Intelligent Tourist System For The 21st Century

Brindha Devi V, Meenaloshini M, Sriram G and Subhashini R

Sri Sairam Institute of Technology, Chennai- 44, India

E-mail: hodit@sairamit.edu.in

Abstract. Travel Industry is always at the forefront of technology adoption. Most of our researchers and town planners have preferred the method of machine learning approach for tour recommendation models. Some traditional methods have achieved certain levels of success in tourism research, but sometimes artificial neural network (ANN) and regression analysis techniques give better results. Our objective is to investigate the different ways in which the machine learning models can be applied in tourism prediction problems and to show the performance of machine learning methods. In this paper we have an intelligent tourist system using ML/AI that has been modeled based on various features using Sentiment analysis and can be performed over these reviews which will be helpful to find tourist place popularity. Based on sentiment analysis results, tourists can easily decide their own tour destination to be visited. We have categorized the places based on their preference like the most visited or least visited family trip or friend’s trip and so on. The dataset was collected from various tourism review websites. We performed a comparative study of feature extraction algorithms that is Count Vectorization, TFIDF Vectorization, along with classification algorithms like Naive Bayes, Support Vector Machine, and Random Forest algorithm. With these results, a recommendation system has been built which would map an individual user’s interests to the highest rated tourist places and generates a tour plan satisfying the user’s needs.

1. Introduction

AI in travel and tourism is used to predict a lot of things like the travel choices, personalize services, bookings and manage in-trip etc. Machine Learning is a subset of AI. In nutshell, ML is about building models which predict the result with high accuracy on the basis of the input data. Using these statistical methods, it would enable machines to improve its accuracy as more data is fed in the system. The machine learning models final output depends on:

- Quality of the data
- Features
- Algorithm

More data is diverse and rich, better the machine can find patterns and more precise the result. For example, here are some ways as to how and which data can be captured by travel industry providers: The datasets of good quality are in high demand and the companies literally have to hunt for the decent datasets. A detail gist in concern with the quality of the data explained in Figure 1.
Features are existing data that contain parameters like user gender / location extension. To build the model, data has more information, so it’s necessary that all the important features are being selected. Either analyst or modeling tool selects or discards the attributes depending on how useful they are for the analysis at the time of this process. Some of the main processes in Feature Validation are explained in Figure 2.

A number of features make the algorithm work slower, so the process of data preparation and to have neat .xlsx and .csv files at the end, takes some time than the whole process of training. Algorithm that analyzes the data looks for patterns, trends and then finds parameters for creating a model. It’s a tough job to choose the best algorithm to solve the specific task, because each algorithm can generate different results and some of them generate more than one result.

Machine learning models can outperform the classical rigid business intelligence where business rules wouldn’t capture the hidden patterns. To dig deep in available data and optimize the flow in their websites and apps, deliver truly superior experience, travel companies are actively implementing Artificial Intelligence & Machine Learning. This is how a ML-powered model has been built. Figure 3 gives a sequential explanation on our model.
2. Literature Survey

Information Technology has always helped to satisfy the needs for accurate and timely matters with the help of clients and the IT domination in the travel and tourism sector has foresaw a growth at an incredible rate. [1]

The contribution of Artificial Intelligence is extremely important for attributes like individualized pricing, recommendations in bundling products, reversed multi attribute auctioning, mobile applications and Semantic Web applications that would have the strength to take the company to great heights. But this process is not a bed of roses and it’s full of thorns and pitfalls. There is a large number of hurdles for Artificial Intelligence that the current techniques don’t take account of, such as the different nature of tourism department, sociology, the constraints on non-technical issues, economics, statistics, law, management science, and psychology which traditional models overlook. [2]

Right from the 2nd part of 20th century, Tourism has become one of the very important economic activity, giving about 11% of the Global Gross Domestic Product with about 1 billion people migrating globally. It is possibly the only department with significant contributions to the indigenous economy.[3] The main purpose of exact tourism forecasting is majorly vital due to the market’s important contribution to the nation’s GDP.

The Internet of Things (IOT) plays the role of bridge to fill the gap between our daily activities and the internet system. So this will create an environment to get all the non-connected things from far off distant coordinates. Changes with the development along with the new activity is inexorable for the human being have created a Smart Car Parking System (SCPS) using the IR (infrared) sensor with the applications of IOT, which allows the person to get the nearest allotted parking area, and provides the no. of vacant places present in that parking arena. This paper mainly cornerstones on removing the actual time involved in finding the parking space and the unwanted travelling. So, we have used this system in our application so that all the tourists will get a proper idea to park their vehicles before moving to the actual location.

Content is the king when it comes to brand-customer interactions, out of which the travel & hospitality industry is not an exception. The engineering team moves towards this task with the help of Deep Learning standpoint.[4] Machine Learning is a superset of Deep Learning that is particularly on the neural network base creation and training set. They deployed a model that chose relevant and attractive photos and then arranged them in order of their priority in their website. [5]

In Comparison with healthcare, e-commerce, and banking it’s crystal clear that the tourism and travel industry does not have an extreme posh vendor landscape for Artificial Intelligence & ML related solutions. The main reason behind this is because of the overall consideration with all the other sectors. It is relatively a very small sector, and the majority of the actual venture capital funds with the center of attention on the startups.

With reference to the recent survey, by 2020 online travel sales are estimated to fore take $800 billion. These applications are regenerating the entire service encountered with the help of Artificial Intelligence & Machine Learning for the TSD (travel software development). Nearly 90% of travelers from America post their travel experiences and reviews on social networking sites and travel agency sites. For example,
consider TripAdvisor, which has a visitor count of nearly 400 million. It is said that nearly 280 reviews are posted per minute on TripAdvisor.

Major benefits of Artificial Intelligence and ML occur behind the actual frontend, that is the main reason [6] highlights more on startups and SMEs leveraging AI-based platforms which could enhance tourism and travel deals in accordance with their own conversions and fixed agreement on the backend section. [6] The very first time when neural networks were connected with tourism demand was by Law et al. [7] to foretell Japanese need to travel to Hong Kong.

There are some social media listening tools engineered for the travel industry. This can perform sentiment analysis and relate with traveler's experience to determine the time of posting status like before travel, after travel, during travel. When a tourist posts his/her irritation in a social networking site, these social media listening tools predict them and provide real-time support to sort out the tourist's issue. Thus, good impressions are created among tourists. Studies indicate that more than 75% of tourists have bad experiences with their travel agency. 27% of such tourists avoided booking in the same agencies. Those agencies that try to make their services good have a higher chance of gaining back trust. AI chat bots can be used for such purposes. They can use Natural Language Processing to aid tourists. [9]. with the help of experimental results, the neural network model predicts the exponential smoothing, moving average, and outperforms multiple regressions. This Law will surely extend to the applicability of neural networks in the fields of tourism.

The system provides a clear overview of an intelligent tourist system. Various packages in the intelligent tourism model system and their points of interest are well described. Some of them are database thinking, recommender systems, information delivery for tourism etc. From their point of view, Vacation Coach's Expert and Travel Hop's Trip Matcher are the top two recommender framework advancements. [8]

Machine learning when applied in the travel industry can perform wonders. These wonders include messaging capabilities and bundling of products. These are essential for travel agencies to provide better services and deals to tourists from their past travel history. A survey conducted by Mind Tree reveals that only 23% of offers are rated excellent as they are based on tourist's preference. Also only 30% of tourists use them. Those who don't utilize such offers have their complaints. Less than 50% of them say that offers are not provided at required time. 35% of them reveal that these offers have a very small span of validity. [9]

3. Implementation Details
The proposed model makes use of suggestions from the native people living near tourist spots. Local businesses near that tourist place can be promoted based on the ratings they have got. Using the model, a travel plan can be created which estimates time duration of trip and suggests places to visit, estimated costs of transport and others. For this, previous trip details of tourists are taken into account. Reviews and ratings of people who had already visited that spot are used in prediction. It also provides central authority contact details in times of emergency. This can act as a virtual guide stating historical facts of tourist places. This can recommend other nearby tourist hotspots to be visited next. Popular Vlogs of tourist spots are presented to them if available. Tourist guides and photographers can be recommended upon request. Also refreshment spots can be located between tourist spots. Details of scammers in the areas near tourist spots are made available to the tourists on safety grounds. Remote places with historic values which are not known by many can also be covered in this model.

Our Application has the following components:
1. **Data framework component** - to provide training & testing data
2. **Natural Language Processing (NLP)** - to process sentiment and semantic analysis
3. **AI/ML** - to make predictions based on trained data to help tourists

Google Maps API is used to gather information about tourist spots. This information includes ratings, distance etc. This also calculates distances between tourist spots. Such information gathered constitutes the data framework. And this forms the first layer. NLP helps users to communicate with bots by understanding their speech. There are some open source libraries that can handle NLP tasks. These are already trained
with some commonly used phrases and sentences. It eases communication of users through their voice itself. And this forms the second layer. Next AI layer is the layer where actual prediction takes place. Models are trained with accurate data for accurate predictions. This is where magic happens. It can recommend people to visit places based on their interests. It is adaptive to the user's interests and can modify itself based on the situations. And this forms the last layer in our system.

4. Process Diagram

Figure 4. Process diagram of the proposed system

Our system will provide suggestions based on various parameters, and yield results accordingly. Our “Intelligent Tourist System For the 21st Century” has two parts while suggesting the locations. A set of attributes are used before the selection of the location and based on all of these filters we will get a set of locations, which will be further classified based on the after selection of the location. Now, in this classification we will rank the locations based on the safety and travel requirements.

Some of the parameters for suggestions are nearby locations, locations based on previous visits of the tourist and locations based on amount. Further considering the safety of the tourist we have included contact details for emergency, scammer alerts within our system. For all the young tourists we have added Vlog prompts and comments from the native people so that they don’t miss out any indigenous spots in their trip. Apart from this we take care of the transportation and food also. We have included all the possible public spots and parking locations for private transportation. Figure 4 gives a brief overview of our system.

5. Performance Evaluation

Following are some machine learning and deep learning algorithms to get models which perform better with greater accuracy. Bernoulli Naïve-Bayes is meant for discrete type of data. In this one, Boolean variables (True/False) are used in order to describe features. These features also describe the inputs for the model. The main purpose of this model is to perform classification of documents. For this document classification, occurrences of terms are made use instead of frequencies. This depends on Bernoulli distribution. Let a Boolean variable $x_i$ represent occurrence of $i^{th}$ term in a collection, then probability of class $y$ to generate term $x_i$ is said to be,

$$P(x_i | y) = P(i | y)x_i + (1 - P(i | y)) (1 - x_i)$$ (1)
This model is suitable for accurate short text classification. This model proved to get accuracy nearly to 83%. This can also model the terms that are not present. Parameters used in prediction can take only binary values.

In Multinomial Naïve-Bayes algorithm a multinomial event model, features are represented in frequencies of events. Let us consider a feature which is represented using a vector \( x=(x_1, x_2, x_3... x_n) \). This can be considered as a histogram where \( x_i \) represents frequency of event \( i \). Like Bernoulli Naïve-Bayes, this can be used for document classifiers.

The probability may become zero, if a class and feature are not occurring together in the dataset used for model training. This is due to loss of information probability when a class and feature never occur together in training data. This problem can be countered by including a pseudo count (little adjustments in the data sample). This counter action is known to be Laplace Smoothing. This model has accuracy up to nearly 84%.

Random Forest Classifier method involves building decision trees for classification. A flow chart like tree structure with each non-leaf/internal node describing a conditional check on attribute and leaf node describing a classification term is called decision tree. The concept behind this classifier is the wisdom of crowds. This consists of a large number of decision trees where each individual tree is used in prediction of a class. The class with the majority of predictions is considered as a final predicted result. This relies on multiple decision trees. Accuracy of this type of classifiers can be up to 85%. Various processes in random forest algorithms are covered in Figure 5.

![Figure 5. Prediction Tree For Random Forest Algorithm](image)

Convolutional Neural Network also known as CNN consists of layers like input layer, output layer along with some hidden layers which include convolutional layers, RELU layers, normalization layers, pooling layers and fully connected layers. This is a deep learning algorithm which takes an image and specifies bias to objects in the image and classifies them. Little amount of preprocessing is required in CNN. Each image is passed through layers with filters for feature detection which in turn can be used for classification. CNN works similar to the visual cortex. CNN works with usage of cross correlation when considered mathematically. The detailed explanation of the process is covered in Figure 6. This is mostly used in image processing, discovery of drugs, NLP and analysis of video. CNN models have accuracy up to 94%.
Long short Term Memory RNNs perform far better than traditional RNNs. This uses feedback for betterment. This can process large data sequences like videos. Gradients from the objective function may explode after multiple rounds of multiplying with network weights. This is a problem faced in traditional RNN based NLP task systems. To avoid this issue, use of traditional RNN is avoided in text classification. Long Short-Term Memory RNN is apt for text classification while avoiding the problem stated above. This is possible due to the presence of input gates, output gates that control information flow. Memory cells are considered important in this one. Applications include handwriting recognition, music generation, language translation and image captioning.

6. Comparison Between The Algorithms
Bernoulli Naive Bayes and Multinomial Naive Bayes belong to Naive Bayes algorithms. Convolutional Neural Network and Long Short-Term Recurrent Neural Network belong to Neural Networks. Random Forest doesn't belong to Naive Bayes algorithms or Neural Networks. Bernoulli Naive Bayes or Random Forest is used when features are binary in nature. Multinomial Naive Bayes is used when discrete values that follow multinomial distribution are involved. Convolutional Neural Network is inspired from the human brain. Long Short-Term RNN is an improvised version of Recurrent Neural Network.

Bernoulli Naive Bayes uses Bernoulli distribution whereas Multinomial Naive Bayes uses Multinomial distribution. Random forest utilizes prediction models. Convolutional Neural Networks contain filters which are their building blocks. Long Short-Term Recurrent Neural Network uses cell, input gate, output gate and forget gate. Among these algorithms, only Random Forest makes use of Decision Trees. Bernoulli and Multinomial Naive Bayes are used for document classification. Random Forest is used for classification and regression applications. Convolutional Neural Network is used for image and video processing. Long Short-Term Recurrent Neural Networks are used for speech/voice and handwriting recognition.

7. Dataset
The dataset that is to be used has the following features: category, distance, duration, nearby Places, rating and count. Category feature defines the type of tourist spot that makes recommendation based on the taste of tourists easy. Distance feature defines the distance of a tourist spot. Duration feature defines the time duration to reach that spot. Nearby Places helps to find nearby spots suitable for tourists thus saving their
time and money. Rating feature helps to recommend spots to tourists. Count feature denotes the number of visitors.

8. Result Analysis
With all the above analysis, we would like to conclude that convolutional neural network and Random Forest Algorithm is comparatively better than all the other methods. And with the number of parameters (the values), the efficiency of the system increases and would produce perfect results according to the recommendations.

![Performance Evaluation](image)

**Figure 7.** Performance Evaluation of Proposed System

9. Conclusions
We investigated how predictions can be made on tourism using different machine learning models. We were focusing on five commonly used machine learning techniques. Additionally, we also investigated the effect of time index when it is being included as an input variable. This study is limited in the number of originating countries, time span, and sample sizes, but findings should be of use to both tourism practitioners and researchers who are interested in forecasting using ML methods. Machine learning (ML) models were not used in tourism demand forecasting except for multi-layer Perceptron models. Our findings reflect that when the available data are very small, machine learning models other than MLP (Multilayer perceptron) models can give better performance. There is no method that can be considered the best one for all the datasets, as it is totally dependent on the dataset. In the existing system many things related to artificial intelligence are available. The existing system is having many applications of Artificial Intelligence but they are all separated. Each and every aspect is not connected which is the major flaw. The main aim of our project is to bring everything together in order to integrate each and every part available into a fully-fledged model in our system.

References
[1] Chandra Mohan, B., 2015 “Restructured Ant Colony Optimization routing protocol for next generation network”, *International Journal of Computer Communication and Control*, 10(4), 493- 500
[2] Chandra Mohan, B. and Baskaran, R., 2012 “A Survey: Ant Colony Optimization based recent research in various engineering domains”, *Expert System with Application, Elsevier*. 
[3] Cenamor, I., de la Rosa, T., Núñez, S., Borrajo, D., 2017, “Planning for tourism routes using social networks”, *Expert Systems with Applications*, 69 1-9.

[4] Belk, R. (1975), “Situational variables and consumer behavior”, *Journal of Consumer Research*, 2(3)157-164.

[5] R. Law and N. Au, , 1999 “A neural network model to forecast Japanese demand for travel to Hong Kong”, *Tourism Management*, 20, 89-97.

[6] R. Law and N. Au, , 2000 “Back-propagation learning in improving the accuracy of neural Network-based tourism demand forecasting”, *Tourism Management*, 21, 331-340.

[7] D. C. Frechtling, 2001 “Forecasting Tourism Demand: Methods and Strategies”, *Oxford: Butterworth-Heinemann*.

[8] Aguilo, E., Alegre, J., AND Sard, M. 2007, “The Persistence of the Sun and Sand Tourism Model”, *Tourism Management*, 26(2), 219-231.

[9] Bigne, E., Sanchez, I., AND Sanz, J. 2001, “Tourism Image, Evaluation Variables and after Purchase Behaviour: Inter-Relationship”, *Tourism Management*, 22(6), 607-616.

[10] Bigne, E, Sanchez, I., AND Sanz, J. 2009, “The functional-psychological continuum in the cognitive image of a destination: A confirmatory analysis”, *Tourism Management*, 30(5), 1-9.