

A Robust Hybrid Learning Framework for Personalized Destination Recommendation in Tourism Analytics

K. Sowmya¹, Avula Architha², Bandela Phanivarma², Nallam Varshitha², Rapaka Charuhasan²

¹Assistant Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering (Data science)

^{1,2}Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India

ABSTRACT

Indonesia's tourism industry has expanded rapidly, leading to the generation of extensive data on traveller interests, destinations, costs, and trip durations. However, traditional recommendation approaches such as those offered by travel agencies or guidebooks often depend on generalized assumptions and lack personalization, resulting in suggestions that may not fully match individual preferences. To overcome these shortcomings, this research presents a hybrid learning-based travel recommendation system that utilizes Indonesian tourism datasets to provide tailored suggestions. The study compares several Classification and Regression Tree (CART)-based techniques, including Linear Logistic Regression (LLR), Support Vector Machine (SVM), and Random Forest (RF). The proposed approach, Neuro Tree Net (NTN), merges an Artificial Neural Network (ANN) with an Extra Trees (ET) model, enabling the system to effectively learn complex and nonlinear relationships within tourism data. Experimental evaluations reveal that this hybrid model delivers higher prediction accuracy than conventional single-model methods. Users can define their preferences, such as budget constraints, available travel time, and minimum ratings, to receive customized destination recommendations. The system's backend manages data preprocessing, model training, and predictions using tools like pandas, scikit-learn, Keras, and Matplotlib, while an interactive interface developed with Django displays the results. In addition to improving recommendation accuracy, the system also provides valuable insights into tourism trends, helping users plan their trips more effectively and make well-informed decisions.

1. INTRODUCTION

In data-driven markets, digital platform competitors reliably offer effective advice. Recommendation engines are a powerful tool for internet companies. While thematic recommendation systems employ deep technologies such as deep and reinforced learning for data streaming in eCommerce applications (platforms such as Spotify, Netflix, and Amazon), other sophisticated sectors such as travel, insurance, logistics, and healthcare are still in their early days of adoption. There is a need for smart recommendation systems. Particularly in the travel sector, travel planning companies that started in the past two decades (e.g., Skiplagged in 2013, Google Flight in 2011, Airbnb in 2008, Kayak in 2005, HomeAway in 2004, SkyScanner in 2003, TripAdvisor in 2000, and Expedia in 1996) could have sustained their services with smart recommendation systems.



Fig. 1: Travel Advice tips

Currently, travel decisions are predominantly made without the full knowledge of several alternatives and options based on user-specific preferences as shown in fig 1. Potential travellers resort to travel agents to recommend rather than make their own decisions due to a lack of information/awareness of new destinations. Furthermore, the information available over the Internet adapted with filtering and selection methods used by search engines can provide limited guidance only. With large amounts of information available online, travelers can get overwhelmed during their decision-making process. With recent developments in technologies and practical tools available to process big data, there exists an abundant opportunity to build automatic or semiautomatic systems that would provide a range of alternatives and suggestions based on user preferences to help travelers in their decision-making process. These form the key motivations of this research work that aims to design and develop a prototype of a travel recommendation system with edge technologies and effective models. However, prior to scoping an effective travel recommendation system, we considered the current state-of-the-art practices in the industry and their limitations. At present, most of the key travel industry players continue to rely on a destination-based recommendation approach in the form of travel packages. However, more inclusive data-driven insights are required for modern travel recommendation systems to provide personalized advice. A smart recommendation system could be developed to meet the varying needs of individual tourists in terms of personal tastes, interests, travel habits, and other contextual preferences influenced by time and monetary constraints. Currently, potential tourists are provided with a plethora of choices from a variety of fixed package-based recommendations that can lead to decision fatigue; furthermore, they fail to meet the consumer requirements effectively.

2. LITERATURE SURVEY

Scharf et al. [1] proposed modeling uncertainty in recommendation data using probabilistic ranking instead of single-score methods. It introduces RankDist, an algorithm that computes the probability of an item's rank in a recommendation list to maximize user satisfaction. The study builds on prior work in probabilistic databases and demonstrates that rank-based methods, newly applied to recommender systems, offer both theoretical optimality and improved empirical performance. Jewpanya et al. [2] proposed algorithm utilized an extended model of the Team Orienteering Problem with Time Windows (TOPTW) to account for mandatory locations and activities at each site. It offered trip planning that included a set of locations and activities designed to maximize the overall score accumulated from visiting these locations. To solve the proposed model, the Adaptive Neighborhood Simulated Annealing (ANSA) algorithm was applied. Tsai et al. [3] approaches

accurately predict user preferences for individual POIs, they struggle to generate complete itineraries that accommodate personalized constraints, such as designated start/end locations and travel time limitations. Halder et al. [4] proposed an adaptive Monte Carlo Tree Search (MCTS) algorithm for personalized POI selection and pruning, effectively handling multiple constraints in itinerary planning. Their approach integrates user preferences and temporal constraints to generate feasible and personalized travel itineraries. Guo et al. [5] Introduced the Top-Personalized-K Recommendation, a task focused on tailoring the number of recommended items to maximize individual user satisfaction. To achieve this, the authors propose PerK, a model-agnostic framework that selects the optimal list size by estimating and maximizing expected user utility using calibrated interaction probabilities. Zhang et al. [6] proposed to enhance travel recommendations, Next-location recommendation techniques predict subsequent POIs based on users' prior trajectories, while top-k location recommendations present a selection of appealing POIs. Shrestha et al. [7] helped provide better suggestions to travelers, promoted decision-making, and raised satisfaction levels all around. The study emphasized the significance of questionnaire design, including demographic data, creating a strong association model using machine learning. Decision criteria were constructed based on the data quality being evaluated. Data from multiple sources were combined to create a comprehensive tourist database, and the recommender system included user feedback and decision-making guidelines.

Zheng et al. [8], looks for a particular destination or attraction, the aim was to predict the next location of the traveler, and a heuristic procedure based on a data mining approach was proposed. This procedure learns the mobility of the tourist against the past movements of visitors; i.e., it uses historical data. For this study, data was collected from travelers using GPS-like tracking applications at the Summer Palace in Beijing, China. The proposed model and study can contribute significantly toward providing better location-based services, promotions of tourist attractions, crowd control, etc. Personalized route predictions are systems that predict routes based on user requirements.

Wolfgang et al. [9] proposed a route recommendation approach that utilizes user-defined points of interest to generate personalized travel paths. In this system, users provide their starting and destination locations through a web-based interface, and the system recommends intermediate places aligned with their preferences. The Dijkstra algorithm is employed to compute the shortest and most efficient routes, ensuring optimal path selection and improved travel planning experience.

Sun et al. [10] developed a personalized route recommendation system based on a two-stage architecture. In the first stage, candidate locations are generated using an SVM model to filter relevant points of interest based on user preferences. In the second stage, these candidates are ranked using a gradient boosting regression tree, which assigns scores and refines the recommendation list to improve accuracy and relevance. Cao et al. [11] proposed a personalized route recommendation system that focuses on identifying and utilizing user-preferred locations. The approach emphasizes capturing user interests to generate routes that are more relevant and tailored to individual preferences, thereby enhancing recommendation quality and user satisfaction. Chen et al. [12] proposed an intelligent route recommendation system that improves both personalization and scalability. The system leverages user preference data to generate more accurate travel recommendations and is implemented using a Hadoop-based framework to efficiently handle large-scale data. This enables improved computational performance and supports real-time recommendation scenarios in practical applications. Pan et al. [13] presented a three-step algorithm for independent travel route recommendation. Initially, a 0–1 knapsack model is used to select optimal landmarks under given constraints. Subsequently, an analytic hierarchy process is applied to evaluate and prioritize these landmarks, followed by optimization using a simulated annealing

algorithm. Finally, the route planning problem is formulated as a traveling salesman problem to determine the most efficient travel sequence.

3. PROPOSED SYSTEM

The proposed system as illustrated in Fig. 2 provides personalized travel recommendations for tourists in Indonesia by leveraging machine learning and hybrid models. Users can upload datasets, which are pre-processed using techniques such as handling missing values, encoding categorical variables, TF-IDF vectorization for textual reviews, and train-test splitting. Features are scaled or normalized when required, ensuring optimal model performance. For predictions, the system implements multiple models for comparative analysis LLR, SVC, RF, and the proposed NTN model, captures complex nonlinear relationships in the data, improving prediction accuracy for both duration and recommendation score. The platform also integrates EDA capabilities, generating interactive visualizations such as histograms, scatter plots, boxplots, correlation heatmaps, and pairplots, helping users gain insights into tourism trends and patterns.

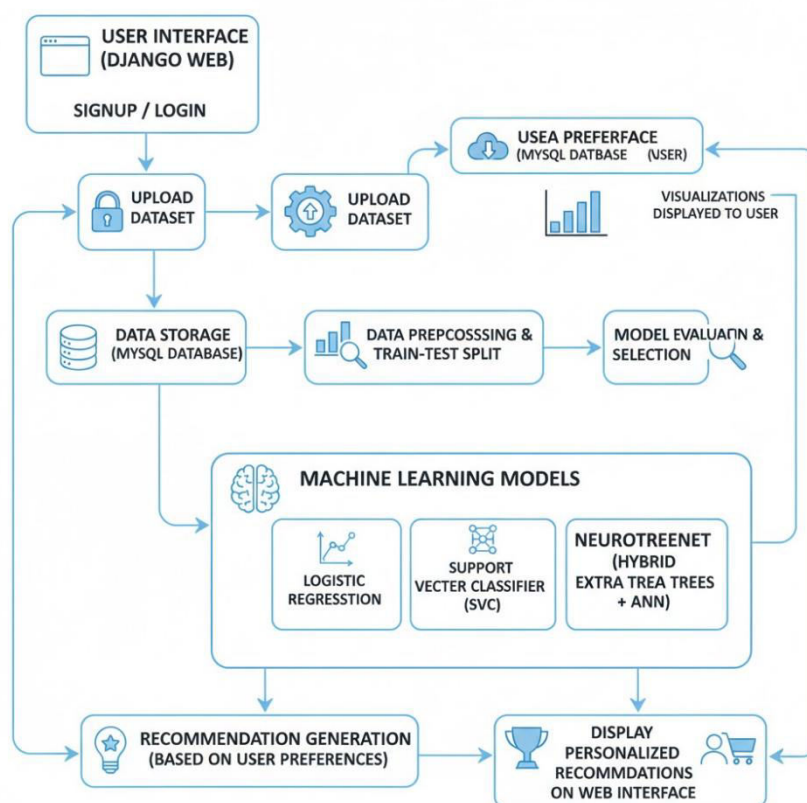


Fig. 2: Proposed system architecture.

Users can input constraints such as budget, travel duration, season, destination, and interests, and the system produces personalized predicted travel duration and recommendation scores as shown in Fig. 4.1. Additionally, the system can generate AI-assisted travel suggestions using the Gemini API, providing tailored recommendations for each traveler.

3.1 Neuro TreeNet CART Model

The Neuro TreeNet Model is a hybrid machine learning architecture that combines the strengths of tree-based ensemble models (like ET or Random Forest) with Artificial Neural Networks (ANNs) as illustrated in Fig. 3. By integrating these two approaches, the model can capture both hierarchical

decision rules and complex nonlinear feature interactions, providing superior predictive performance compared to single-model approaches. Tree-based models excel at handling categorical features, missing values, and capturing hierarchical relationships, while ANNs are powerful at learning high-dimensional nonlinear representations, such as patterns from TF-IDF vectors of reviews. The Neuro Tree Net model fuses these capabilities to leverage the best of both worlds.

Tree-Based Component Training (ET): A forest of ET is trained on the input features. Each tree identifies feature splits that best separate classes (or predict regression targets). The output of this step includes predicted values or transformed features for each training sample.

Feature Transformation: The outputs from the tree ensemble (predictions or leaf indices) are converted into numerical representations. These transformed features capture hierarchical relationships and patterns from the tree component that the neural network can further process.

Neural Network (NN) Training: The NN is trained using tree outputs combined with original input features. It learns complex nonlinear relationships and interactions that the tree-based component alone cannot capture, enhancing predictive performance.

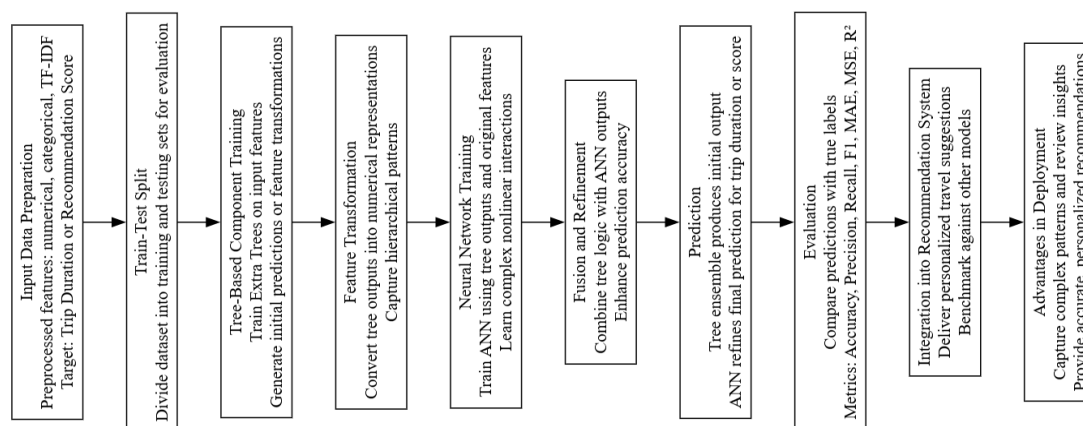


Fig. 3: Internal working of Neuro TreeNet model

Fusion and Refinement: The ANN refines the predictions from the tree component, combining hierarchical decision logic with nonlinear feature interactions. This fusion enables the model to handle high-dimensional, mixed-type features effectively.

Prediction: For a new user input, the tree ensemble first produces initial predictions, which are passed to the ANN. The final output is the predicted trip duration category (classification) or recommendation score (regression).

Evaluation: The model's predictions are compared with the true labels in the testing set. Performance metrics like accuracy, precision, recall, F1-score (for classification) and MAE, MSE, RMSE, R^2 (for regression) are computed to assess effectiveness.

Integration into the System: The Neuro TreeNet predictions are integrated into the recommendation system. Users receive personalized travel suggestions based on their preferences, and the model's outputs are combined with other models for performance benchmarking.

Advantages in Deployment: The hybrid structure allows the system to capture complex patterns in tourism data, leverage user reviews via TF-IDF, and deliver more accurate recommendations than single models like LLR, SVM, or RF.

9.3 Results description

This section presents the outcomes obtained from the implementation of the Travel AI system developed for tourism prediction and recommendation. The results demonstrate the performance of the ML models used for travel duration classification and recommendation score prediction. Various visualizations and evaluation metrics highlight the effectiveness of the implemented algorithms in analyzing tourism data. The generated results illustrate the system’s ability to process travel attributes, produce accurate predictions, and deliver intelligent travel recommendations through the integrated AI framework.

Fig. 4 Neuro-Tree-Net: The confusion matrix generated by the Neuro-Tree-Net CART model illustrates the classification output obtained from the hybrid tree-based learning approach implemented in the Travel AI system. The matrix shows a strong concentration of values along the diagonal, indicating a high number of correct predictions for travel duration categories. This result demonstrates the effectiveness of the Neuro-Tree-Net CART model in learning relationships between tourism features, textual review representations, and travel duration outcomes within the dataset

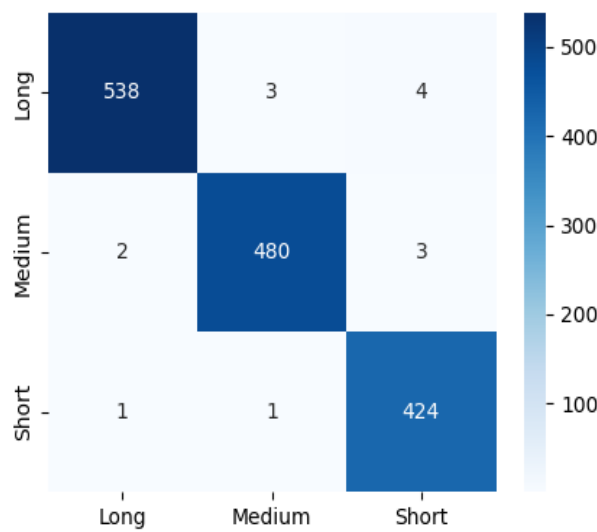


Fig. 4: Obtained confusion matrix of duration attribute for Neuro-Tree-Net

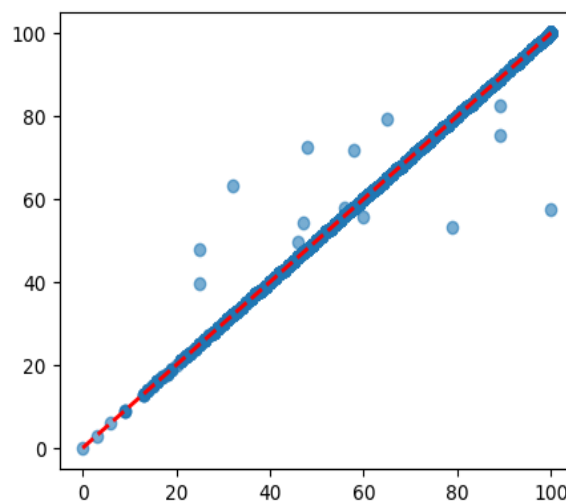


Fig. 5: Obtained scatter plot of Recommendation Score attribute Neuro-Tree-Net

Fig. 5 Neuro-Tree-Net: The scatter plot produced by the Neuro-Tree-Net CART model presents the predicted recommendation scores relative to the actual values in the dataset. The concentration of points near the diagonal reference line reflects strong agreement between predicted and actual scores. This distribution demonstrates the effectiveness of the Neuro-Tree-Net CART approach in modelling relationships between tourism features, review text representations, and recommendation score outcomes within the Travel AI recommendation system.

Fig. 6 presents the prediction output generated by the Travel AI system when test data is processed through the trained machine learning models. The prediction interface displays the estimated travel duration category and the recommendation score generated from the classification and regression models. The system processes input attributes such as destination, season, budget category, travel interest, and review information. These attributes undergo preprocessing operations including label encoding and TF-IDF transformation before entering the trained models. The Neuro-Tree-Net CART model produces the final predictions, which appear on the interface along with the recommendation score value. The system also generates intelligent travel suggestions through the integrated AI recommendation module, providing personalized tourism guidance based on the predicted results.

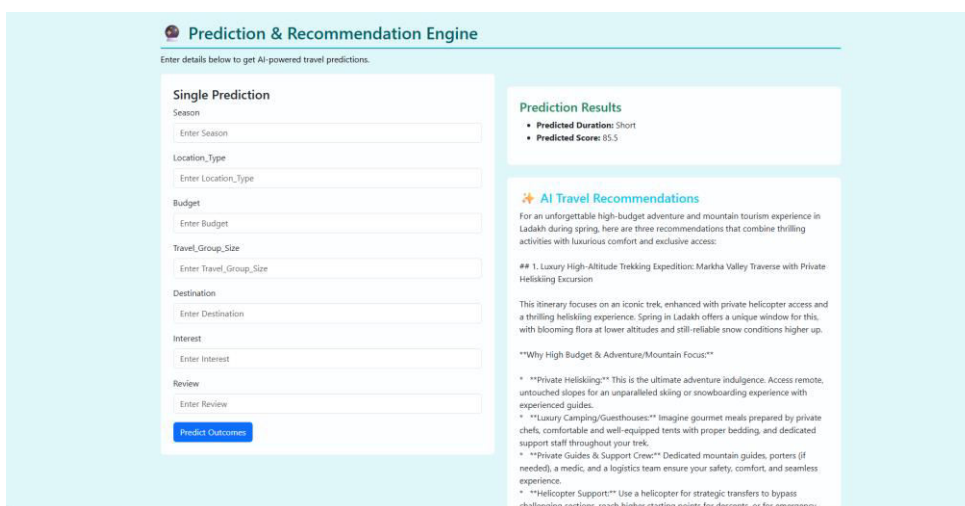


Fig. 6 Prediction on Test Data after user login

Table 1: Classification performance comparison for travel duration prediction

Model	Accuracy	Precision	Recall	F1-Score
LR model	0.35	0.34	0.35	0.33
SVM model	0.36	0.35	0.365 36	0.35
RF model	0.48	0.48	0.48	0.48
Neuro-Tree-Net	0.994	0.99	0.99	0.99

Table 1 presents the classification performance comparison of different machine learning models used for predicting the travel duration attribute in the Travel AI system. The evaluation metrics include Accuracy, Precision, Recall, F1-Score, and AUC. LR and SVM produce moderate

classification performance with accuracy values around 0.35. The RF classifier shows improved performance with higher accuracy and balanced precision and recall values. The Neuro-Tree-Net CART model demonstrates the highest performance among all models with an accuracy of 0.99 along with very high precision, recall, and F1-score values. These results indicate that the Neuro-Tree-Net CART model provides highly accurate classification results for predicting travel duration based on tourism features.

Table 2 presents the regression performance comparison of different models used for predicting the recommendation score in the Travel AI system. The evaluation metrics include MAE, MSE, RMSE, and R² Score. LR and SVM Regressor produce higher error values and very low R² scores, indicating weaker prediction performance. RF Regressor shows better results with reduced error values and an improved R² score. The NTN CART model achieves the best regression performance with extremely low MAE, MSE, and RMSE values along with an R² score of 0.99. These results demonstrate that the Neuro-Tree-Net CART model produces highly accurate recommendation score predictions within the Travel AI system.

Table 2: regression performance comparison for recommendation score prediction

Model	M AE	MS E	RMS E	R² Score
LR model	19. 98	552. 81	23.5 1	0.01
SVM model	19. 94	553. 88	23.5 3	0.008
RF model	18. 46	494. 67	22.2 4	0.11
Neuro-Tree- Net	0.1 55	3.77	1.94	0.99

5. CONCLUSION

The Travel AI system presents an intelligent tourism recommendation platform that integrates machine learning techniques with a web-based interface for predicting travel duration and recommendation scores. The system processes tourism-related attributes such as destination, season, budget, travel interests, and textual reviews to generate accurate predictions and personalized travel suggestions. Data preprocessing techniques including label encoding and TF-IDF vectorization transform categorical and textual information into numerical representations suitable for machine learning models. Several classification and regression algorithms were implemented and evaluated to determine the most effective model for tourism prediction tasks. The classification models predict travel duration, while the regression models estimate recommendation scores for destinations. Performance evaluation using metrics such as accuracy, precision, recall, F1-score, MAE, MSE, RMSE, and R² score demonstrates that the NTN CART model achieves the highest predictive performance compared with other models. The system also integrates an AI-based recommendation module that generates travel suggestions based on user preferences and predicted results. The developed system provides a comprehensive platform for tourism data analysis, predictive modeling, and intelligent travel recommendation. The integration of data visualization, machine learning

models, and AI-generated suggestions creates a decision-support tool that improves travel planning and enhances the tourism experience.

REFERENCES

- [1] Scharf, C.; Domshlak, C.; Gal, A.; Roitman, H. A Rank-Based Approach to Recommender. System's Top-K Queries with Uncertain Scores. *Proc. ACM Manag. Data* 2025, 3, 5.
- [2] Jewpanya, P.; Nuangpirom, P.; Pitjamit, S.; Nakkiew, W. Optimized Travel Itineraries: Combining Mandatory Visits and Personalized Activities. *Algorithms* 2025, 18, 110. <https://doi.org/10.3390/a18020110>
- [3] Tsai, C.Y.; Chuang, K.W.; Jen, H.Y.; Huang, H. A Tour Recommendation System Considering Implicit and Dynamic Information. *Appl. Sci.* 2024, 14, 9271.
- [4] Halder, S.; Lim, K.H.; Chan, J.; Zhang, X. A survey on personalized itinerary recommendation: From optimisation to deep learning. *Appl. Soft Comput.* 2024, 152, 111200.
- [5] Yu, J.; Guo, L.; Zhang, J.; Wang, G. A survey on graph neural network-based next POI recommendation for smart cities. *J. Reliab. Intell. Environ.* 2024, 10, 299–318.
- [6] Zuo, J.; Zhang, Y. Collaborative trajectory representation for enhanced next POI recommendation. *Expert Syst. Appl.* 2024, 256, 1248.
- [7] Shrestha, D.; Wenan, T.; Shrestha, D.; Rajkarnikar, N.; Jeong, S.-R. Personalized Tourist Recommender System: A Data-Driven and Machine-Learning Approach. *Computation* 2024, 12, 59. <https://doi.org/10.3390/computation12030059>.
- [8] Zheng, W.; Huang, W.; Li, Y. Understanding the tourist mobility using GPS: Where is the next place? *Tour. Manag.* 2017, 59, 267–280.
- [9] Wolfgang, W.; Hefele, A.; Herzog, D. Recommending a sequence of interesting places for tourist trips. *Inf. Technol. Tour.* 2017, 17, 31–54.
- [10] Sun, X.; Huang, Z.; Peng, X.; Chen, Y.; Liu, Y. Building a model-based personalised recommendation approach for tourist attractions from geotagged social media data. *Int. J. Digit. Earth* 2018, 661–687.
- [11] Cao, L.; Tao, J.; Chen, B. Implementation of Personalized Scenic Spots Route Recommendation System. In Proceedings of the 2018 13th International Conference on Computer Science & Education, Colombo, Sri Lanka, 8–11 August 2018.
- [12] Chen, X.; Zhou, L. Design and implementation of an intelligent system for tourist routes recommendation based on Hadoop. In Proceedings of the 2015 6th IEEE International Conference on Software Engineering and Service Science, Beijing, China, 18–20 October 2015.
- [13] Pan, Q.; Wang, X. Independent travel recommendation algorithm based on analytical hierarchy process and simulated annealing for professional tourist. *Appl. Intell.* 2018, 48, 1565–1581.

