



4<sup>th</sup> National Conference on Advances in Engineering and Applied Science  
Organized by : Anjuman College of Engineering and Technology (ACET) Nagpur,  
Maharashtra, India, In association with  
International Journal of Scientific Research in Science and Technology



## Data Mining for Sales Prediction in Tourism Industry

Prof. Sadia Patka, Abdul Kalam, Rajik Sheikh, Rohit Raut, Swarnabh Gokhale

Computer Science and Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra,  
India

### ABSTRACT

The tourism industries have widely adopted the information technology (IT) to reinforce their operation efficiency and to enhance service quality and customer experience. Tourism industry often use the web as a channel of communication with the targeted customers, because Internet are often easily accessible, cheaper and friendly to user. A system is provided for accurate and reliable prediction to assist the choice makers to plan more effectively and efficiently by providing an admin with previous sales results and therefore the user with positive or negative reviews for better deciding . this technique are going to be using ID3 algorithm to build a decision tree supported a review data and illustrate how this procedure are often used to make a choice on an action supported the present data using previously collected data.

**Keywords :** Sales Prediction, Tourism Industry, Decision Tree, Reviews

### I. INTRODUCTION

Tourism, within the past few decades, has emerged together of the main industries within the world economy by benefiting transportation, accommodation, catering, entertainment, retailing, employment and lots of other sectors. The tourism industries have widely adopted information technology (IT) to reinforce their operation efficiency and improve service quality and customer experience. Tourism firms often use the web as a channel of communication with the targeted customers, because Internet are often easily accessible, cheaper and friendly to user. Internet usages on the private computers (PCs) are shifted to smart phone which are capable and more versatile. Traveler behaviors tend to vary when new technologies come. they're going to plan less and look for information at the purpose of activities. Previously, they might search information from PC then switch to look on a mobile device

instead because it's more flexible and convenient. this is often according to a study by Hyde (2000) indicated that traveler avoids vacation planning because flexibility of action and experiencing the unknown are essential among-est the hedonism experience they're seeking.

The system works basically on predicting the trends occur within the tourism industry on basis of review of the tourist visited to explore the planet . This review will predict the potential of the places to hike within the business for the service provider .The system provide useful information of knowledge to predict the foremost promising event within the future response. The analytics diagram, graph, curve trends will help to know the scenario of industry.

### II. LITERATURE REVIEW

#### 1.1. Existing System

Tse and Poon (Tse and Poon, 2015) [1] used one year long reservation data provided by Hotel ICON in Hong Kong . They observed certain trends in data and developed a regression model employing a second order equation. Their proposed system predicts hotel reservations using reservation curves from the past data. To predict a selected day's reservation, they use a 90-day window. Authors do daily and weekly predictions, and observe that their model performs better on weekly predictions.

Ellero and Pelegrini (Ellero and Pellegrini, 2014) [2] assess the performance of various widely-adopted models from literature to forecast Italian occupancy rate . They find that exponential smoothing, advanced pick-up, and moving average models show the simplest success within the compared models.

Shenoy et al.(Shenoy et al., 2017)[3] demonstrate their estimation of reservation information supported user activity and search results using the info provided by Expedia. Their studies show that acquisition of serious results becomes possible through clustering and ensemble operations.

Song and Li (2007)[4] included in their review was "A practitioners guide to time-series methods for tourism demand forecasting -a case study of Durban, South Africa " by Burger, Dohnal, Kathrada and Law (2001). the target of the study was to conduct a forecast folks demand for visit Durban, South Africa. during a review of the planet Tourism Organization in 1995 about African tourism, South Africa was considered to be "one of the foremost promising tourism destinations on the African continent" but it's not been ready to realize its full potential yet.

### III. PROPOSED WORK

#### 3.1. Problem Definition

one among the most purposes of the web site is to facilitate the offline customer online because customer cannot spend their precious times in marketing research . therefore the customer take help outsourcing work which help to enhance the industry business.

There will be many users visiting the web site and hence we require a robust and reliable frontend which may withhold the users on our site. the info we'll be consisting is extremely important which will help everyone

#### 3.2. Aims and Objectives

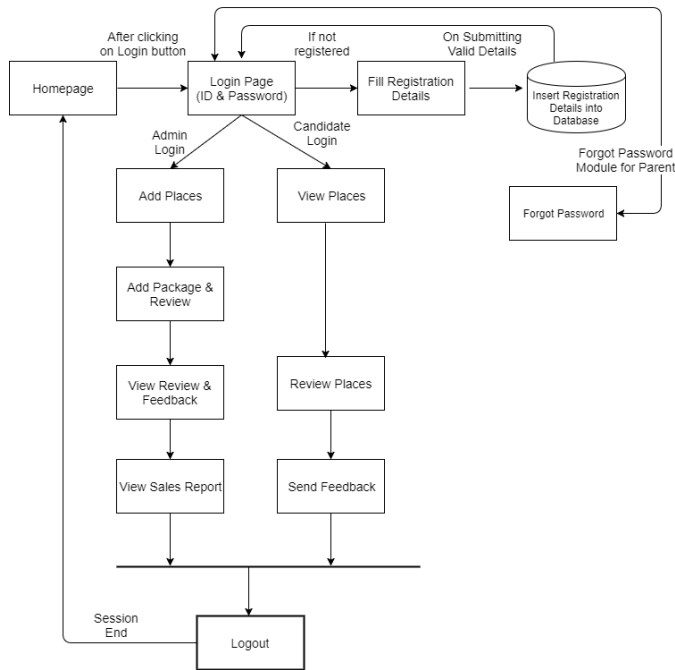
The main idea behind the project is to predict the sales from analyzing the previous data and giving the anticipated value of sales.

This project have been proposed to bring more detailed and promising features of Tourism industry at the finger tips of our users at some ways imaginable

#### 3.3. PROPOSED SYSTEM

Considering the anomalies within the existing system computerization of the entire activity is being suggested after initial analysis .The android application is developed using Visual Studio with Asp .Net with C# as a programing language.

Proposed system is accessed by two entities namely, Admin and User. Admin got to login with their valid login credentials first so as to access the online application .



**Fig 1.** System Architecture

If a user visits an area, he/she can write a review about the visited place. supported user’s reviews, system will predict sales for a specific place.

This will be done by a resulting graph which can be generated by order count and therefore the user reviews. the essential keywords of the reviews are assigned a score upto 10 within the database which can help within the generation of the graph.

word	score	flag
not bad	5	1
not good	-1	0
too bad	-2	0
very good	8	1
very impressive	10	0
very nice	7	0
very beautiful	7	0
very excellent	10	0
most beautiful	9	0
NULL	NULL	NULL

**Fig 2.** Some Keywords with its associated score

### 3.3.1. Algorithms Used For Proposed System

#### Decision Tree :

An Overview of Decision Tree :

Decision Tree may be a prediction model representing the mapping relation between object properties and object values. Given the certain probability of all situations, it's specifically utilized in decision analysis to calculate the probability of zero-or-greater arithmetic mean of net present value, assess the project risks and judge whether a project is desirable or not.

#### Basic Strategy of ID3 Algorithm :

Starting with one node of the coaching examples, if the examples are within the same class, the node is known as a leaf and delineate by a corresponding image. Otherwise, the rule employs the entropy-based activity info[of data|of knowledge] gain as heuristic information, associate degree selects an attribute that may most satisfactorily classify the examples. The attribute represents “test” or “decide” on the node. during this version of the rule, all attributes are classified, i.e. separate values. for every famous worth of the take a look at attributes, a branch is made, supported that examples are divided. Following constant method, the choice tree of every division is generated recursively. Once associate degree attribute seems in an exceedingly node, it mustn't be thought-about once more in any descendants of the node.

ID3 rule can perform following tasks recursively:

1. produce a root node for the tree
2. If all examples are positive, come leaf node ‘positive’
3. Else if all examples are negative, come leaf node ‘negative’
4. Calculate the entropy of current state  $E(S)$
5. for every attribute, calculate the entropy with relevance the attribute ‘A’ denoted by  $E(S, A)$

6. choose the attribute that has the utmost worth of  $IG(S, A)$  and split the present (parent) node on the chosen attribute

7. take away the attribute that provides highest IG from the set of attributes

8. Repeat till we tend to run out of all attributes, or the choice tree has all leaf nodes

Entropy :

A decision tree is made top-down from a root node and involves partitioning the information into subsets that contain instances with similar values (homogeneous). ID3 rule uses entropy to calculate the homogeneity of a sample. If the sample is totally solid the entropy is zero and if the sample is equally divided then it's entropy of 1.

Entropy is that the live of the quantity of uncertainty or randomness in information. Intuitively, it shows sure thing of a definite event. If associate degree outcome of a happening incorporates a chance of 100 percent, the entropy is zero (no randomness exists), associate degree if an outcome is five hundredth, the entropy takes the utmost worth (i.e. equals to one since it's the log base 2) because it comes good randomness. as an example, think about a coin toss whose chance of heads is zero.5 and chance of tails is zero.5. The entropy here is that the highest potential worth (i.e., equals 1), since there's no likelihood to exactly confirm the end result. as an alternative, think about a coin that has heads on each the edges, the end result of such a happening is foretold dead since we all know beforehand that it'll continually be heads. In alternative words, this event has no randomness, thus its entropy is zero. ID3 follows the rule: a branch with associate degree entropy of zero could be a leaf node (endpoint). A branch with associate degree entropy over zero desires more cacophonous. just in case it's impracticable to attain zero entropy within the leaf nodes, the choice is

formed by the tactic of a straightforward majority.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Information Gain:

ID3 chooses the cacophonous attribute with the best gain in data, wherever gain is outlined as distinction between what quantity data is required once the split. this is often calculated by decisive the variations between the entropies of the initial dataset and also the weighted add of the entropies from every of the divided datasets.

The formula used for this purpose is:

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

This application includes the following five modules:

**Admin:**

- **Login:** In order to access the system, admin need to login first using valid id and password.
- **Add Places:** Admin can add Places on which users can provide their review.
- **Add Package:** By selecting the added places, admin can create a package with its details.
- **Add Reviews:** Admin can upload an Excel Sheet, Which will consist of Place, Month, and Review.
- **View Reviews:** Admin can have a look on the ratings provided to reviews, for a particular Place.
- **View Sales Report:** Admin will be shown a Bar Graph which will have Ratings for all the places out of 10.
- **View Feedback:** Admin can have feeds from users and he can even delete them if required

**User:**

- **Register:** User can register themselves and into the system using basic details.
- **Login:** User can Login into the system, using his email id and password.
- **Forgot Password:** If user forgets his password, he can press forgot password and password will be sent to his E - Mail.
- **View Places:** User can have a look on different Places added by admin and view comments on them.
- **Review Places:** User can add his review on a place only once, with details of Place Name and Month they visited.
- **Send Feedback:** User can send feeds to admin mentioning any issues they face, or any new places they want to add etc.

**III. CONCLUSION**

This project proposed the domain knowledge of a tourism industry during a single window. The admin/customer will have great advantage in increasing the efficiency of their business model. Since this status of the planet having large competition in tourism industry so it's worth to figure on data for accountability, liability and for broadening the boundaries of the tourism industry. The current status of the project is analyzing the review given by the tourist with module work of user feedback, view places for the various location.

**IV. ACKNOWLEDGEMENT**

This research paper is possible due to the extreme efforts taken by everyone. Without the support and help put together by each member including the

valuable guidance, profound advice and encouragement by guide Prof. Sadia Patka.

**IV. REFERENCES**

- [1]. Candel, A., Parmar, V., LeDell, E., and Arora, A. (2016). Deep learning with h2o. H2O. aiInc
- [2]. Shenoy, G. G., Wagle, M. A., and Shaikh, A. (2017). Kaggle competition: Expedia hotel recommendation
- [3]. Ellero, A. and Pellegrini, P. (2014). Are traditional forecasting models suitable for hotels in Italian cities
- [4]. <https://www.ijltet.org/journal/146475292884.pdf>
- [5]. [http://www.springer.com/cda/content/document/cda\\_downloadaddocument/9783319442624-c1.pdf?SGWID=0-0-45-1592857-p180195151](http://www.springer.com/cda/content/document/cda_downloadaddocument/9783319442624-c1.pdf?SGWID=0-0-45-1592857-p180195151)
- [6]. TianLinqin, Application of Data Mining Technology in Tobacco Industry, Journal of Agricultural Science and Technology, March ,2012
- [7]. Wang Ying, Li Renwang, Li Bin, Zhang Zhile, Costume Sales Forecasting Model Based on CURE Algorithm and C4.5 Decision Tree, Journal of Textile Research, September, 2008
- [8]. Zhang Gefu, Ou yang, Hao nan, Xu qi, Application of Decision Tree in Apparel Marketing Based on Appearance of Consumers, Journal of Computer Applications, July,2010