With the feedback received from LLAMA, sentiment analysis is done to provide feedback to the existing ML system to fine-tune. Converting the system from a feed-forward to a system that receives active feedback

4.2. The Hybrid Approach: Logistic Regression + LLAMA AI

When combined, Logistic Regression and LLAMA AI complement each other in powerful ways, creating a recommendation system that leverages both data-driven optimization and human-like contextual understanding.

- Personalized Data Feed for LLAMA: Logistic regression models can process user data to identify the most relevant destinations for each user. This data can then be fed into LLAMA, which can generate highly personalized, insightful responses based on these recommendations. For example, if logistic regression identifies that a user is likely to enjoy beach destinations, LLAMA can elaborate on this by offering detailed information about the best beach spots, unique activities, and local tips.

ISSN: 2576-8484
Vol. 10, No. 4: 146-159, 2026
DOI: 10.55214/2576-8484.v10i4.12575

- Combining Predictive Power with Insightful Descriptions: Logistic regression optimizes for the locations most likely to appeal to the user, while LLAMA explains why a specific location is a good fit, what makes it unique, and why the user might enjoy it. This combination enhances the user's experience by providing actionable insights and context that aid decision-making.
- Handling Complexity: Logistic regression handles structured data well, while LLAMA excels at understanding unstructured data. Together, they form a system capable of handling both structured and unstructured inputs, leading to more accurate and contextually aware recommendations.

Sentiment analysis is a key component for feedback and tuning of the system. It is also the factor that binds LLAMA with logistic regression. For instance, the price parameter is tuned such that the user inquires about the cost-effective method to travel to each destination and estimates spending. LLAMA provides an estimation of cost and how it can be reduced with different activities of choice. This allows the user to plan their costs ahead and reduce excess spending. While doing so, the system records the sentiment the user has with the interaction and updates the rating for pricing accordingly. From the start of a pricing-related inquiry, 3 to 5 consecutive interactions are analyzed within a stipulated timeframe to deduce the need to update the dataset pertaining to pricing.

Similarly, for distance sentiment analysis on travel modes, the analysis is conducted. Initial distance quantification is always from the current location to the nearest spot. It is constantly updated based on places the user is more comfortable traveling to, considering the mode of transport and convenience. Similar to the price analogy, the interaction of the user with the system is analyzed, and data is adjusted accordingly based on the outcome.

With a mixture of machine learning models and AI, this study was able to overcome the cold start related to learning an individual before recommending certain places. The system retains all the information of the user and their preferences related to a certain tourist spot. When prompted, the machine learning aspect of the system immediately heads up 10 locations to LLAMA, which describes the pros and cons of visiting the place. Opposed to traditional LLAMA, which will need prompts from the user to refine its search further before selecting tourist spots relevant to the user's needs.

5. Results

A comparative study with and without the hybrid AI introduction is as follows: The initial study was conducted without LLAMA and sentiment analysis for feedback. MSE or MAE-based evaluation of the system is explored by multiple scholars. An example in research is the precision of predicted item evaluations, quantified through regression-type error metrics such as mean absolute error (MAE) or root mean square error (RMSE) [22]. Another study, which uses a similar approach, quantifies the MAE and RMSE based on different data sizes [23]. It concludes that RMSE is more accurate than MAE for detecting prediction mistakes in a hybrid collaborative model. Based on this, the current system was fine-tuned with MSE values while considering other components for an accuracy study.
The results from each prediction are listed in the following tables.

Table 1.
Yields of Distance Prediction.

	System	
Type	**Method 1**	**Method 2**
R^2	-6.503	0.724
Mean absolute error	2.745	0.823
Mean squared error	10.804	2.357
Root mean squared deviation	3.287	0.986
Mean signed difference	-2.745	-0.674

Table 2.
Yields of User Rating Prediction.

	System	
Type	**Method 1**	**Method 2**
R^2	-3.128	0.724
Mean absolute error	2.247	0.823
Mean squared error	7.918	2.357
Root mean squared deviation	2.814	0.986
Mean signed difference	-2.247	-0.674

Table 3.
Yields of Price Prediction.

	System	
Type	**Method 1**	**Method 2**
R^2	-6.503	0.452
Mean absolute error	2.745	0.549
Mean squared error	10.804	2.161
Root mean squared deviation	3.287	0.657
Mean signed difference	-2.745	-0.549

Table 4.
MSE Results.

System	Mse		
	DISTANCE	**PRICE**	**RATING**
Method 1	10.804	10.804	7.918
Method 2	3.802	2.161	2.357

The MSE data listed is individually trained and collected for each feature
The formula for cumulative MSE is:
Cumulative MSE $= \frac{1}{k}\sum_{i=1}^{k} MSE_i$
where:

- k is the number of predictions.
- MSE_i is the MSE for the i-th prediction.

Applying the formula:

Method 1

Cumulative MSE $= \frac{1}{3} \cdot (10.804 + 7.918 + 10.804)$
Cumulative MSE = 9.842

Method 2

Cumulative MSE $= \frac{1}{3} \cdot (3.802 + 2.357 + 2.161)$
Cumulative MSE = 2.773

Table 5.
RMSE Results.

SYSTEM	RMSE		
	DISTANCE	**PRICE**	**RATING**
METHOD 1	3.287	2.814	3.287
METHOD 2	1.083	0.657	0.986

The RMSE data listed is individually trained and collected for each feature
The formula for cumulative RMSE is:

ISSN: 2576-8484
Vol. 10, No. 4: 146-159, 2026
DOI: 10.55214/2576-8484.v10i4.12575

Cumulative RMSE $= \frac{1}{k}\sum_{i=1}^{k} \mathrm{R}MSE_i$
where:

- k is the number of predictions.
- $RMSE_i$ is the RMSE for the i-th prediction.

Applying the formula:

Method 1
Cumulative RMSE $= \frac{1}{3} \cdot (3.287 + 2.814 + 3.287)$
Cumulative RMSE = 3.129
Method 2
Cumulative RMSE $= \frac{1}{3} \cdot (1.083 + 0.657 + 0.986)$
Cumulative RMSE = 0.909

Table 6.
MAE Results.

SYSTEM	MAE		
	DISTANCE	PRICE	RATING
METHOD 1	2.745	2.247	2.745
METHOD 2	0.823	0.549	0.823

The MAE data listed is individually trained and collected for each feature
The formula for cumulative MAE is:
Cumulative RMSE $= \frac{1}{k}\sum_{i=1}^{k} MAE_i$
where:

- k is the number of predictions.
- MAE_i is the MAE for the i-th prediction.

Applying the formula:

Method 1
Cumulative MAE $= \frac{1}{3} \cdot (2.745 + 2.247 + 2.745)$
Cumulative MAE = 2.579

Method 2
Cumulative MAE $= \frac{1}{3} \cdot (1.083 + 0.657 + 0.986)$
Cumulative MAE = 0.732

Based on the first study with Method 1 classification, it was imperative that distance and price are correlated to each other in most cases. The effect of the 438 data points on the MSE produced by distance and pricing cannot be altered because of their static nature. However, the user's opinion on places is subjective and can fluctuate based on preference. Hence, the Hybrid AI or Method 2 study attempted to reduce the MSE produced by user ratings for more precise prediction. The opinion of the user was also aggregated with distance and pricing classification, as explained in the methodology section.

The study uses a unique cumulative form of MSE to aggregate and study the system's reliability, which is not explored. It is found that by cumulating MSE of different factors, it can still improve the system's precision for user recommendations. This is due to the ever-evolving nature of the weightage given to pricing, distance, and user ratings according to Hybrid AI, which adjusts preferences based on the user, making it ideal for personalized responses.

The distance and price parameters were tuned using sentiment analysis through user feedback to suggestions by LLAMA. For price, the user inquires about the cost-effective method to travel to each destination and estimates of spending. Based on their response to the AI's suggestion, the cost profile is updated to reflect the comfort of users spending in certain locations. Similarly, for distance, sentiment analysis on travel modes is conducted. Initial distance quantification is always from the current location to the nearest spot. It is constantly updated based on places the user is more comfortable traveling to due to mode of transport and convenience.

The variance and improvement of the system are close to 70% to 80%. The system vastly improved in relatable location recommendation by introducing the concept of Hybrid AI. A graphical depiction is shown below:

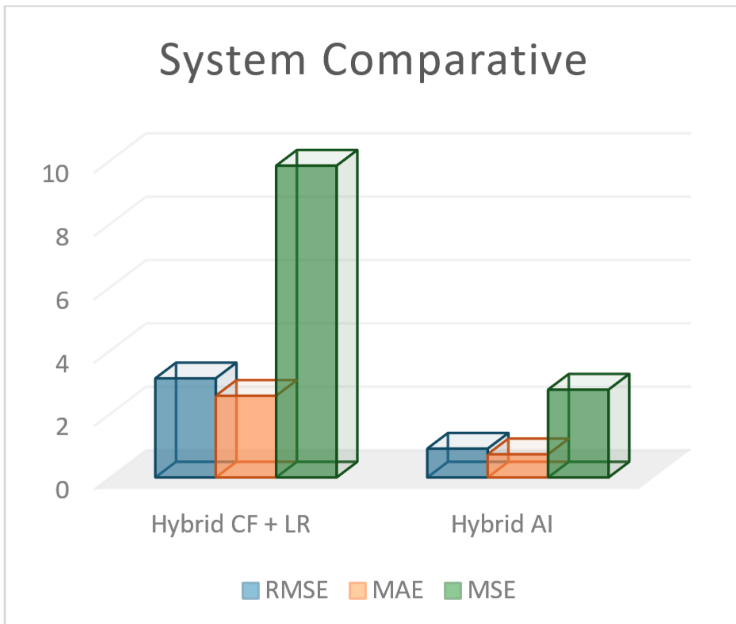

Figure 3.
Comparing Hybrid AI with Hybrid CF+ LR.

The R^2 value after implementing the Hybrid AI indicates a stronger correlation between predicted and actual user ratings. This suggests that the system is more accurate than traditional logistic regression with hybrid collaborative filtering.

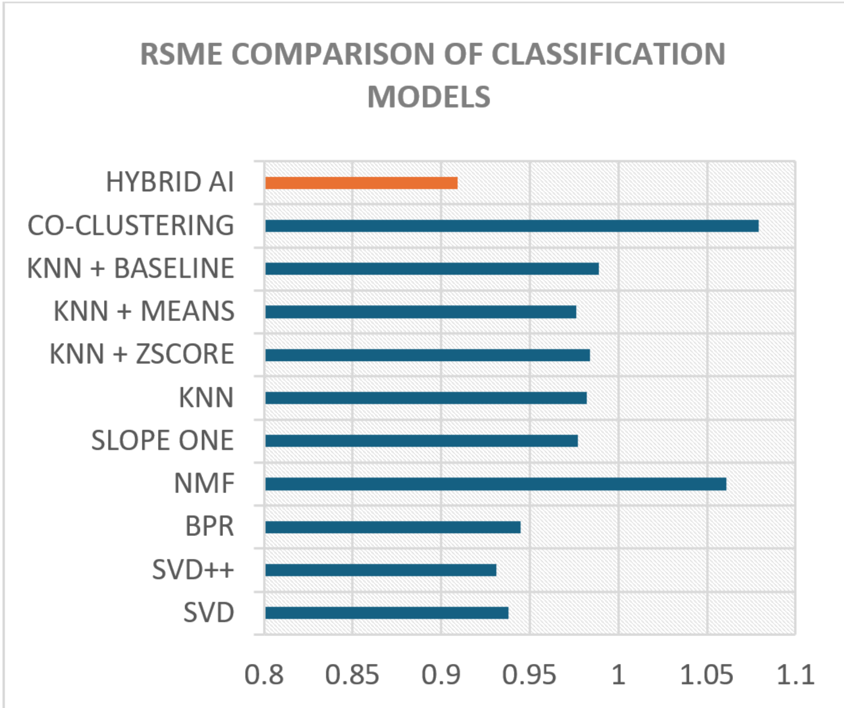

Figure 4.
Comparative Study with Machine Learning Methods.

A comparative study was conducted between Hybrid AI and other systems using enhanced machine learning methods for classification. The RMSE values of the proposed Hybrid AI algorithm decrease by approximately 3.09%, 2.36%, 3.81%, 14.32%, 6.96%, 7.43%, 7.62%, 6.86%, 8.08%, and 15.76%, respectively, compared with SVD, SVD++, BPR, NMF, Slope One, KNN, KNN + ZSCORE, KNN + MEANS, KNN, and Co-Clustering [24-30].

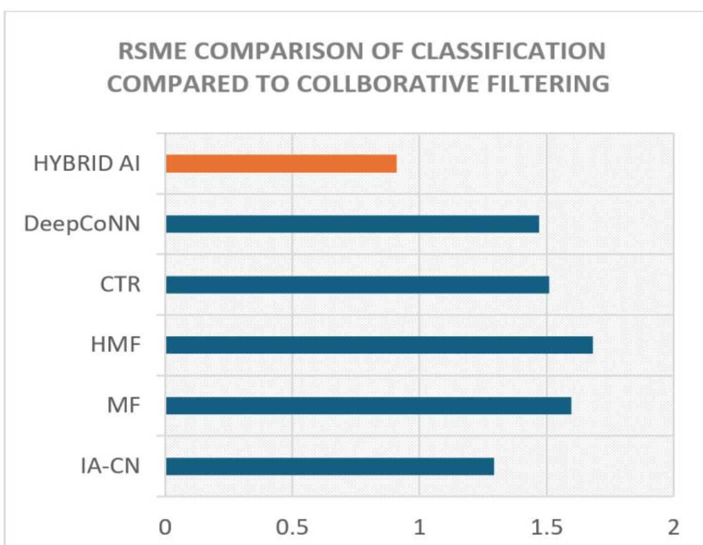

Figure 5.
Comparative Study with Collaborative Filtering-based Systems.

The system was also compared with other similar Collaborative Filtering-based systems. The results are as above. The RMSE values of the proposed Hybrid AI algorithm decrease by approximately 29.75%, 43.08%, 45.93%, 39.84%, and 38.21%, respectively, compared with IA-CN, MF, HMF, CTR, and DeepCoNN. Based on the RMSE comparison, it is proven that Hybrid AI comparatively outperforms other studies [31].

6. Conclusion

Studying the current system setup proves that it is possible to improve the system with Hybrid AI by quantifying the cumulative MSE instead of traditional logistic regression.

Improved accuracy has lowered the MSE value for both price ratings by actively suggesting comfortable traveling distances, cheaper travel options, and user ratings. The lowered R^2 value indicates a stronger correlation between predicted and actual user ratings.

With the hybrid system, a system with lowered underestimation of bias can be observed with more accurate predictions. The current best cumulative MSE achieved by the system is 0.732. It is worth noting that a cumulative form of MSE to aggregate decisions of different factors and study the reliability is explored. The proposed quantification method is able to give reliable feedback on how the system responds.

In conclusion, this research demonstrates that an intelligent tourist attractions recommender system with hybrid AI significantly improves upon traditional methods. The study successfully integrated Logistic Regression and Llama AI to create a system that provides personalized, data-driven recommendations while offering qualitative insights to users. By actively analyzing user sentiment from their interactions with Llama AI, the system dynamically adjusts its recommendations, addressing the "cold start" problem and making it more accurate and adaptable to individual preferences. This approach, which moves beyond simply processing structured data, proves to be a more effective way to recommend tourist attractions by considering factors like travel comfort, budget, and personal interests.

A key contribution of this study is the introduction of a cumulative Mean Squared Error (MSE) to evaluate the system's performance across different prediction factors: distance, price, and user rating. The research found that the Hybrid AI system achieved the best cumulative MSE of 2.773, a cumulative RMSE of 0.909, and a cumulative MAE of 0.732. These results represent a significant improvement over the initial system that lacked the AI and sentiment analysis feedback loop. The R^2 values also showed a notable improvement, indicating a stronger correlation between the predicted and actual user ratings, thus confirming the system's enhanced accuracy.

Thus, the Hybrid AI model for creating a scalable and efficient tourism recommender system improves accuracy by lowering MSE values for all three factors. It actively suggests comfortable travel distances, cheaper options, and explains the pros and cons of locations using Natural Language Processing (NLP). By providing reliable feedback on system responses and outperforming other machine learning and collaborative filtering methods in comparative studies, this research demonstrates that combining traditional machine learning with modern AI and sentiment analysis results in a more robust, dependable, and user-friendly platform for trip planning.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

References

[1] T. Huang and Z. Tang, "Estimation of tourism carbon footprint and carbon capacity," *International Journal of Low-Carbon Technologies*, vol. 16, no. 3, pp. 1040-1046, 2021. https://doi.org/10.1093/ijlct/ctab026

[2] J. Li and F. Di, "Application of artificial intelligence technology in smart tourism," in *2021 2nd Artificial Intelligence and Complex Systems Conference*, 2021.

[3] J. Yun and X. Youyuan, "Design of personalized leisure tourism route recommendation system based on deep learning," in *Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence*, 2020.

[4] I. Czarnowski, P. Jędrzejowicz, K.-M. Chao, and T. Yildirim, "Overcoming "big data" barriers in machine learning techniques for the real-life applications," *Complexity*, vol. 2018, p. Article 1234390, 2018. https://doi.org/10.1155/2018/1234390

[5] W. Yu, "Design of a social network data deep learning recommendation system guided by user preference," in *2021 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI) (pp. 33-39). IEEE*, 2021.

[6] A. Peña, J. Mesias, A. Patiño, J. Carvalho, and G. Gomez, "Evolution of the perceptions about tourist destinations affected by risk events using a panas-TDL Deep Learning Model," *Journal of Information Systems Engineering and Management*, vol. 6, no. 3, p. em0144, 2021.

[7] M. Flórez, E. Carrillo, F. Mendes, and J. Carreño, "A context-aware tourism recommender system using a hybrid method combining deep learning and ontology-based knowledge," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 20, no. 3, p. 194, 2025. https://doi.org/10.3390/jtaer20030194

[8] R. Kirubahari and S. Miruna Joe Amali, "A hybrid deep collaborative filtering approach for recommender systems (Version 1) [Preprint]," *Research Square*, 2021. https://doi.org/10.21203/rs.3.rs-651522/v1

[9] J. Karthiyayini and R. J. Anandhi, "Personalized travel recommendations system using hybrid filtering and deep learning," *Journal of Information Systems Engineering and Management*, vol. 10, no. 13s, pp. 118–129, 2025. https://doi.org/10.52783/jisem.v10i13s.2011

[10] D. Das, L. Sahoo, and S. Datta, "A survey on recommendation system," *International Journal of Computer Applications*, vol. 160, no. 7, pp. 6-10, 2017. https://doi.org/10.5120/ijca2017913081

[11] W. Lu *et al.*, "Intelligence in tourist destinations management: Improved attention-based gated recurrent unit model for accurate tourist flow forecasting," *Sustainability*, vol. 12, no. 4, p. 1390, 2020. https://doi.org/10.3390/su12041390

[12] M. Al-Hassan, H. Lu, and J. Lu, "A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system," *Decision Support Systems*, vol. 72, pp. 97-109, 2015. https://doi.org/10.1016/j.dss.2015.02.001

[13] M. Li and N. Fan, "Research on night tourism recommendation based on intelligent image processing technology," *Scientific Programming*, vol. 2021, no. 1, p. 2624621, 2021. https://doi.org/10.1155/2021/2624621

[14] M. Lenzen, Y.-Y. Sun, F. Faturay, Y.-P. Ting, A. Geschke, and A. Malik, "The carbon footprint of global tourism," *Nature Climate Change*, vol. 8, no. 6, pp. 522-528, 2018. https://doi.org/10.1038/s41558-018-0141-x

[15] A. Fatemeh, A. Khodadadi, and M. Yaghmaei, "A grouping hotel recommender system based on deep learning and sentiment analysis," *Journal of Information Technology Management*, vol. 11, no. 2, pp. 60–75, 2019.

[16] D. Bokde, S. Girase, and D. Mukhopadhyay, "Matrix factorization model in collaborative filtering algorithms: A survey," *Procedia Computer Science*, vol. 49, pp. 136-146, 2015. https://doi.org/10.1016/j.procs.2015.04.237

[17] M. Nilashi, O. bin Ibrahim, N. Ithnin, and N. H. Sarmin, "A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS," *Electronic Commerce Research and Applications*, vol. 14, no. 6, pp. 542-562, 2015. https://doi.org/10.1016/j.elerap.2015.08.004

[18] C. A. Martín, J. M. Torres, R. M. Aguilar, and S. Diaz, "Using deep learning to predict sentiments: Case study in tourism," *Complexity*, vol. 2018, no. 1, p. 7408431, 2018. https://doi.org/10.1155/2018/7408431Digital

[19] A. Kulkarni, P. Barve, and A. Phade, "A machine learning approach to building a tourism recommendation system using sentiment analysis," *International Journal of Computer Applications*, vol. 178, no. 19, pp. 48-51, 2019. https://doi.org/10.5120/ijca2019919031

[20] H. U. Rehman Khan, C. K. Lim, M. F. Ahmed, K. L. Tan, and M. Bin Mokhtar, "Systematic review of contextual suggestion and recommendation systems for sustainable e-tourism," *Sustainability*, vol. 13, no. 15, p. 8141, 2021. https://doi.org/10.3390/su13158141

[21] V. Kumar, Y. B. Chin, and L. W. Soong, "Intelligent tourist attractions recommender system with hybrid collaborative filtering," presented at the In 2024 Multimedia University Engineering Conference (MECON) (pp. 1-4). IEEE, 2024.

[22] D. Massimo and F. Ricci, "Building effective recommender systems for tourists," *AI Magazine*, vol. 43, no. 2, pp. 209-224, 2022. https://doi.org/10.1002/aaai.12057

[23] M. Sarkar, A. Roy, M. Agrebi, and H. AlQaheri, "Exploring new vista of intelligent recommendation framework for tourism industries: An itinerary through big data paradigm," *Information*, vol. 13, no. 2, p. 70, 2022. https://doi.org/10.3390/info13020070

[24] X. Zhou, J. He, G. Huang, and Y. Zhang, "SVD-based incremental approaches for recommender systems," *Journal of Computer and System Sciences*, vol. 81, no. 4, pp. 717-733, 2015. https://doi.org/10.1016/j.jcss.2014.11.016

[25] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756-2779, 2007. https://doi.org/10.1162/neco.2007.19.10.2756

[26] D. Lemire and A. Maclachlan, "Slope one predictors for online rating-based collaborative filtering," in *Proceedings of the 2005 SIAM International Conference on Data Mining (pp. 471-475). Society for Industrial and Applied Mathematics*, 2025.

[27] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Annals of Translational Medicine*, vol. 4, no. 11, p. 218, 2016. https://doi.org/10.21037/atm.2016.03.37

[28] P. Bruce and A. Bruce, *Practical statistics for data scientists: 50 essential concepts*. Sebastopol, CA: O'Reilly Media, 2017.

[29] H. Nicolas, "Surprise: A Python library for recommender systems," *Journal of Open Source Software*, vol. 5, no. 52, p. 2174, 2020.

[30] G. Li *et al.*, "Novel recommendation system for tourist spots based on hierarchical sampling statistics and svd++," *Mathematical Problems in Engineering*, vol. 2019, no. 1, p. 2072375, 2019. https://doi.org/10.1155/2019/2072375

[31] F. Ming, L. Tan, and X. Cheng, "Hybrid recommendation scheme based on deep learning," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 6120068, 2021. https://doi.org/10.1155/2021/6120068