CHAPTER 1

INTRODUCTION

Great journeys are not planned, they are made. Road trips are the best time to bond with family, and friends, and there's plenty that can be done. Moving away from a steady Internet connection means long conversations find a place to happen; songs that stir up nostalgia get a chance to be heard; embarrassing and funny old stories that you thought were buried deep get to relive (and admit it, you, kind of, like it); and more importantly, you get a chance to create new memories.

After all, on a road trip, the journey is always the real destination. And what's a little detour if you can end up having the time of your life in some town you would have otherwise never heard of? Taking your time means allowing for the spontaneous to happen, and allowing for some awesome unexpected experiences to take shape. Going on a road trip is really a very fascinating and fantastic experience in one's life. It gives the excitement and adventure of the place being travelled to as well as of the entire journey on the road.

However, the most common difficulty faced by people is to decide which would be the most optimal path to be taken to reach their destination while planning for a trip. Also people would want to know whether there are any site-seeing places that could be visited on the way to their destination and what would be the best time to travel to their destination. The existing works have been focusing on how to find routes that minimize a single kind of trip cost such as trip time or distance, amongst others. However, it has been noticed that the existing systems have not considered the user interest in their recommendation systems. Some only suggest the best time to travel to the destinations. Therefore, taking all these factors into consideration we have developed an application that provides suitable solutions for the above mentioned problems so that people can make their travel experience a memorable one. We have tried to efficiently plan a route that includes most of the sites of the user's interest based on the geographical location of these sites. We have focused on generating an aesthetic travel route, which covers the astounding tourist sites. We also focused on suggesting the best time to travel to the destinations so that the user will not have to browse through multiple sites to acquire all the information required to plan his/her travel. We have therefore designed an application where information about the route to be taken as well as the best time to travel is obtained at one place.

Objectives:

1. The application takes the source and destination as input from the user. It also asks the user to select his/her interest. An optimal path from the source to the destination is created which covers all the site-seeing places according to the user's interest. It displays this optimal path that is generated based on the user's interest on a map for better user readability.
2. It also displays a directions panel where the distance from every source to every other destination is mentioned so that the user can have a clear idea of the distance he/she will have to travel between source and destination by covering the intermediate places. It also recommends the best time to travel by analysing the weather data.

CHAPTER 2

LITERATURE SURVEY

The proposed project draws ideas and conclusions from several other papers published on topics concerning travel and prediction of prevailing as well as upcoming conditions of an area. The surface details of those papers which were used to derive the solutions and ideas related to various relevant concerns are described in the upcoming sections.

2.1 Machine Learning Based Short-Term Travel Time Prediction

According to a recent study [1] , the selected freeway segment travel time data were collected from the RITIS. RITIS is an advanced traffic analysis system that includes probe data analytics, segment analysis, and signal analytics. The raw travel data gathered from a series of selected road segments along the I-485 freeway in Charlotte, North Carolina, were used in the case study. As one of the most heavily travelled interstate freeways in the Charlotte metropolitan area, I-485 encircles the city and the last segment was completed in June 2015. The city of Charlotte has experienced a significant increase in daily traffic on many of its freeway segments in the past 25 years as the population of the Charlotte area increased from 688,000 to 1.4 million; more than 500,000 more residents are expected over the next 20 years. Charlotte has the largest population in the state and is also one of the fastest-growing metro areas in the U.S. The rapid population growth has caused traffic congestion on major roads. I-485 freeway segments in the southern Charlotte area experience massive recurrent congestion during weekdays due to heavy commuter and interstate traffic, which can seriously affect the travel and further economic development in this area. The I-485 Express Lanes project that began in the summer of 2019 will be completed in 2022 (the estimated cost is 346 million dollars) with one express lane added in each direction along I-485 between exit 67 (I-77) and exit 51 (U.S. 74). Travel time reliability and traffic flow in these freeway segments are therefore expected to improve. Figure shows the satellite map of the selected sections.

TTP Methods

2.1.1. Ensemble Learning

Ensemble-based learning is a supervised learning algorithm obtained by combining diverse models. In this paper, we focus on tree-based ensemble learning, which consists of multiple base models (i.e., DT model), each of which provides an alternative solution to the problem. Diversity among the models tends to make the prediction results more accurate . A single DT always suffers from high variance, which may cause instability in the prediction results. It is instructive to look at the psychological backdrop to this otherwise statistical inference. In our daily lives we use such an approach routinely by asking the opinions of several experts before making a decision (e.g., asking the opinions of several doctors before a major surgery, reading multiple user reviews before purchasing a car, or a paper that needs to be reviewed by several experts before being accepted for publication).

2.1.1. Random Forest

RF algorithm is built upon the idea of ensemble learning, which is a large collection of uncorrelated decision trees, each of which is capable of generating a result when estimated by a set of predictor values. The randomness in RF is to generate multiple datasets from the sample set, and the method is named bootstrap aggregating (bagging). Bagging is an ensemble algorithm designed to increase the randomness and improve the accuracy of machine learning algorithms. In the bagging process, the algorithm builds multiple models from the same original sample dataset to reduce the variance (shown in Figure 3). RF is an application of bagging in addition to building trees based on different bagging samples from the original training data. RF algorithm constrains the features that can be used to build the trees which forces trees to be different. To date, RF models have been widely applied to various research fields.

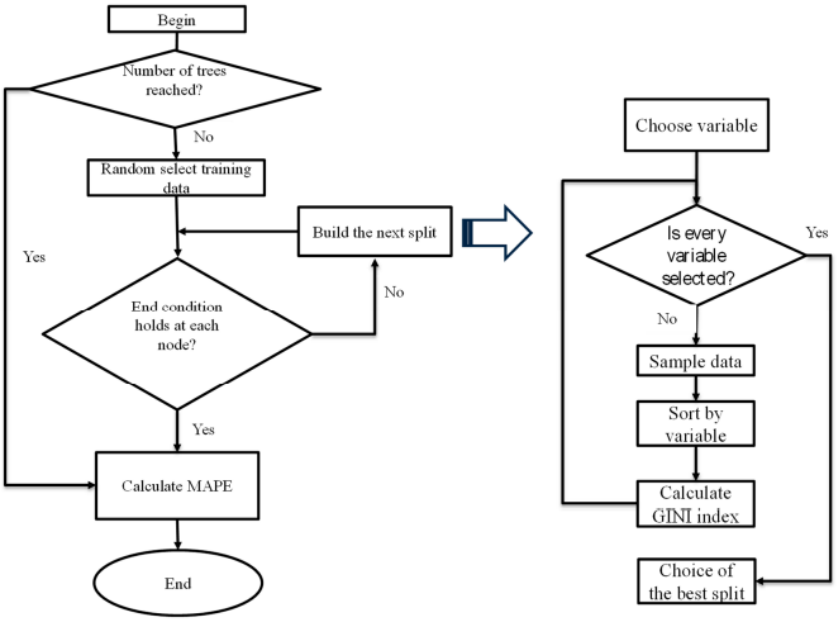

Fig:2.1 RF algorithm flow chart

2.2 Tourist prediction using Machine Learning algorithms

Research [2] states that the importance of tourism in the world helps the tourists to get used to the culture, customs, and conventions, language, and mode of living of the people at the destination. The advantages of tourism include job creation, foreign currency earnings, infrastructure development, poverty eradication, inequality reduction, and balanced regional development. In addition, tourism is important for creating world peace. Machine Learning is leading to technological innovation in all fields, with a great impact on the tourism sector. There is no difficulty in change diametrically the forms of commercialization and the way the travel industry works. In the last, tourism forecasting has attracted the study of various researchers mainly due to the importance of tourism in national economies. Using the time-series and regression methods have mostly dominated forecasting models of the current research approaches. Although these traditional techniques have proved some success in tourists forecasting new methods such as machine learning methods can very contribute to this area.

Algorithms of Machine learning especially for prediction

In this section, we discuss the different types of machine learning techniques and explain how they have been used for analysing data related to tourism. Usually, two types of machine learning activities are common in tourism – association learning and classification learning. In association learning, the learning method searches for associations or relationships between features of tourist behaviour. Also, this known as unsupervised learning. The second type of machine learning is classification learning. This learning scheme takes a set of classified examples from which it discovers a way of classifying unseen examples. This is a form of supervised learning. In the next, we explain the various machine learning techniques used in tourism. There are three used uses of machine learning techniques in tourism are (1) forecast expenses of tourists, (2) analysing profiles of ICSG 2020 K O C H I 2 0 2 0 tourists, and (3) forecast the number of tourist arrivals. In this section brief for the ten machine learning techniques are used to support these activities.

1. Logistic Regression: In logistic regression, there is a lot of data whose classification is prepared by building an equation. It is applied to predict discrete values "Binary values like 0/1, yes/no, true/false" based on a set of independent variables. Because it predicts the probability, its output values between 0 and 1 as expected. Logistic regression all generates the coefficients to predict a logit transformation of the probability.

2. Linear Regression: In linear regression, to build a model (equation) based on the data. Then use a linear regression model to make predictions about one variable based on particular values of the other variable. The variable is making predictions about is called the (dependent variable) also referred to as (y), the response variable. The variable that is using to make these predictions is called the (independent variable) also referred to as (x), the predictor variable.

3. Decision Tree: it is one of the supervised learning algorithms. The decision tree can also use in classification and regression. In a decision tree building algorithm first the best attribute of the dataset is placed at the root, then the training dataset is split into subsets. Splitting of data depends on the features of datasets. This process is done until all data is classified and find the leaf node at the various branch. Information gain can be calculated to find which feature is giving us the highest information gain. Decision trees are created for making a training model that can be used to predict class or the value of the target variable.

4. Support vector machine: The support vector machine algorithm is a binary classifier. Support Vector Machine is an algorithm announced by Vapnik in 1995. SVM is so popular a machine learning technique that it can be a group of its own. It uses a separating hyper plane or to define decision boundaries among a set of data points classified with different labels. It is a strictly supervised classification algorithm. In other words, the algorithm improves an optimal hyper-plane utilizing input data or training data and this defines decision in turn classifications new examples. Based on the kernel in use, SVM can perform both linear and nonlinear classification.

5. Naive-Bayes: It is a supervised algorithm for building classifiers based on Bayes theorem used classification methods. Naive Bayesian model is easy to build and especially useful for ICSG 2020 K O C H I 2 0 2 0 large data sets. In this algorithm occurrence of each feature is independent of occurrence various features. It needs a small number of training data for classification, and all terms can be pre computed thus classifying becomes easy, fast and effective. Also, Naive Bayes is an advanced classification method. Bayes theorem provides a method of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c).

2.3 Machine Learning based Tourism recommendation system

Previous publications that are briefly covered in this section illustrate how recommendation systems are applied in the tourism industry. Many systems employ machine learning, while some apply Deep Neural Networks for their working. Lucas et al in their study, applyclassification based on association in a recommender system for tourism, implementing a hybrid recommendation technique. It is named Personalized Sightseeing Planning System . A. Umanets and others developed an application called Guide Me and its primary features include its ability to integrate with social networks. This prototype works on both Android and iOS in helping suggest unexplored tourist destinations based on user ratings . Employing the Amazon Reviews dataset,. Kulkarni et al rank tourist places based on the positive and negative reviews about them. They use a deep learning algorithm in arranging the sequence of POIs . Proposing an RS using Social Network analysis for the city of Seoul in South Korea, Jeong et al argue that the personality type of the tourist plays a major role in selecting a tourism destination . Wang processed users' information like demographic variables viz age, gender, profession or occupation and city, and review data to provide a system to

recommend personalized travel products. The researcher used a total of 1,283,715 reviews for his study [13]. G and H. Verma et al focussed on rural tourism in India using opinion mining. They employed a supervised machine learning method to categorize the sentiments from a variety of travel-related companies and websites, reviews of hotels, and tourism agencies left by foreign travellers. “Term Frequency - Inverse Document Frequency (TF-IDF)” metrics are calculated to find the relevance of the words identified for sentiment analysis, and a robust model is proposed [5]. Offering a combined Lexicon-based and rule-based NB sentiment analysis approach to extract tourist characteristics from mobile apps, Muthukrishnan and others in their paper collect different reviews on Twitter connected to travel and processes them to identify the polarity and categorize each into different sentiments [6]. In their paper, Zelenka et al. present a conceptual framework for offering validated destination assessments and verified ratings of tourism services. Tripadvisor and Booking.com are the two case studies used in the main research. A trust model is created after conducting a SWOT analysis of the review and verification processes of the aforementioned websites [7]. To define the geographical, temporal, and demographic tourist flows throughout a tourist region, the research by Paolanti and others presents and evaluates a deep learning geo data framework. The dataset included Electronic copy available at: https://ssrn.com/abstract=4286575

2.4 Machine Learning Algorithms for building Recommender Systems.

A. Collaborative filtering (CF)CF refers to user-to-user association [8-9]. It follows the concept that if two or more folks have identical interests in one area then there is a likelihood that they will get attracted towards similar products or items of some other category as well [3-4]. Implicit and explicit user ratings are considered to compute similarity between two or more users. While implicit ratings are derived from user browsing pattern and click- through rate, explicit ratings are delivered by the user himself. The options of people you may know, suggested posts, similar pages you may like, suggested pokes, displayed on Facebook, are the examples of collaborative filtering. These are nothing but recommendations, based on features like number of mutual friends, similar pages liked or number of mutual groups, locations a user have been to or belongs to etc. For example, if two users have mutual friends, then the possibility is that they two may know each other as well.

B. Content-based filtering (CBF) CBF follows the idea “Show me more of what I have liked”. These systems recommend only those products to a user, which are alike the ones they enjoyed in the past . The similarity between two or more products is computed based on the features they have in common. While browsing through videos on YouTube, the browsing pattern of the user is observed to find out the types of videos he or she prefers and based on that, he or she is recommended similar kind of content under the feature suggested videos. Therefore, CBF based systems assume that if a user likes an item from some specific category, probability is that he or she might show his interest in another item from that category as well.

C. Knowledge based systems (KBS) KBS follows the notion “Tell me what fits my needs” . These systems generate recommendations based on a particular domain knowledge or domain expert . The user specifies his needs to the system which further compares those needs with its knowledge base and provide the most relevant suggestions accordingly. While buying anything from an e- commerce site, the users are expected to specify the desired features of the product they wish to buy, like the price range, color, size etc. and based on the resemblance between the features specified by the user and the properties of the product, the most suitable products are recommended.

D. Hybrid recommender systems Hybrid systems are the combinational systems, the inspiration following which is to merge the characteristics of two or more recommendation techniques in such a way that the limitations of a lone recommendation approach are conquered by the other. Netflix is the most popular hybrid recommender based on collaborative and content-based approach. It suggests movies or series to a user as per his interests, view history and the similarity between him and other Netflix users. Consider a user likes PS I Love you, The Notebook and The Fault in our stars then as he subsequently uses Netflix, he is recommended movies belonging to Romantic genre. Similarly, if two users have liked or viewed similar content on Netflix then each of them would be suggested what the other views next

2.5 Tourism Recommender System using Machine Learning

A. Recommender System

There are 2 main purposes for developing recommender systems [3]. Firstly, the capability to "predict" a user's interests and preferences by investigating the behavior of this user and/or the behaviour of other users in the same context to generate personalized recommendations. Secondly, the ranking version of the problem, it is also referred to as the top-k recommendation problem. This method is not directly predicting the user with the specific answer but recommended the top-k items to the user. Aggarwal [3] concluded that there are 5 basic models of the recommender systems as shown in Fig. 1. The collaborative filtering model is making a recommendation based on the rating of user-item from multiple users. On the other hand, the content-based recommender system analyses the attribute information of the users and items which focus only on a single user rather than those of all users. In a knowledge-based recommender system, the recommendations are based on explicitly specified user requirements. Instead of using external knowledge bases or historical data. The demographic recommender systems using the demographic information about the user to learn and create classifiers that can map specific demographics to ratings or buying propensities. In a more complex recommender system, it uses a combination of the different aspects to create hybrid systems. It combines the strengths of various types of recommender systems to create techniques that can perform more robustly in a wide variety of settings.

B. Machine Learning Framework

Machine learning (ML) can be generally defined as "a computational method using experience to improve performance or to make accurate predictions [9], [10]". In this context, experience refers to past information as the electronic data collected. Therefore, its quality and size are the keys to the success of the predictions made by the learner. There are 3 categories of data in ML.

1) Training data: ML's algorithm will use this dataset to learn how to perform the specific tasks given.

2) Validation data: the data used to adjust the hyper parameters of a learning algorithm.

3) Test data: the data used to test the results from ML regarding the trained ML's model.

To date, there are several companies built the pre-trained ML frameworks to use for predicting specific tasks. It included the library, platform, models, and anything required to run the ML. The developer could access those ML via API(Application Programming Interface) or micro services.

CHAPTER 3

SYSTEM ANALYSIS

3.1 Existing System

The existing works have been focusing on how to find routes that minimize a single kind of trip cost such as trip time or distance, amongst others. Some only suggest the best time to travel to the destinations. The users will have to browse through different sites to obtain all information required for planning their travel.

- TripAdvisor: TripAdvisor is a popular travel recommendation platform that uses machine learning algorithms to provide personalized recommendations to travellers. It analyzes travellers' past behaviour, preferences, and reviews to suggest hotels, restaurants, and attractions.

- Booking.com: Booking.com uses machine learning algorithms to provide personalized recommendations to travellers based on their past bookings, searches, and reviews. The system also incorporates user feedback to improve its recommendations.

3.2 Proposed System

Therefore, taking all these factors into consideration we are developing an application that provides suitable solutions for the above mentioned problems so that people can make their travel experience a memorable one. We have trying to efficiently plan a route that includes most of the sites of the user's interest based on the geographical location of these sites. We have focusing on generating an aesthetic travel route, which covers the astounding tourist sites. We are focusing on suggesting the best time to travel to the destinations so that the user will not have to browse through multiple sites to acquire all the information required to plan his/her travel. Our application will allow the user to enter multiple preferences. The accuracy of the path that is obtained by our algorithm will be high and also the weather predicting classifier is able to predict the best months to travel from source to the destination.

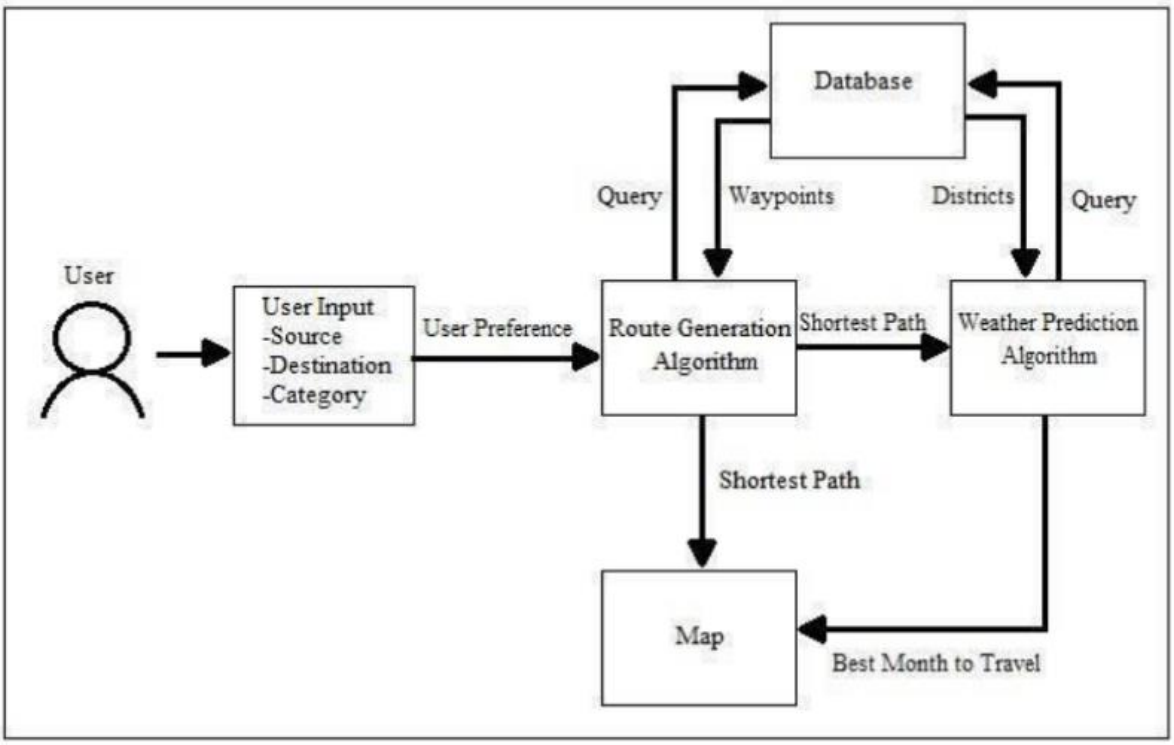

Fig:3.1 Proposed architecture design

CHAPTER 4

SYSTEM REQUIREMENTS SPECIFICATION

4.1 Software Requirements:

Software requirement specifies the minimum software that is required for the application to run smoothly. The following are the software requirement for our project:

- Windows 7 or higher
- Python
- Django framework
- MySQL database

4.2 Hardware Requirements:

Hardware requirement specifies the minimum hardware that is required for the application to run smoothly. The following are the hardware requirement for our project:

- Processor –Core i3
- Hard Disk – 160 GB
- Memory – 1GB RAM
- Monitor

CHAPTER 5

ADVANTAGES

- Personalization: To produce personalised recommendations, machine learning algorithms can examine a traveller's preferences, previous actions, and demographics. As a result, the recommendations are personalised to the person's interests, increasing the likelihood that they would enjoy their vacation experience.
- Accuracy: Machine learning algorithms can process large volumes of data and detect patterns that humans might overlook. As a result, the recommendations are probably more precise and pertinent.
- Time-saving: Travellers and tourism businesses can save time and effort by utilising machine learning to evaluate data and rapidly and effectively provide travel recommendations.
- Financial Gain: Travellers are more likely to book or buy when given accurate recommendations, which boosts income for travel agencies.
- Customized suggestions based on machine learning can raise consumer satisfaction, which can encourage loyalty and repeat business.
- Machine learning suggestions can be adaptive and dynamic, which means that they can be changed in real-time based on modifications to a traveller's tastes or behaviour as well as alterations to the tourist industry.
- Cost-Effective: Machine learning can eliminate the need for human labour and resources by automating the suggestion process, making it a cost-effective alternative for travel agencies.

CHAPTER 6

DISADVANTAGES

- Limited data: Machine learning algorithms depend on data to generate suggestions; therefore, if the data is scarce or of low quality, the recommendations may not be accurate or pertinent.
- Bias: If the data used to train machine learning algorithms is biased, it may result in suggestions that are unreliable or unfair.
- Lack of transparency: It might be tricky to comprehend how some machine learning algorithms arrived at their suggestions because they can be complicated and difficult to understand. Users may be hesitant to rely on them due to this lack of openness.
- Overreliance on technology: When making recommendations, tourism operators who depend too much on machine learning algorithms run the risk of overlooking other crucial aspects like intuition and human expertise.
- Lack of adaptability: Machine learning algorithms may not be able to respond to quick changes or unexpected events, resulting in suggestions that are out-of-date or irrelevant.
- Absence of human touch: Machine learning-based personalised recommendations can be missing the human touch that can make a trip genuinely special and unforgettable.

CHAPTER 7

EXPECTED OUTCOMES

- Personalized recommendations: The system can provide personalized recommendations to travellers based on their preferences, past travel history, and behaviour. This can help travellers discover new destinations, activities, and accommodations that they might not have considered otherwise.

- Increased engagement and satisfaction: By providing tailored recommendations, the system can increase traveller engagement and satisfaction, leading to more positive travel experiences.

- Improved efficiency: The system can help travel service providers, such as airlines, hotels, and tour operators, improve their efficiency by matching their offerings with the needs and preferences of travellers.

- Enhanced revenue: By offering personalized recommendations, the system can also help travel service providers increase their revenue by up selling or cross-selling relevant products and services.

- Better data analysis: The system can gather and analyze large amounts of data on traveller behaviour and preferences, providing valuable insights for travel service providers to make informed decisions.

- Increased trust and loyalty: By providing relevant and accurate recommendations, the system can build trust and loyalty with travellers, leading to repeat business and positive reviews.

CHAPTER 8

REFERENCES

- Article [1] Year: 2021, Authors: Bo Qiu and Wei (David)
- Article [2] Year : 2020, Authors: Bilal Adualgalil, Sajimon Abraham
- https://shsu-ir.tdl.org/shsu-ir/bitstream/handle/20.500.11875/1164/0781.pdf?sequence=1
- https://ieeexplore.ieee.org/document/6208293/
- https://ieeexplore.ieee.org/document/4679917/
- https://link.springer.com/article/10.1007/s11042-022-12167-w#:~:text=Tourism%20Recommendation%20System%20(TRS)%20is,the%20trip%20a%20memorable%20one
- https://www.ijert.org/tourism-recommendation-system-a-systematic-review
- https://onlinelibrary.wiley.com/doi/full/10.1002/aaai.12057
- https://www.hindawi.com/journals/cin/2022/1424097/
- https://ieeexplore.ieee.org/document/9310777
- Machine Learning Algorithms for building Recommender Systems

  Author : Richa Sharma Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India richa.sharma@chitkara.edu.in

- Deep Learning based Tourism recommendation system

  Author : Ismat FathimaResearch Scholar, Department of Computer Science & Information Technology, Maulana Azad National Urdu University, Hyderabad. India.

- Tourism Recommender System using Machine Learning

  Author : Charnsak Srisawatsakul * Faculty of Computer Science Ubon Ratchathani Rajabhat University Ubon Ratchathani, Thailand