IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 4, April 2020

Recommender System based on Learning Techniques for Tourists

I.Harika[1], K.Tejaswini Reddy[2], J.Bhavya Suma[3], N.Maheswari[4], P.G.K.Sirisha[5]

UG Students, KKR & KSR Institute of Technology and Sciences[1,2,3,4]

Associate Professor, KKR & KSR Institute of Technology and Sciences[5]

Abstract: Choosing a tourist destination from the knowledge that's available on the web and thru other sources is one among the foremost complex tasks for tourists when planning travel, both before and during travel. Previous Travel Recommendation Systems have attempted to solve this problem. This paper proposes a novel Travel Recommendation System that recommends destinations to tourists that are mostly visited based on the tourists dataset taken. It considers both technical and practical aspects using a real world data set we collected. The system is developed using a two-steps feature selection method to reduce number of inputs to the system and recommendations are provided by decision tree C4.5. The experimental results show that the proposed Travel Recommendation System can provide recommendation on tourist destinations that are mostly visited.

Keywords: Travel Recommendation System, Destination, c4.5Decision tree, Feature selection, Tourists.

I. INTRODUCTION

The Internet is now considered to be the main information source of tourists for information on products and services. Due to the huge volume of heterogeneous information available on the Internet, the search for destinations, as known as travel planning can overwhelm tourists. The travel-planning task is complex and dynamic such that there are many factors involved when making a decision, for examples, the quality of the attractions, travel routes, hotels, numbers of traveler, leisure activities, weather, etc. Recently, tourism has substantially benefited from ICT, and especially from Internet technology. With the development of decision support tools, also known as Recommendation Systems (RS), tourists and tourism providers can search, select, compare, and make decisions more efficient than ever.

Most of the previous Travel Recommendation Systems have focused on estimates of choosing the destination, activities, attractions, tourism services (e.g. restaurants, hotels, and transportation) based on the user's preferences and interests.

The main theme of this paper is to propose a recommendation system for tourists that provide the most visited place to them based on the given dataset. This can be proposed by overcoming all the drawbacks in the previous travel recommendation systems. In the proposed recommendation system for tourists we have reduce the no. of parameters that reduces the complexity of the system. In return, the performance level of the system can be increased.

II. LITERATURE SURVEY

L. Sebastia et.al… [1] described that "Nowadays, most of the people who plan a visit or a day-out will first initiate an enquiry through the web. Travelers usually have a limited knowledge of the town to go to and that they are unaware of the local artistic, social or entertainment places. A user may find an outsized amount of data about the town, but he may invest an extended time selecting the activities he prefers and organizing them to profitably spend a day-out. E-Tourism may be a web application that generates recommendations about personalized tourist tours within the city of Valencia (Spain). It's intended to be a service for foreigners and locals to become deeply conversant in the town and plan leisure activities. E-Tourism makes recommendations supported the user's tastes, his demographic classification, the places visited by the user in former trips and, finally, his current visit preferences.

E. Pitoska [2] described that "Tourism is a combination of activities, services, and industries such as means of transportation, accommodation and entertainment places, sports centers, restaurants, shops, etc. Tourism is one of the biggest industries in the world and it is based on information. Information and Communication Technologies and the Internet are the most efficient ways to disperse any kind of information. Some of the advantages of e-tourism are the decline of seasonality, the more successful communication with the customers, and the rise in reservations and sales in general. By using Information and Communication Technologies, the customer can easily choose a destination, compare prices, and complete financial arrangements."

E. Pantano and L. D. Pietro [3] described that “This concept is to advance their knowledge on the extent to which tourists use social networks for both achieving information on possible destinations and for expressing adverse judgments, to figure out the main consequences for tourism marketers and possible solutions, as well as to describe and detail the prevailing shifting from e-tourism to f-tourism (from tourism-based one-channel to new tourism based on Face book).

Prof. P. A. Manjare et.al… [4] described that “-The recommender system is constructed as an online application which is capable of generating a personalized list of preference attractions for the tourist. Tourists can find tourism information on blogs, forums, websites of cases of interest, etc. However, erudition overflow can occur on the internet as there is still a loss of focus on the use of recommender technology in the tourism field. Recommendation of tour information is vital for users, for the recommendation system to succeed; it must be able to provide tourism information based on the user’s preferences and current location. There is also raising demand for more information on social area attractions, such as local food, shopping spots, places of interest, and so on during the tour. The goal of this research is to propose a suitable recommendation method for use in a Recommendation System Based on Tourist Attraction to provide personalized tourism information to its users.

R. Anacleto [5] described that “Tourist recommendation systems have been growing over the last few years, mainly because of the use of mobile devices to obtain user context. This work addresses some of the most important systems on the field and presents PSiS Mobile, which is mobile credentials and planning application designed to support a tourist during his vacations. It provides suggestions about points of interest to visit based on tourist choices and user and sight context. Also, it recommends a visit planning which can be dynamically adapted based on modern user and sight context. This tool serves also as a journey dairy since it records the tourist moves and tasks to help him retain how the trip was like. To conclude, some field experiences will be presented.”

G. Häubl and V.Trifts [6] described that “Recommendation system interactive effects are of two decision aids, each designed to assist consumers in performing one of the above tasks, on purchase decision making in an online store. The premier interactive device, a recommendation agent(RA), allows customers to more efficiently screen the (possibly very high) collection of alternatives accessible in an online shopping ecosystem. Based on self-explicated data about a consumer's service function (attribute charge weights and minimum tolerable attribute levels), the RA creates a personalized list of suggested alternatives. The second decision aid, a comparison matrix (CM), is designed to help consumers make in-depth correlations among selected alternatives. The CM enables consumers to make trait data about multiple products in an alternative × traits matrix and to have alternatives ordered by any trait. Based on analytical and practical work in marketing, examination and decision making, psychology, and decision support systems, we develop a collection of hypotheses about the results of these two choice aids on many aspects of consumer decision making. In particular, we focus on how the use of the RA and CM influences consumers' search for product information, the size and quality of their consideration collections, and the feature of their purchase decisions in an online shopping environment”.

F. Ricci, L. Rokach, and B. Shapira [7] described that “Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user. The suggestions relating to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read. “Item” is the archetypical term employed to signify what the operation sustains to users. An RS normally concentrates on a specific type of item (e.g., CDs, or news), and accordingly, its scheme, its graphical user interface, and the core recommendation method used to generate the recommendations are all customized to give useful and effective instructions for that specific type of item. RSS is primarily aimed towards individuals who lack adequate personal experience or competence to judge the potentially overwhelming number of alternative details that a Web site. A case in point is a book recommender system that assists users to pick a book to read. In the popular Web site, Amazon.com, the site employs an RS to personalize the online store for each customer. Since instructions are usually personalized, different users or user groups receive diverse suggestions. Besides, there are also non-personalized recommendations. These are much easier to produce and are normally featured in magazines or newspapers. Illustrative examples include the top ten selections of books, CDs, etc. While they may be beneficial and effective in specific situations, these types of non-personalized instructions are not typically addressed by RS research. In their simplest form, personalized recommendations are offered as ranked list”.

P. Resnick and H. R. Varian [8] described that “A Recommender System refers to a system that is capable of predicting the future preference of a set of items for a user, and recommend the top items. One key purpose of why we need a recommender system in modern society is that people have too several options to apply due to the prevalence of the Internet. In history, people used to shop in a physical store, in which the items available are inadequate. For example, the number of movies that can be placed in a Blockbuster repository depends on the size of that repository. By variation, nowadays, the Internet allows people to locate abundant resources online. Netflix, for example, has a huge collection of movies. Although the amount of available information increased, a new problem arose as people had a hard time selecting the items they want to see.

G. I. Alptekin and G. Buyukozkan [9] described that "The accelerating interaction between technology and tourism has changed radically the efficiency and effectiveness of tourism organizations, as well as how consumers interact with organizations. In this study, an internet-based intelligent framework for travel agencies is proposed that gives customers a quick and reliable response service during a less expensive manner. The suggested framework integrates case-based reasoning (CBR) system with well known multi-criteria deciding (MCDM) system, particularly Analytic Hierarchy Process, to increase the performance and speed just in a case resembling in tourism destination planning. The integration of the two techniques enables taking advantage of their strengths and complements each other's weaknesses. Case research is conducted to demonstrate how this framework can aid intelligent decision support by reclaiming best-fitted responses for customers".

J. B. Schafer, J. A. Konstan, and J. Riedl [10] described that "Recommender systems are being used by an ever-increasing number of E-commerce sites to help consumers find products to purchase. What began as a novelty has changed into a serious business tool. Recommender systems use result erudition—either hand-coded erudition produced by experts or "dug" erudition determined from the performance of consumers—to lead consumers through the often-overwhelming task of organizing products they will like. In this article, we present an analysis of how recommender systems are related to some classical database analysis techniques. We analyze how recommender systems help E-commerce sites improve sales and examine the recommender systems at six market-leading sites. Based on these cases, we create a taxonomy of recommender systems, including the data required from the consumers, the extra erudition required from the database, the ways the recommendations are introduced to consumers, the technologies adopted to create the recommendations, and the level of personalization of the recommendations. We place five commonly used E-commerce recommender purpose patterns, describe several open examination complexities in the field of recommender systems, and examine the privacy hints of recommender systems technology."

III. EXISTING SYSTEM

Many travel recommendation systems provide a destination to tourists based on user's preferences and interests. In the existing travel recommendation systems, they perform a basic matching mechanism between the user's constraints and the items by filtering and sorting mechanisms. But there are many technical aspects not present in them like system accuracy, scalability, user satisfaction, usability, etc…

3.1 Drawbacks in the existing system

- More no. of parameters is considered.
- Less accuracy.
- User satisfaction is less.

3.2 Proposed system

To overcome the above problem, we are using C4.5 decision tree algorithms that take experiences of previous users and then build a model and using a decision tree that will predict the best location based on his given input. Decision trees don't need new users' experience data. To implement a decision tree model we need to have a dataset and this dataset sometimes will have empty or garbage values and these values will have a bad effect on the decision tree model so we can remove such empty or garbage values by applying to pre-process techniques. Sometimes to predict or build model no need to use all columns (attributes) values from the dataset and these unnecessary attributes can be removed by applying features selection algorithms and here we are using MRMR features selection algorithms to remove unnecessary attributes to reduce the execution time of building model and to increase system accuracy.

3.2.2 Advantages of the proposed system

- Removes unnecessary data.
- Gives accurate results.
- Saves time.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 4, April 2020

3.3 Modules

Data acquisition

To understand tourist's search behavior in assessing travel information and decision-making processing for destination choice, we use a questionnaire as a data collection method due to its effective mechanism for collecting information from tourists. Pre-study on a variety of factors that influence tourist's preferred destinations were identified for questionnaire design.

Data Preprocess

Feature selection or variable selection is a process of selecting subsets of relevant features that describes the output classes. It is a very important process for not only utilization and usability, but also accuracy improving. It is to reduce the number of necessary user inputs as well as to increase the performance of the classification model. We propose a Feature selection method based on to rank the features and remove irrelevant and redundant features from the dataset.

Data Analysis

Decision trees are chosen as a classifier/model for the proposed TRS because it provides several benefits for decision-makers such as simplicity, interpretability. The decision tree consists of nodes and leaves. The first node is called the root node, where the instances from the test set start to navigate down to a leaf. To recommend a destination to tourists, we must traverse the decision tree from the root to the leaf. C4.5 supports two types of splitting criteria, including the information gain and the entropy-based criterion. It also supports both nominal and scale variables. To avoid the over-fitting problem, C4.5 supports tree pruning (e.g., confidence-based and error-based pruning). Moreover, C4.5 allows attributes to be missing.

IV. ALGORITHM USED

Algorithms are very important in computer science. The best-chosen algorithm makes sure the computer will do the given task in the best probable manner. In cases where ability matters a proper algorithm is vital to be used. An algorithm is important in optimizing a computer's performance according to the available resources.

C4.5 Decision tree is used, which is an extension of the ID3 Algorithm. This algorithm is taken because C4.5 tried to solve the drawbacks of the ID3 Algorithm. ID3 is the most simple decision tree algorithm but has many drawbacks such as that the optimal solution is not guaranteed, the over-fitting problem with the training data set, and it supports only nominal variables. On the other hand, C4.5 supports two kinds of splitting criteria that include the information gain and the entropy-based criterion. It also supports both nominal and scale variables. To withdraw the over-fitting problem, C4.5 helps tree pruning (e.g., confidence-based and error-based pruning). Further, C4.5 allows attributes to be missing. C4.5 Decision tree algorithm is implemented to the provided dataset to analyze the dataset and predict the destination that is mostly visited.

C4.5 algorithm builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set {\displaystyle S={s_{1},s_{2},...}} of already classified samples. Every sample {\displaystyle s_{i}}contains of a p-dimensional vector {\displaystyle (x_{1,i},x_{2,i},...,x_{p,i})}, where the {\displaystyle x_{j}}represent attribute values or features of the sample, as well as the class in which {\displaystyle s_{i}} falls.

At each node of the tree, C4.5 chooses the attribute of the data that most efficiently divides its set of samples into subsets enriched in one class or another. The splitting pattern does the normalized information gain (difference in entropy). The attribute by the highest normalized information gain is taken to make the decision. The C4.5 algorithm later recurses upon the partitioned sub lists.

This algorithm has a few base cases.

- All the samples in the list relate to the same class. When this occurs, it simply creates a leaf node to the decision tree saying to choose that class.
- None of the features gives any information gain. In that case, C4.5 builds a decision node higher up the tree using the expected value of the class.
- Instance of the previously-unseen class found. Again, C4.5 builds a decision node higher up the tree using the expected value.

 DOI 10.17148/IJARCCE.2020.9414

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 4, April 2020

Pseudocode

In pseudo-code, the general algorithm for constructing decision trees is:
1. Check as those above base cases.
2. For every attribute, determine the normalized information gain ratio from splitting upon a.
3. Let a_best mean that attribute with the highest normalized information gain.
4. Construct a decision node that splits on a_best.
5. Recurse upon the sub lists obtained by dividing on a_best, and add those nodes as children of the node.

Improvements from ID.3 algorithm

C4.5 made some enhancements to ID3. Some of those are:

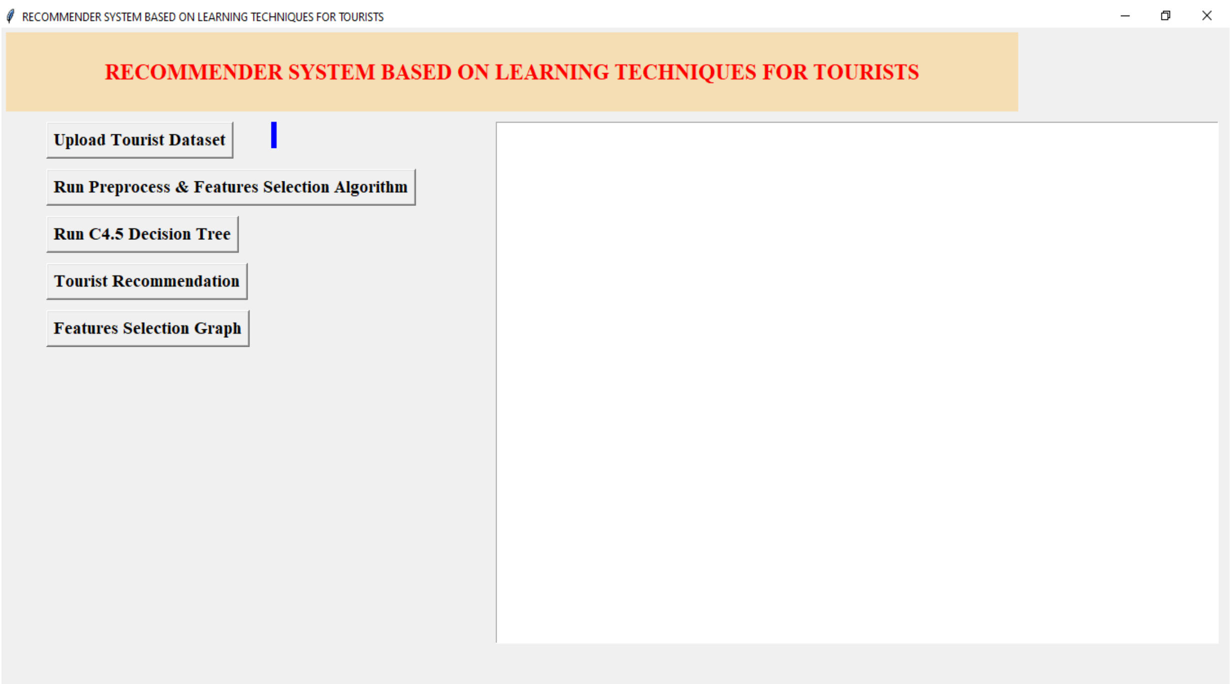
- Managing both continuous and discrete attributes - To manage continuous attributes, C4.5 generates a threshold and then divides the list into those whose attribute value is high the threshold and those that are smaller than or equal to it.
- Handling training data by missing attribute values - C4.5 allows attribute values to be marked as ? as missing. Missing attribute values exist simply not used in gain also in entropy calculations.
- Handling attributes by differing costs.
- Pruning trees after creation - C4.5 moves back through the tree once it's done creating and tries to eliminate branches that do not help by replacing them by leaf nodes.

MRMR Feature selection algorithm is also used in the proposed system. Sometimes we do not require using all the attributes and removing unnecessary data from the given dataset. This can be done by using the MRMR Feature selection algorithm. This reduces the complexity of the proposed recommendation system and increases the performance of the model.

4.1 Procedure

- Upload the dataset.
- Run preprocess and feature selection algorithm.
- Run the c4.5 decision tree algorithm.
- Predict the location.

V. RESULTS



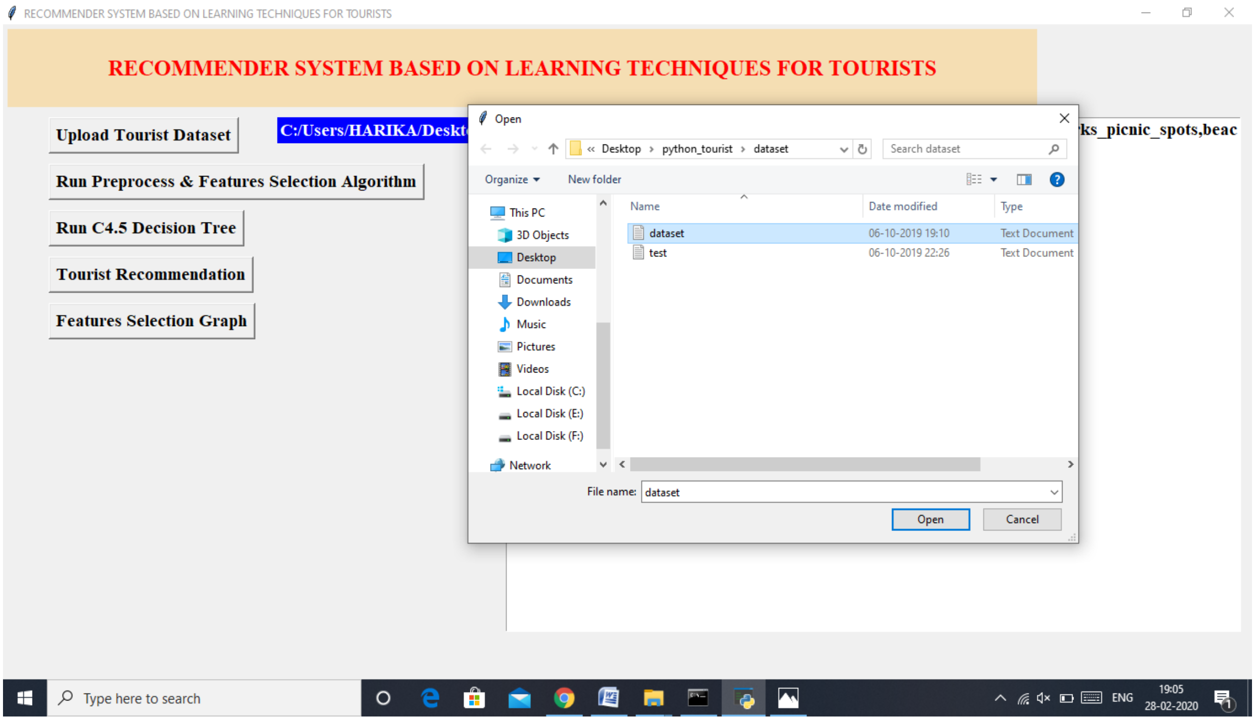
In the above screen click on ‘Upload Tourist Dataset’ button to upload the dataset file.

IJARCCE

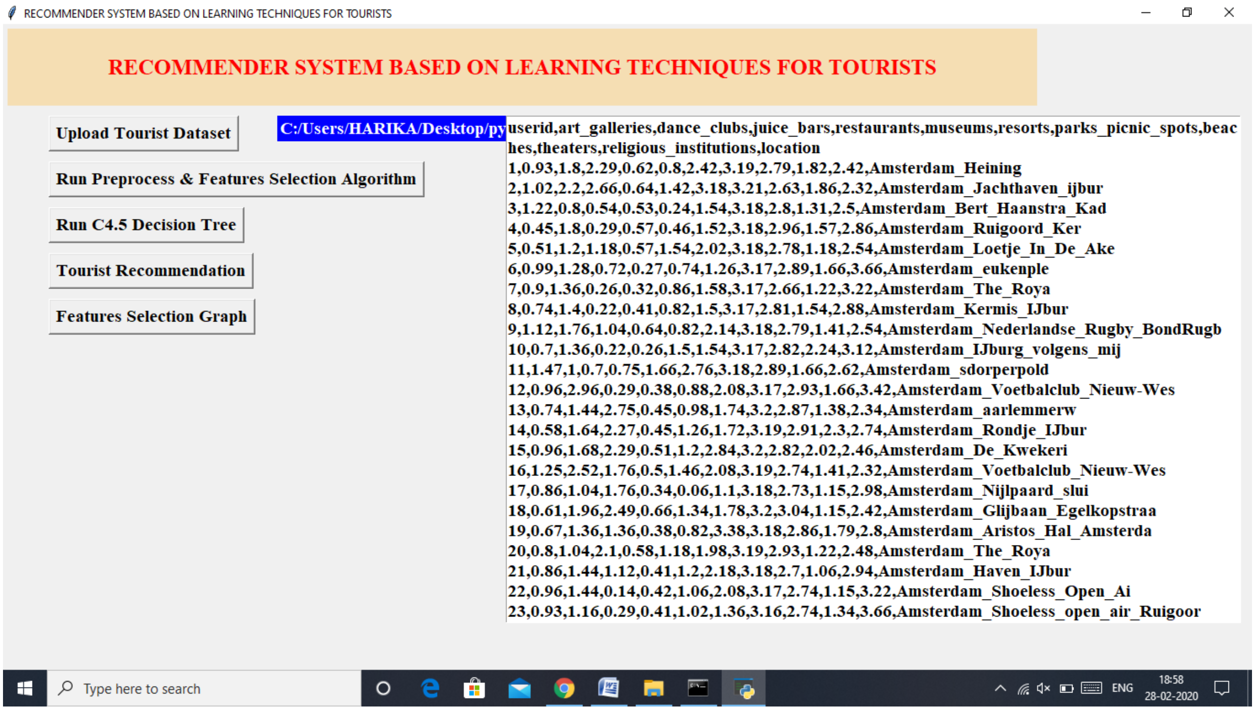
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 4, April 2020



After file upload we will get the below screen with dataset details.



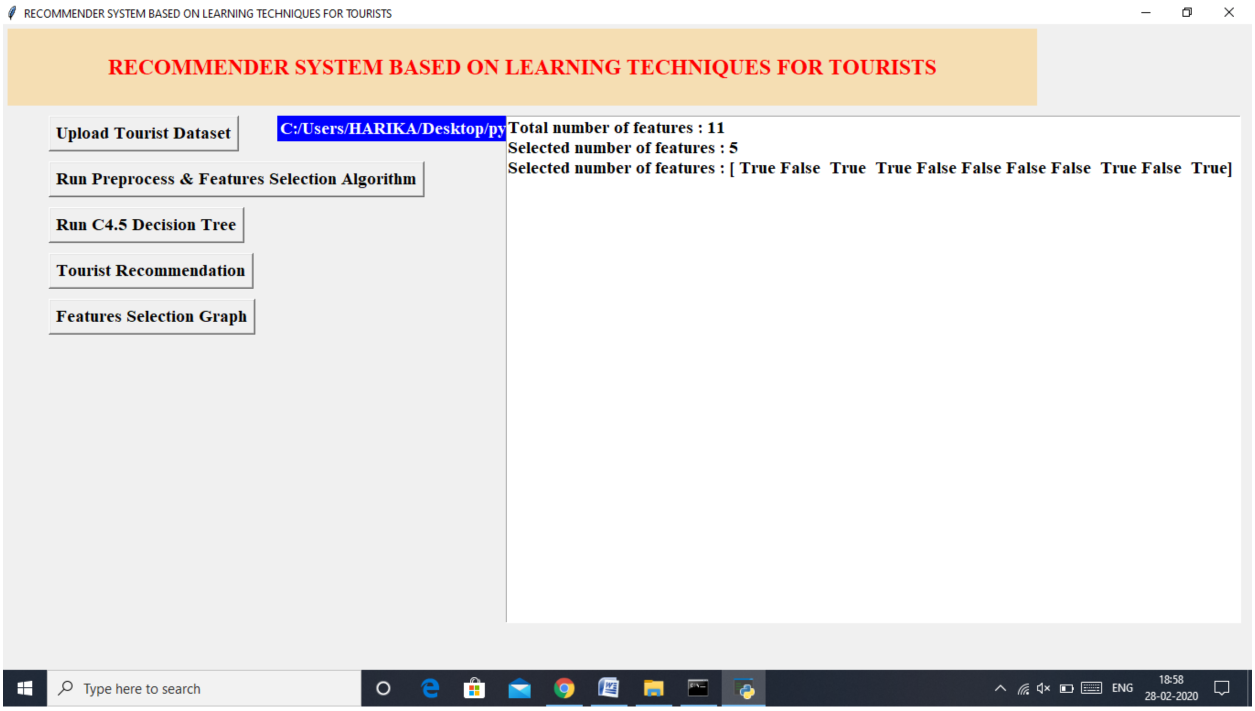
In the above screen all users past experience dataset loaded and total 11 attributes are there in the dataset. Now click on the 'Run Preprocess & Feature Selection Algorithm' button to remove empty values and reduce attributes size.

IJARCCE

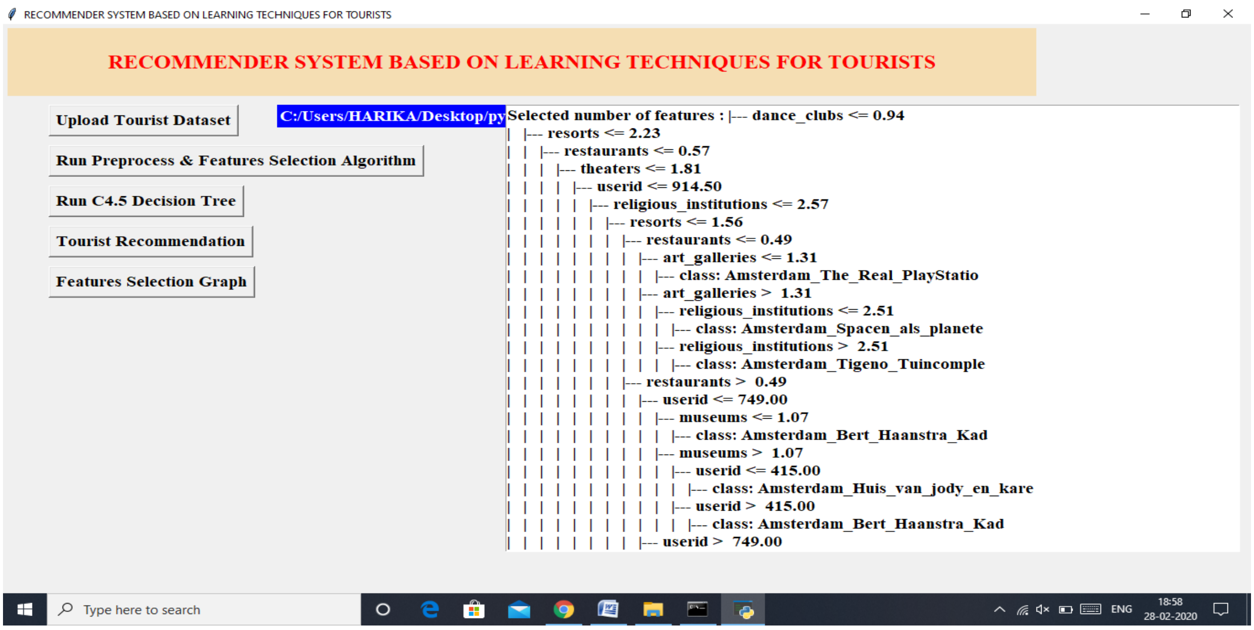
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

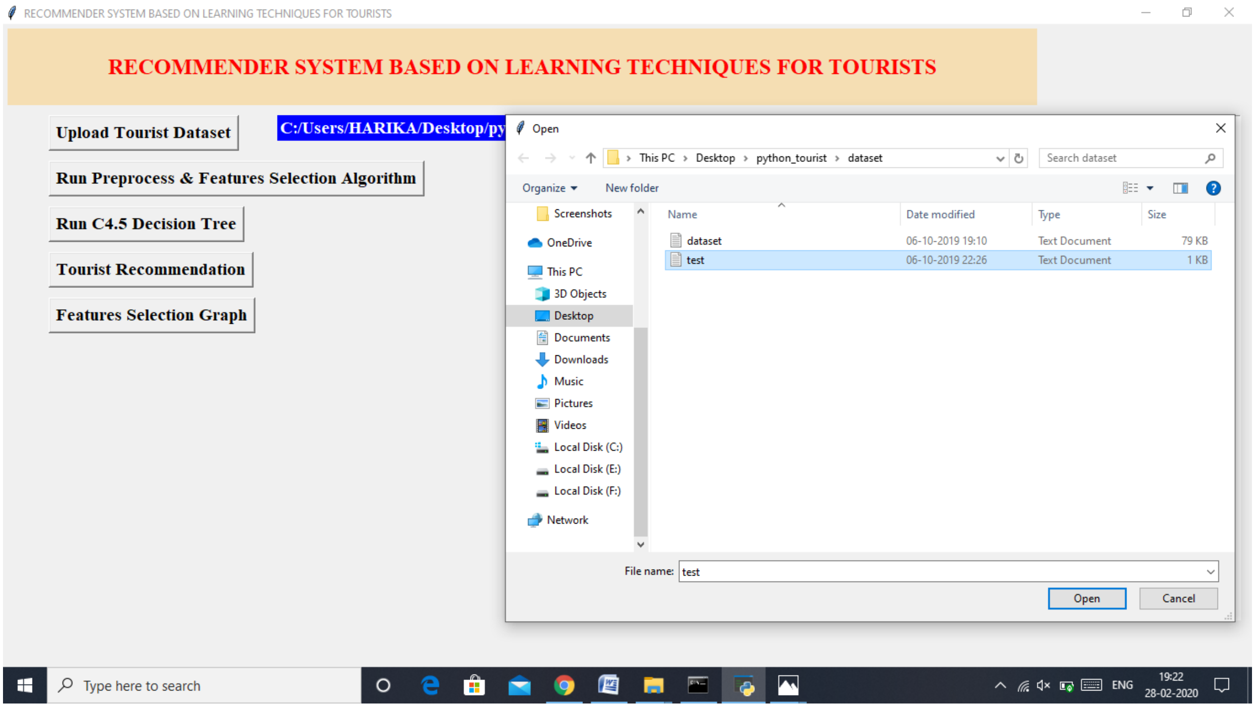
Vol. 9, Issue 4, April 2020



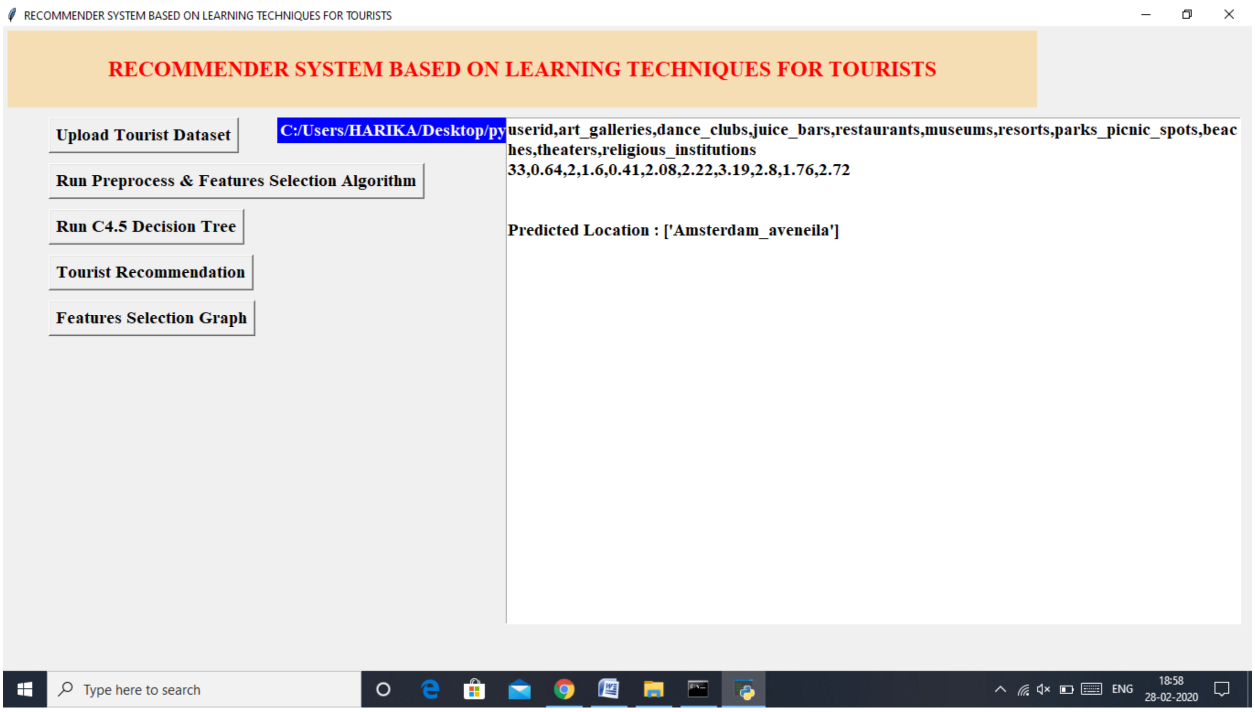
In the above screen after applying MRMR feature selection algorithm features size reduces to 5 and only those attributes will be used whose column is TRUE and FALSE column will be ignored. Now click on ‘Run C4.5 Decision Tree Model’ to build a model.



In the above screen we can see using IF and ELSE statement decision trees have generated models. If > it will choose some decision if < it will choose some other decision. Now click on the ‘Tourist Recommendation’ button to upload a test file with no location name and application will predict it.

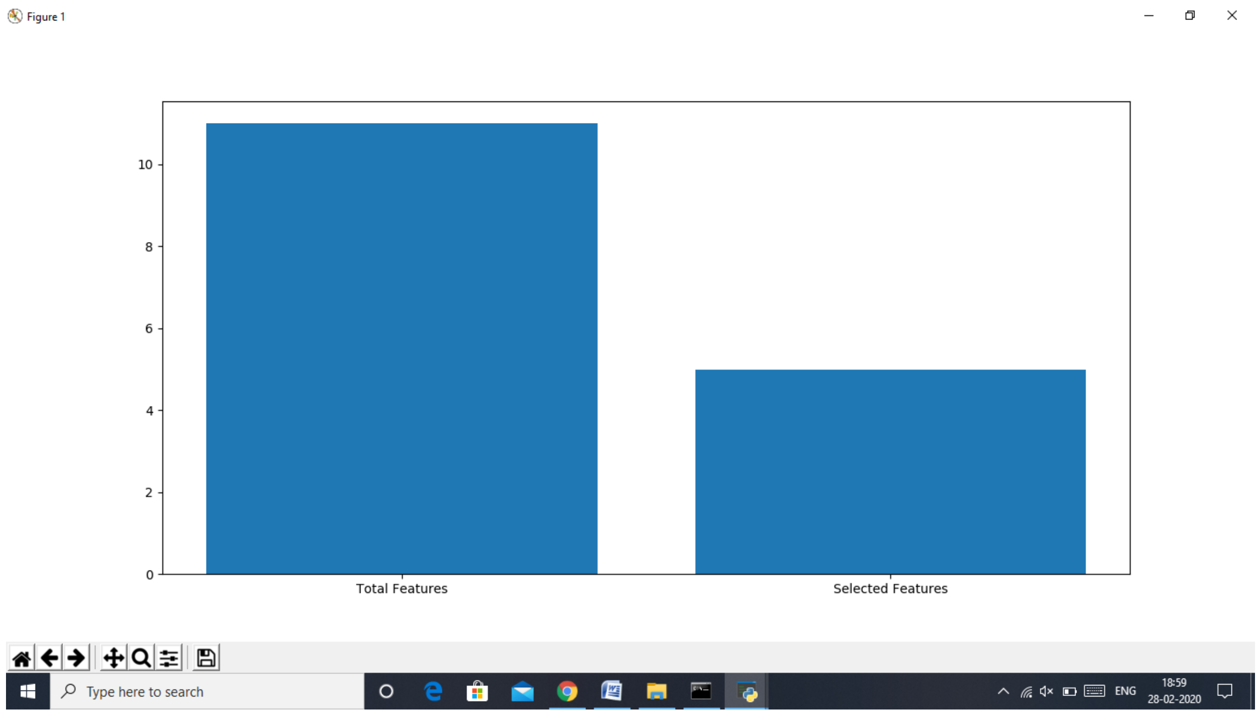


In the above screen we are uploading a 'Test file' now click open to get predicted or recommended location. In test file location name is not there application will give predicted location.



In the above screen after uploading test data we can see all values are there in test data but it does not have location name and base on test values application predicted or recommend location name.

Now click on the Features Selection Graph button to get below graph.



In above graph x-axis contains total features and MRMR selected features and y-axis represents count of features and in above graph we can see after applying MRMR technique features size reduces to 5.

VI. CONCLUSION AND FUTURE ENHANCEMENT

In this research, a decision tree based tourist recommendation system has been presented in attempt of solving the current challenge of the destination Travel Recommender System. The data set has been decomposed into two sub data sets using relevant tourism domain knowledge. This was done to increase classification accuracy rate and to reduce the complexity of the decision tree. The optimal decision trees from MRMR feature selection algorithm with the highest accuracy rate and simplicity (i.e. less number of leaf and tree size) have been constructed for destination choice. The decision rules from decision trees were extracted. It can be seen that it is the optimum method because it uses fewer numbers of features for the data sets. Finally, the experimental results confirm the applicable of the proposed a Travel Recommender System.

REFERENCES

[1].L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, “e-TOURISM:: A TOURIST RECOMMENDATION AND PLANNING APPLICATION,” Int. J. Artif. Intell. Tools, vol.18, no. 5, pp. 717–738, Oct. 2009.
[2]. E. Pitoska, “E-Tourism: The Use of Internet and Information and Communication Technologies in Tourism: The Case of Hotel Units in Peripheral Areas,” Tour. South East Eur., vol. 2,pp. 335–344, Dec. 2013.
[3]. E. Pantano and L. D. Pietro, “From e-tourism to f-tourism: emerging issues from negative tourists ' online reviews,” J. Hosp. Tour. Technol., vol. 4, no. 3, pp. 211–227, 2013.
[4]. Prof. P. A. Manjare, Miss P. V. Ninawe, Miss M. L. Dabhire, Miss R. S. Bonde, Miss D. S. Charhate, Miss M. S. Gawande “Efficient Recommendation System Using Decision Tree Classifier and Collaborative Filtering” Volume: 03 Issue: 08 | Aug-2016
[5]. R. Anacleto, L. Figueiredo, A. Almeida, and P. Novais, “Mobile application to provide personalized sightseeing tours,” J. Netw.Comput. Appl., vol. 41, pp. 56–64, May 2014.
[6].G. Häubl and V. Trifts, “Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids,” Mark. Sci., vol. 19, no. 1, p. 4, Winter 2000.
[7]. F. Ricci, L. Rokach, and B. Shapira, “Introduction to Recommender Systems Handbook,” in Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor,Eds. Springer US, 2011, pp. 1–35.
[8].P. Resnick and H. R. Varian, “Recommender Systems,”Commun ACM, vol. 40, no. 3, pp. 56–58, Mar. 1997.
[9]. G. I. Alptekin and G. Buyukozkan, “An integrated case-based reasoning and MCDM system for Web based tourism destination planning,” EXPERT Syst. Appl., vol. 38, pp. 2125–2132, 2011.
[10].J. B. Schafer, J. A. Konstan, and J. Riedl, “E-Commerce Recommendation Applications,” in Applications of Data Mining to Electronic Commerce, R. Kohavi and F. Provost, Eds.Springer US, 2001, pp. 115–153.