A Data Mining Technique for Tourist Destination Brand Image Building

Rahul Kaul, Manmohan Singh, Sweta Gupta, Geetanjli Khambra, Prashant Pathak

Abstract: *The destination image branding is the domain of tourism industry where the facts and information is collected and evaluated for finding the credibility of a target tourist destination. Manual collection and processing of collected information accurately is a complicated and time consuming task therefore a data mining model is suggested ,in this presented work that collect and evaluate the destination image accurately and based on evaluation can make the recommendations about visits of tourist. In order to perform this task data mining techniques are applied on text data source. In first the data is extracted from the Google search engine and it is preprocessed for make it impure. In further the data is labeled based on the positive and negative words available in the collected facts. Finally the clustering and classification of text is performed. For clustering of data FCM (fuzzy c means) clustering algorithm and for classification the Bayesian classifier is used. Based on final classification of text data the decision is made for the destination visits.*

Keywords: Brand building, destination image building, data mining techniques, clustering.

I. INTRODUCTION

Data mining and their techniques provide us the ability to analyze the data automatically using the computational algorithms. Additionally grab the outcomes of analysis for decision making, classification, predictions or other essential task. In this presented work the data mining is performed on the unstructured data i.e. text documents. Therefore the proposed work is intended to demonstrate the technique of text mining.

The data mining is conducted in this presented research on u nstructured data, i.e. text documents. Hence the proposed w ork is intended to demonstrate the text mining technique. Te xt mining is the data mining subdomain which deals with th e text data. Using the text mining approaches in this work the destination branding of the tourist places is obtained using text mining techniques. Basically when someone plans to visit some place as tourist he/she not know all the prospects of the particular place.

Therefore sometimes the visitor is trapped in various kinds of issues such as misleading place, inappropriate visiting conditions, risk of thief and others.

Therefore the visitors collect the information from web to know basics of the particular place. But in most of online resources only the common or basic overview about the places are available. That information is not complete in terms of to make decision to visit the place strongly recommended or not.

Therefore in this presented work using the different source of data analysis a new model is proposed that investigates about the places to visit. Additionally by analysis of the data it produces the strong recommendations to the users to visit the place or not. In this context different source of data such as news, blog and other source of data is investigated and analyzed using the data mining algorithms. The analysis of the data results the patterns of data and using the recovered patterns the suggestions are made to visit the place or not.

II. LITERATURE REVIEW

A. Data Mining: Data mining is a data based analysis approach[1], we have data as a collection, which is no doubt a historic in nature[2]. To understand the concept we can say that it is a process which have several filters of condition y which we get an approximate information. Now this process can be consist of several technique like artificial intelligence and statics formulas to identify the real fact. Data mining is also terms as "knowledge discovery" The data mining system is responsible to establish relationship among the data[3], also which describes the steps that must be taken to ensure meaningful results. "Fig.1" illustrates main components of our prototype for knowledge discovery in databases.

Mr. Rahul Kaul*, Assistant Professor, Department of Computer Science & Engineering, CDGI, Indore, India.
Dr.Manmohan Singh, Professor, Department of Computer Science and Engineering, CDGI, Indore, India.
Mrs Sweta Gupta, Assistant Professor, Department of School of Engineering &Technology, Jagran Lakecity University, Bhopal, India.
Ms. Geetanjli khambra Raghuwanshi , Assistant Professor, Department of Computer Science & Engineering, JLU, Bhopal, India.
Mr. Prashant Pathak , Assistant Professor, Department of Computer Science, Jagran Lakecity University, Bhopal, India.
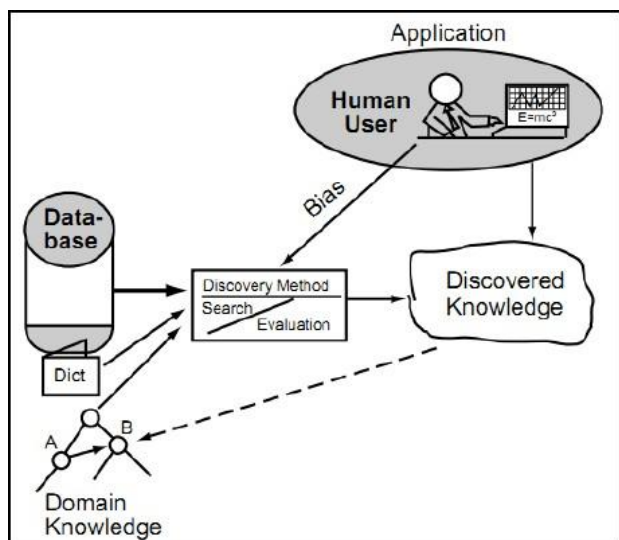
Figure 1. A Framework for Knowledge Discovery in Database[4]

We input raw question and selects sample data and apply algorithm, process data using AI algo and identify finally our output. The basic and central point of the system is the discovery of information, which computers can computes. Take an instance-

We need to identify a tourist place on basis of its popularity, so how should we initiate? so

1. We need to gather information about no of tourist visits.
2. We need to identify their source, from where they come
3. Their visit timing, with respect to months, and arrival at tourist place.
4. Their booking type like either by tourism agent or by selecting individually. If by tourist then portfolio of tourist like dig any other place agent is involved?
5. If they are selecting their own hen how they are planning? Like by internet or by any suggestion from previous visitor.
6. Their spending habits i.e. cash or credit card, Debit card etc.

After collecting all the data we need to question in this collection to get answer like, "How many visitors in april month visit through agent every year" and evaluates patterns. Now here we must apply some technique of data mining like "Association Rule, Clustering, Decision Trees[5].

If data identify that bookings of hotels are going by online platform then, we need to identify which is the popular platform for online booking. This case might differ by city to city and country to country. Online booking of hotels and travelling is grown by 90% world wide, especially in Europe and US. South asia visitors also seeking for this methods. Now how we got this result of identifying booking behavior of tourist? To identify this, it is best showed in Emel[6]. This papers is also suggest to identify tourist profile. You can not sent a tourist of newly married (based on European native) to any religious place. You need to select the romantic package fr them, which may help them to understand. On the contrary, you can not replicate the same pattern with newly Indian couple. For branding of tourist place several peoples play vital role, but marketing company can do their best using data mining. For instance - Cox's Bazaar is one of the destination to be included by all the tours manager but due to unknown or less know we could say that due to less brand building it is still not in the list of tour manager[18], the better they can market their destination[17]. By using several data mining technique we can identify different types of tourist profiling behavior[16].Wong, Chen, Chung, and Kao (2006)[7] adopted data mining techniques to analyze the tourist plan patterns of travelers and his conclusion was that DMOs in Asian countries should promote their destinations which are new and untouched. In the tourism industry, To take marketing decision and strategy, tourist company can take help of data mining technique for their collected data of past tourists[8]. Now in current scenario, several peoples done bookings in mobile using app or web application, we can co-relate the profile of traveler with his search keywords for deciding tour patterns by tour manager[15].

For profiling of tourist first we understand the hospitality business. What is hospitality business and what type of product are on sale. Some examples are as follows-

1. Reservation (Rail, Bus, Flight Local Conveyance).
2. Hotel
3. Meal arrangements
4. Local market and their product
5. Local guide, and interpreter (In case)

Now considering all these variables we need to establish of one person over all these variables[9]. Now we can do some meal price discount for repeated customer in hotel, but in case of travel we can discount them in hotel reservation or providing them extra site to visit in the same packege.

B. Consumer Behavior: What is a consumer behavior? To know this first understand the consumer, who is he? He is a person having capacity to spent in lieu of any services or product and service to deliver that product.

Some research shows that due to globalization there are a lot of changes happened in tourist behavior. Like a tourist from America can purchase shawl in Jammu and book it to his country through cargo, now seller of shawl can also do the business of cargo[10]. Some research shows that tourist are also going to the remote place where nobody went in decades and these types of fashions is now in trends[11],[12] like visit to Udakmandalam in Tamilnadu, India is increased in late 90's. Some time past travelers suggest to future travelers for exploring new places[9], like visits to COX bazaar, Bangladesh is increased[18].

Research shows that new generation travelers are interested in

1. Planning at their own
2. Visit a place which is un-predictable
3. They don't want to take a plan route

So the new generation travelers are more complicated to be handled. New generation tourists does not believe in waiting queue for a well known destination, rather then they will visit to uncover area[13], like they might not visit Tajmahal, due to heavy rush or crowd, but they will prefer to go Himanchal in winter season when a little bit tourist visits.

They also reluctant to make their plan in vacation only, now the plan according to less crowd time

In literature a number of applications and examples are exist where the data mining approaches are employed for reviewing the text for obtaining the review about product and services. In this presented work the sentiment analysis is applied on NEWS data and blog data which are collected using the Google search results to develop strong recommendation about the place to visit. This process is termed here as destination image branding. Basically when we search about any tourist destination from web the available source of information provides us the overview of the particular destination.

Additionally data is not updated frequently or regularly in available sources such as blog or other informative sites. But the place and their conditions are influenced by the different political event, natural events and other human made disasters (i.e. terrorism) and others. Therefore the image of particular place is also changing over time. In this context for obtaining the clear and strong review about the tourist destination the fresh data is required to process or analyze. In this presented work the Google search technique is implemented for finding fresh information from web data source.

Additionally by using this fresh data the analysis is made to recommend the destination image. In addition of that a hybrid data model is proposed for performing the data analysis.

The proposed technique is a combination of two data mining algorithms FCM (fuzzy c means) clustering and the Bayesian classifier. Using both the technique the data is processed and the essential features are recovered for making recommendations about the tourist destinations. In this section the need of proposed system is described.

III. PROPOSED WORK

The sentiment analysis and supervised learning techniques help to investigate the user's mood and the reviews about the different products. In this presented work the review analysis of visiting places are performed using the data mining techniques for developing effective branding of any tourist place.

A. Proposed Methodology

The proposed methodology of the system design to find recommendations about the tourist places are described in figure 2. In addition of the components of the given model is explained in same section.

User Destination: the data mining and machine learning system requires some initial inputs to process data and recovers the essential information. The proposed system also involve the concept of data mining techniques therefore to find the fresh data about the places a provision is developed for accepting the initial input. Using this input provision user can provide the place name (destination) for visit.

Google Search: the user input place name is produced to the Google search API for finding the information from web. Google search collect the information from internet source and generate a list of results.

Search results: the generated results from Google search is collected in a file for utilizing with the further process.

Data preprocessing: The preprocessing of data is an essential step in the exploration of data mining and information. It is performed for cleaning the target data and makes it suitable for use. Therefore using the preprocessing here we remove the unwanted data from the generated results through the Google search. In this context the stop words and special characters from the text data is removed. To remove the stop words and special characters from collected data a list of words and characters are prepared. Additionally a function is developed that remove all the data from text which is available in the prepared stop word list and special character list.
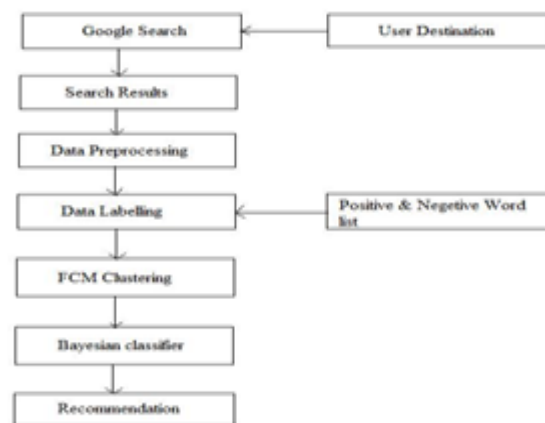

Figure 2 Proposed System Architecture

Positive & negative word list: a list of positive and negative words that are used in sentiment analysis or orientation mining is taken from web. This list of word contains both kinds of words negative as well as positive labels.

Data labeling: using the positive and negative word list initial labels of the collected data is prepared. In order to perform the labeling of data all sentences are evaluated using this word list and if the sentence contains multiple positive words then the sentence is labeled as positive otherwise the sentence is labeled as negative sentence.

FCM clustering: after successfully labeling of data t is used with the FCM (fuzzy c means) clustering algorithm. That clusters all the sentences more precisely. The FCM is an unsupervised learning technique that evaluates data using the fuzzy optimization function to make more clear clusters from the data. The process of data clustering is given as follows:

The fuzzy c means clustering works on the basis of the

$$O_n = \sum_{i=1}^{n} \sum_{j=1}^{k} P_{ij}^n \left\| d_i - k_j \right\|^2$$

following objective function:

Where n is a real number and must be grater then

1 P_{ij}^n is the degree of membership

d_i is i^{th} element of data object or instance of data (i.e sentences)

k_j is cluster centroid

For computing the partitions the iterative optimization process is called. Therefore it is required to compute the membership of data and new cluster centers.

$$P_{ij} = \frac{1}{\sum_{l=1}^{k}\left(\frac{\|d_i - k_j\|}{\|d_i - k_l\|}\right)^{\frac{2}{n-1}}}$$

εThe algorithm is terminated when the following condition reached.

After clustering the data is subdivided into two groups one group contains the positive sentences and other group contains the negative sentences.

Bayesian classifier: the categorized data using the FCM (fuzzy c means) clustering is used with the Bayesian classifier. The classifier basically cross validates the clustered data instances by again classification of all the data. That improves the quality of prediction which is performed after carry out the classification of Bayesian classifier. The Bayesian classifier works in the following manner.

IV. RESULT ANALYSIS

The given section includes the performance analysis of the implemented algorithms for the proposed web content mining approach of destination branding and image. Therefore some essential performance parameters are obtained and listed with their obtained observations.

A. Positive Probability

Positive probability is the probability of the words which are belonging to outcome of the positive reviews. Positive probability can be calculated by using following formula:

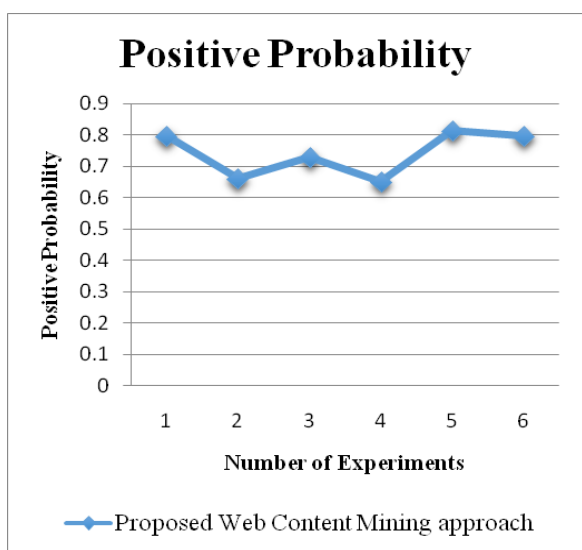$$\text{Positive Probability} = \frac{\text{Number of Positive Words}}{\text{Total Number of Words}}$$

Figure 3 Positive Probability

The positive probability of the proposed algorithm of destination branding image is represented using table 3 and

figure 3. The given graph figure 3 contains the positive probability of the implemented algorithms. The X axis of the diagram shows the different experiments and Y axis contains the obtained performance. To demonstrate the performance of the proposed technique is representing using blue line. According to the obtained results the performance of the proposed model provides more accurate results. Additionally the positive probability of the proposed model is varying as the number of experiments increase.

$$k_j = \frac{\sum_{i=1}^{N} P_{ij}^{n} * d_i}{\sum_{i=1}^{N} P_{ij}^{n}}$$
$$\left\{ |P_{ij}^{(k+1)} - P_{ij}^{k}| \right\} < \varepsilon$$

Table.1 Positive Probability

Number of Experiments	Proposed Web Content Mining Approach
1	0.7954
2	0.6595
3	0.7272
4	0.65
5	0.8125
6	0.7955

B. Negative Probability

The negative probability which is shows the negative reviews of the outcome of the classification process. The negative probability can be estimated by using following formula.

$$\text{Positive Probability} = \frac{\text{Number of Negative Words}}{\text{Total Number of Words}}$$
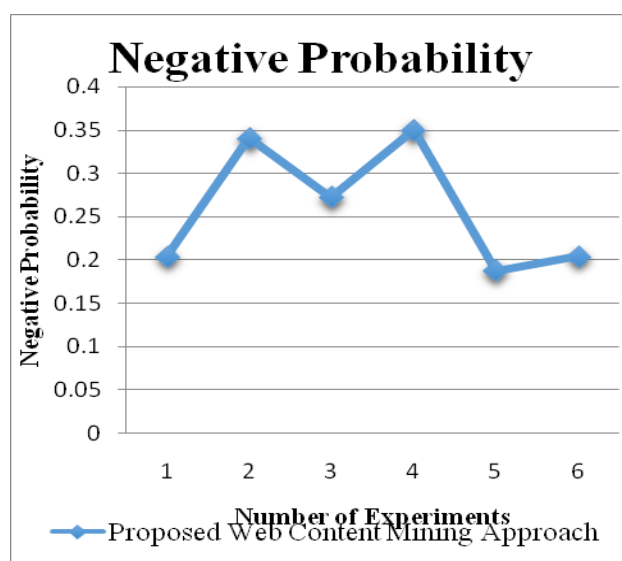

Figure 4 Negative Probability

The figure 4 and table 2 shows the error rate of implemented proposed approach. In order to show the performance of the system the X axis contains the experiments and the Y axis shows the performance in terms of negative probability. The performance of the proposed web content mining technique is given using the blue line. The performance of the proposed word classification is effective and efficient during different execution and reducing with the amount of data increases.

Thus the presented classifier is more efficient and accurate than the other text classifier of classification

Table 2 Negative Probability

Number of Experiments	Proposed Web Content Mining Approach
1	0.2045
2	0.3404
3	0.2727
4	0.35
5	0.1875
6	0.2045

C. Usage Memory (Memory Usage)

Any algorithmic performance is based on memory consumption of the system which determines the space complexity and the consumed memory can be calculated by the following formula:

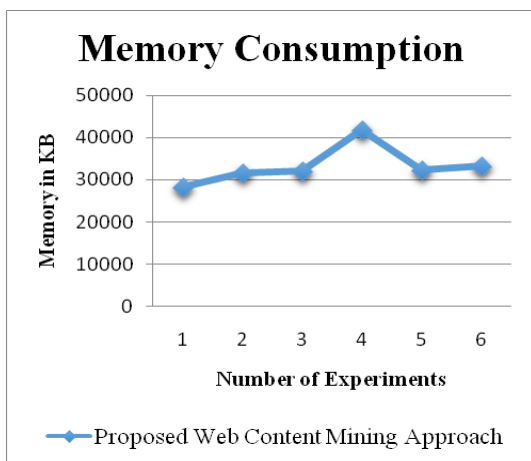$$\text{Memory Consumption} = \text{Total Memory} - \text{Free Memory}$$

Figure 5 Consumed Memory

Algorithmic statics in relate to memory consumption depends how much of volume data has been resides in RAM. The performance of the implemented web content mining approach of destination branding is given using figure 5 and data is numerically show by table 5. For clarification of the result, X-axis contains the different amount of program/ software execution and the Y-axis shows the memory consumed during execution in terms of KB. The result we got shows that algorithm execution

behavior is similar with increasing of volume of data and on the contrary memory utilization is less.   This consumed memory represents the required space by which algorithm of word classification of dictionary is executed and produces efficient output.

V. CONCLUSION

The intension behind the technique proposed is to design and develop a data mining based technique that helps to suggest the credibility of the tourist destination. Therefore a model is developed using the data mining algorithm. This chapter provides the summary of conducted efforts for developing the required data model.

Destination branding is a subject of tourism industry. In this subject the different source of information is collected for finding the facts about the particular tourist destination. Additionally a brand image is developed on the basis of collected facts about the place goodness. In this context various manual efforts are observed in literature. But the data analysis in manual mode is complicated and time consuming task. Therefore the proposed work include a data mining technique that collect and evaluate the target tourist place for computing the recommendations about the tourist destination brand image. The proposed technique is a data mining model that first employs the Google search engine API for collecting fresh information from web. In next using the positive and negative word list the labeling of data is performed. In final phases two different data mining algorithms are applied for mining and exploring the information and facts. The data mining methods includes the FCM and Bayesian classifier. First using the FCM (fuzzy c means) clustering the data is categorized in two clusters and finally it is classified for finding accurate information about the target place.

The implementation of the proposed tourist destination image branding technique is performed using the JAVA technology. After implementation the performance of the model is also computed. Based on the experimental results the following outcomes are observed as demonstrated in .

Table 3

S. No.	Parameters	Remark
1	Positive probability	The positive probability is fluctuating between 0.65-0.81 which is highly acceptable for recommending the visitors place
2	Negative probability	Negative probability is varies between 0.34-0.09 therefore it is acceptable for recommendations

3	Memory usages	The main memory requirements are depends on the data to be process and it is observed between 28K-41K KB

VI. FUTURE WORK

The aim of this work is to extend a data mining model which can be recommended for the tourist destination brand acceptability according to the fresh data mining. The implementation and their performance evaluation are performed successfully. Probably the proposed model can be extended for the following domains for future prospects. Current system only extract the data that is appeared on the search results the relevant inner links are not downloaded. Therefore in near future the inner link data extraction is also involved in this technique

1. The proposed system implements the classifier in normal mode in near future the ensemble learning concept is used for improving the classification performance of the system.

REFERENCES

1. MICHAEL, J.A., AND LINOFF, G. (2011). "DATA MINING TECHNIQUES FOR MARKETING, SALES, AND CUSTOMER RELATIONSHIP MANAGEMENT". WILEY PUBLISHING.
2. CABENA, P., HADJINIAN, STADLER, R., VERHEES, AND ZANASI, A. (1997). "DISCOVERING DATA MINING: FROM CONCEPT TO IMPLEMENTATION". PRENTICE HALL.
3. EDELSTEIN H. "DATA MINING: EXPLOITING THE HIDDEN TRENDS IN YOUR DATA". DB2 ONLINE MAGAZINE. HTTP://WWW.DB2MAG.COM/9701EDEL.HTM.
4. FRAWLEY, W.J., PIATETSKY-SHAPIRO, G., AND MATHEUS, C.K. (1992). "KNOWLEDGE DISCOVERY IN DATABASE: AN OVERVIEW". AI MAGAZINE: 57-70.
5. HOONTRAKUL, P. AND SAHADEV, S. (2005). "APPLICATION OF DATA MINING TECHNIQUES IN ON-LINE TRAVEL INDUSTRY: A CASE STUDY FROM THAILAND". WEBSITE. WWW.PONGSAK.HOONTRAKUL.COM
6. EMEL, G. G., TA?KIN, Ç., & AKAT, Ö. (2007). "PROFILING A DOMESTIC TOURISM MARKET BY MEANS OF ASSOCIATION RULE MINING". ANATOLIA, 18(2), 335-343.
7. WONG, J., CHEN, H., CHUNG, P., & KAO, N. (2006). "IDENTIFYING VALUABLE TRAVELERS AND THEIR NEXT FOREIGN DESTINATION BY THE APPLICATION OF DATA
8. MINING TECHNIQUES". ASIA PACIFIC JOURNAL OF TOURISM RESEARCH, 11(4), 355-373.
9. DANUBIANU, M. AND HAPENCIUC, V. (2008). "IMPROVING CUSTOMER RELATIONSHIP MANAGEMENT IN HOTEL INDUSTRY BY DATA MINING TECHNIQUES", PROCEEDING OF "COMPETITIVENESS AND STABILITY IN THE KNOWLEDGE-BASED ECONOMY", VOL: CD, 30-31 MAI, 2008, CRAIOVA, ROMANIA, ISSN/ISBN: 978-606-510-162-3, PAGE: 2444-2452
10. BUHALIS, D. (2002). "ETOURISM: INFORMATION TECHNOLOGIES FOR STRATEGIC TOURISM MANAGEMENT". FINANCIAL TIMES PRENTICE HALL.
11. KOTLER, P., BOWEN, J. AND MAKENS, J. (1999). "MARKETING FOR HOSPITALITY AND TOURISM", 2ND ED., PRENTICE HALL, USA
12. SMITH, VALENE., (1977). "HOSTS AND GUESTS: THE ANTHROPOLOGY OF TOURISM", UNIVERSITY OF PENNSYLVANIA PRESS, PHILADELPHIA.
13. KING, BRIAN AND HYDE, G., (1989). "TOURISM MARKETING IN AUSTRALIA", HOSPITALITY PRESS, MELBOURNE.
14. POON, A. (1993). "TOURISM, TECHNOLOGY AND COMPETITIVE STRATEGIES". NEW YORK: CABI PUBLISHING.
15. PAIRAYA JUWATTANASAMRAN 2, SARAWUT SUPATTRANUWONG , 3, SUKREE SINTHUPINYO, 2013 "APPLYING DATA MINING TO ANALYZE TRAVEL PATTERN IN SEARCHING TRAVEL DESTINATION CHOICES".
16. GIRISH KUMAR SHARMA, PROMILA SHARMA, 2016, "A STUDY ON DATA MINING ALGORITHMS FOR TOURISM INDUSTRY"
17. PARIKSHAT SINGH MANHAS, DEEPAK RAJ GUPTA, JEET DOGRA,2012," DESTINATION BRAND BUILDING, PROMOTION & BRANDING: IMPACT ANALYSIS OF BRAND BUILDING ELEMENTS"
18. BIPLAB ROY, RAJIDUL HOQUE,2015, BUILDING A STRONG BRAND IMAGE OF COX'S BAZAR AS A TOURIST DESTINATION: AN EMPIRICAL ANALYSIS ON COX'S BAZAR

AUTHORS PROFILE

Mr. Rahul Kaul, is an Assistant Professor in the Department of Computer Science & Engineering, CDGI Indore. He has a Master Degree in Computer Science Engineering. His area of interest is Data Mining. He has written original research articles in various international journals and he is interested in academics and research.

Dr. Manmohan Singh, is a Professor in Department of Computer Science and Engineering at CDGI Indore M.P. He has 11 years of experience in teaching and research. He has published more than 25 research paper in international journals and conferences. He has authored many book published by national and international publisher.

Mrs Sweta Gupta, is an Assistant Professor in School of Engineering &Technology, Jagran Lakecity University, Bhopal. She is an academician having 9+ yrs of teaching experience. She is a PhD scholar in computer science engineering department UIT RGPV Bhopal.
By profession she has been a trainer for Network & Cyber Security and Information Security. Her area of research includes blockchain.

Ms. Geetanjli khambra Raghuwanshi , is working as an Assistant Professor Having 6 years of teaching experience in the field of computer science & engineering. Pursuing Ph.D in CSE from JLU, Bhopal. She has written original research articles in various international journals and she is interested in academics and research..

Mr. Prashant Pathak , is working as an Assistant Professor Having 12+ years of teaching experience in the field of computer science. He also holds two copyright in computer science area. He has written original research articles in various international journals and is interested in research related to AI and human behavior.