INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Index Copernicus Publishers Panel, Poland,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)]

as well as in Cabell's Directories of Publishing Opportunities, U.S.A.

Circulated all over the world & Google has verified that scholars of more than Hundred & Twenty One countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

www.ijrcm.org.in

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email addresses: infoijrcm@gmail.com or info@ijrcm.org.in.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF** _____.

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, centered, separately numbered & self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:** These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITE

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

A FRAMEWORK FOR MINING BUSINESS INTELLIGENCE – A BOON TO NON MINING EXPERTS

B. KALPANA
RESEARCH SCHOLAR, BHARATHIAR UNIVERSITY, COIMBATORE
ASST. PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS
PSG COLLEGE OF ARTS AND SCIENCE
COIMBATORE

DR. V. SARAVANAN
PROFESSOR & DIRECTOR
DEPARTMENT OF MCA
DR. NGP INSTITUTE OF TECHNOLOGY
COIMBATORE

DR. K. VIVEKANANDHAN
PROFESSOR
SCHOOL OF MANAGEMENT
BHARATHIAR UNIVERSITY
COIMBATORE

ABSTRACT

Any business today requires additional intelligence for its own decision making process. The data available should act as leverage to business intelligence that could be used for decision making purpose. This paper outlines a framework for an automated data mining tool for business operators. The required knowledge to use a data mining tool by the business user is far out of his/her domain knowledge. Activities like selecting required attributes for the query, selecting appropriate mining technique for business intelligence can be done by software agents. The proposed framework for automated data mining proposes to solve the problem of business operators by use of intelligent agents. This paper shows the preliminary work done on an online ticket booking website database.

KEYWORDS

Automated data mining, agents, business intelligence.

INTRODUCTION

Any software bought today requires a certain degree of customization. Either we have to setup a wizard or we have a company representative to install it in our system or office. After some basic configuration is done, the system runs as required. Consider a business operator, who requires some business intelligence for marketing, for enhancing customer relationship or for cross selling purposes. The business operator has a repository of data like who bought what, customer type, customer searches, click patterns, queries, product sold dates etc. The business operator uses a business tool for analysing the required trends or patterns. The settings for the tool can be semi automated by a wizard. The business operator has no knowledge of how the settings can be changed or whether the settings are set properly initially. The proposed framework helps to solve these problems by making the tool more user friendly and with the aid of agents, make the mining process more automated. The following are major advantages a building an business intelligence tool. [1]

- follow profitability of their products sold;
- analyse expenditures;
- monitor corporate environments; and
- discover business anomalies and frauds
- discover new trends and novel information based on historic data

BUSINESS OPERATOR AND THE MINING TOOL

Consider the scenario where an online business uses mining tool for business analysis. The data to be mined is collected and made ready for mining. After pre-processing the data is made ready for mining. The business operator will not have a clue of what to do with the accumulated data or how to process it so as to get a optimized result for his query. The job of the tool is to present different results based on the configuration given by him initially.

FIG. a. BUSINESS OPERATORS IN THE WEB AND A DM TOOL

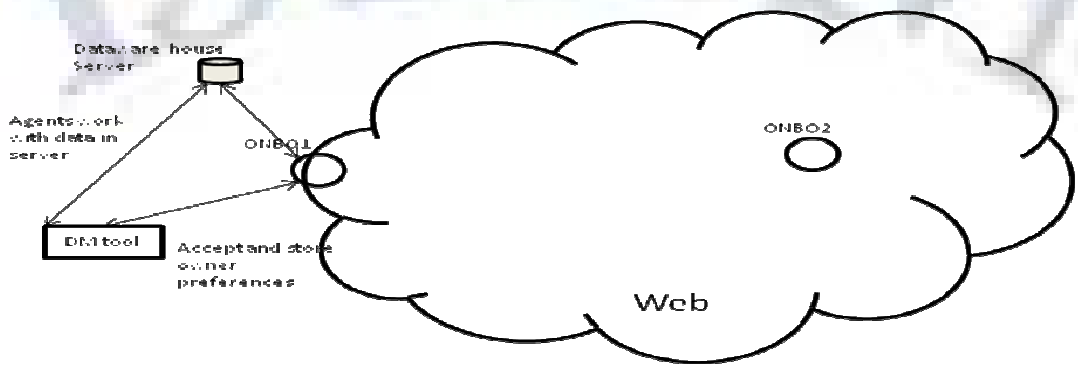
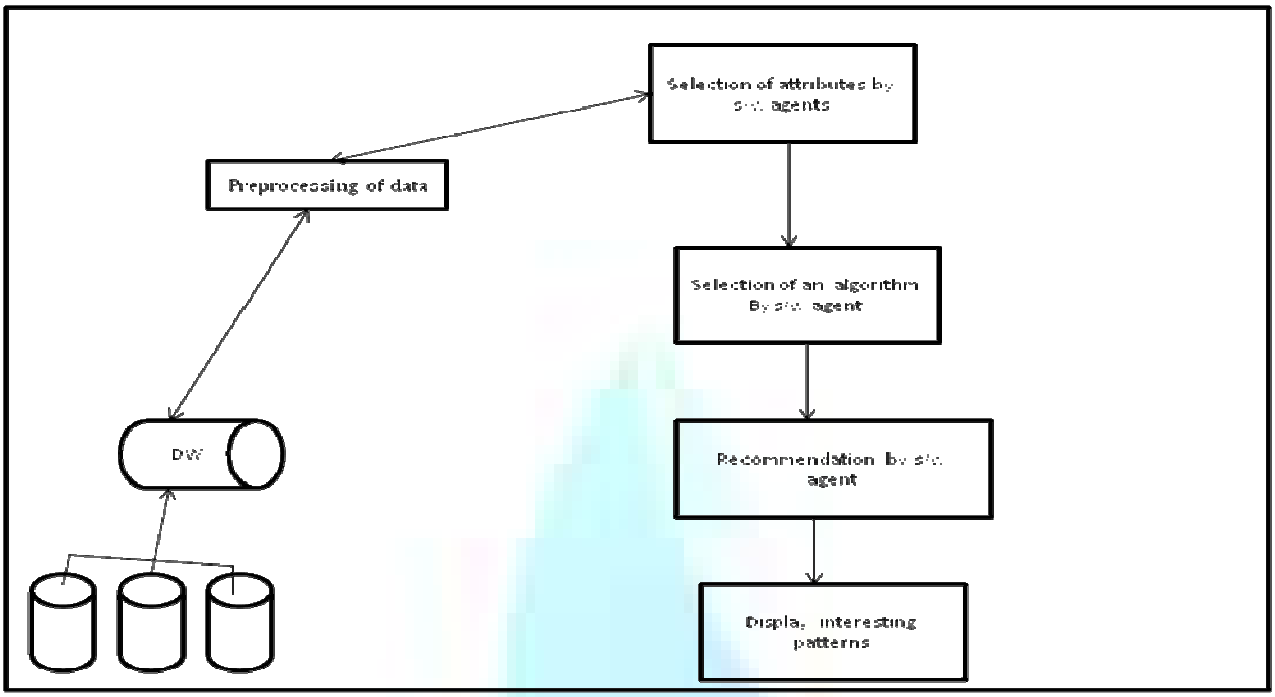


FIG b: THE FRAMEWORK OF AN AUTOMATED TOOL



The DM tool keeps collecting the business operators preferences based on their clicks. In fact the framework suggests that, agents can show the structure of the query to be processed in a more general term than calling it association, classification or clustering. The tool starts storing the preferences of the business operators of what type of analysis he/she is interested in, for future use. The proposed framework is shown in Fig b.The theoretical model is explained in following sections.

BUILDING A WAREHOUSE

A ware house is built based on the transaction database. A copy of transactions can be periodically updated in a data warehouse. Loading a data ware house with information for mining is an iterative step. According to W. Inmon [2] , the father of data warehouse, first the data from different sources are extracted and loaded into a data ware house. How the subset of data should be loaded for meaningful mining?. Should the data be transformed based on unit of time, or location or product type or customer class or activity type?

FIG c. DATAWARE HOUSE

There are practically speaking, an infinite number of ways to subset data. What is a meaningful subset depends entirely on who the first user of the first iteration of the data warehouse will be. The data architect needs to ask the question - who will be the first user of the first iteration of data and knowing whom the first user is, what data would be meaningful to them? A data model is also selected so as to make mining less cumbersome.

PREPROCESSING OF DATA

Data cleaning and transformation of instances to suit the mining technique is done in the pre-processing stage.

Data cleaning deals with missing values, noise and redundant data. Missing values could be due to data not available or accessible due to privacy issues. If attribute value is missing we can infer the value of attribute based on other attributes or other instances or just ignore the instance .If class label is missing we can omit that instance or use a semi supervised learning and use estimation maximization technique to find if the classifier classifies accurately. Noise in data could be due to unknown encoding, inconsistent formats, out of range values etc. Noise can be removed using consistency checks, canonicalization, prior domain knowledge or statistical measures. We can handle noisy data by binning, regression, clustering to find outliers.

THE ROLE OF AGENTS IN AUTOMATED DATA MINING FRAMEWORK

Agents are defined as software or hardware entities that perform some set of tasks on behalf of users with some degree of autonomy [4]. In order to work for somebody as an assistant, an agent has to include a certain amount of *intelligence*, which is the ability to choose among various courses of action, plan, communicate, adapt to changes in the environment, and learn from experience. Other attributes that are important for agent paradigm include mobility and learning. An agent is *mobile* if it can navigate through a network and perform tasks on remote machines. A *learning* agent adapts to the requirements of its user and automatically changes its behaviour in the face of environmental changes [5].

For learning or intelligent agents, an *event-condition-action* paradigm can be defined [6].In the context of intelligent agents, an *event* is defined as anything that happens to change the environment or anything of which the agent should be aware. For example, an event could be the arrival of a new mail, or it could be a change to a Web page. When an event occurs, the agent has to recognize and evaluate what the event means and then respond to it. This second step, determining what the *condition* or *state* of the world is, could be simple or extremely complex depending on the situation. If mail has arrived, then the event is self-describing, the agent may then have to query the mail system to find out who sent the mail, and what the subject is, or even scan the mail text to find keywords. All of this is part of the recognition component of the cycle. The initial event may wake. Here the software agent plays a role of decision maker. Based on the query string given the agents starts preparing the data and use a technique most suitable for the data and return the result. For example in classification many algorithms are available. The most suitable ones in the domain of application can be selected and stored in a repository. The agent can classify the instance and the results can be returned.

The results of the agent can be stored and if there is spike upward or downward than the previously stored results then it can be reported with the variation information.

SELECTION AGENT

The agent can be doing a selection of appropriate attributes for data mining process. The agents can also select a particular mining technique. Thus we can have two type of selection agents.

1. Attribute selection agent
2. Algorithm selection agent

The job of the agent is to choose the required attributes that will help the mining technique to predict the class into which an instance of data will belong. Given N number of attributes, the job of the agent is to choose a set of N' attributes, called the "qualifying attributes" that will contribute mostly in classifying the given data into a particular class in minimum amount of time and efficient utilization of resources. For example the customers can be classified according to their spending habits as "high spender" or "moderate spender". Given a set of n transactions it will be useful to find, the spending pattern of customers on their each transaction. The proliferation of feature selection algorithms, however, has not brought about a general methodology that allows for intelligent selection from existing algorithms. In order to make a "right" choice, a user not only needs to know the domain well (this is usually not a problem for the user), but also is expected understand technical details of available algorithms. Therefore, the more algorithms available, the more challenging it is to choose a suitable one for an application. Consequently, a big number of algorithms are not even attempted in practice and only a couple of algorithms are always used. Therefore, there is a pressing need for intelligent feature selection that can automatically recommend the most suitable algorithm among many for a given application [3].

Feature selection is a process that selects a subset of original features. The optimality of a feature subset is measured by an evaluation criterion. As the dimensionality of a domain expands, the number of features N increases. Finding an optimal feature subset is usually intractable [8] and many problems related to feature selection have been shown to be NP-hard [9]. A typical feature selection process consists of four basic steps namely, subset generation, subset evaluation, stopping criterion, and result. Knowledge and data about feature selection are two key determining factors. Currently, the knowledge factor covers purpose of feature selection, Time concern, expected Output Type, and M=N Ratio—the ratio between the expected number of selected features M and the total number of original features N. The data factor covers Class Information, Feature Type, Quality of data, and N=I Ratio—the ratio between the number of features N and the number of instances I.

The algorithm selection agent can choose between algorithms if association between objects required or a classification based on some attributes is required. Clustering techniques can be used for grouping items into different categories. A prediction model can be computed based on previous training test or by cross validation.

RECOMMENDATION AGENT

The user navigation pattern can be identified and be supplied to indecisive users. For example if the state space consists of all possible combination of navigation, then the job of the agent is to group all similar interest users and give them directions for recommendation. To develop mining tasks for navigation pattern, we need to estimate how similar two mining patterns are. We introduce a metric that estimates the similarity based on the total number of common categories that coexist to the total number of distinct categories.According to [11] the sequential subsequence can be obtained from decisive users and help the indecisive users.

For example, if association between an item-X with that of n number of other item. We can use this metric to identify users who share common navigation behaviour and search interests. These are people who search for new information in a similar way. They form the "decisive group". We can use k-means algorithm to group similar interest users and the agent help the indecisive users with the clusters obtained to recommend popular predictions.

EXPERIMENTS AND RESULTS

We conducted mining on the data base of an online ticket booking operator, ticketgoose [7]. An SQL dump was made available whose information from 2007 to 2010. The database had 128 tables each with a minimum of 2 attributes to a maximum of 5 attributes. A set of questions were prepared to find what the operator was interested in. Equipped with this information the database was pre-processed and the selected attributes were stored in separate files. Some pre-processing task had to be performed. For example the locations were all given ids which were integer attributes. So they were converted to nominal attributes initially. Some of the tables contained null attribute values for example the comment attribute of the user feedback table. The value was not going to affect the classification criterion. So it was left as such. Summary statistics of attributes helped in evolving the results.

A subset of questions and the results are submitted herewith. Clustering was done based on travel date to find the number of persons who travelled through ticket goose between 2007 to 2010 (in fig d). The number of bus operators, buses, trading partners, agents registered with ticketgoose in fig. e . Summary statistics helped to achieve the number of hits per year to the website (in fig f).To predict the most sought after destination during peak holiday period such as pongal holidays from Chennai (in fig g). How did you know about ticket goose?(In fig h). Relation between booking and browsing (in fig. i).Browsing can be made not only for booking but also for cancellation and browsing the ticket status or casual browsing. This statistics may help website designers for advertisement and other deals.

FIG d. THE NO. OF PERSONS TRAVELLED

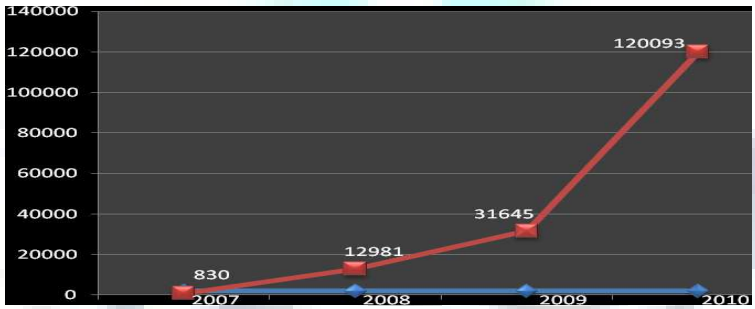


FIG e. THE STAKE HOLDERS OF THE COMPANY

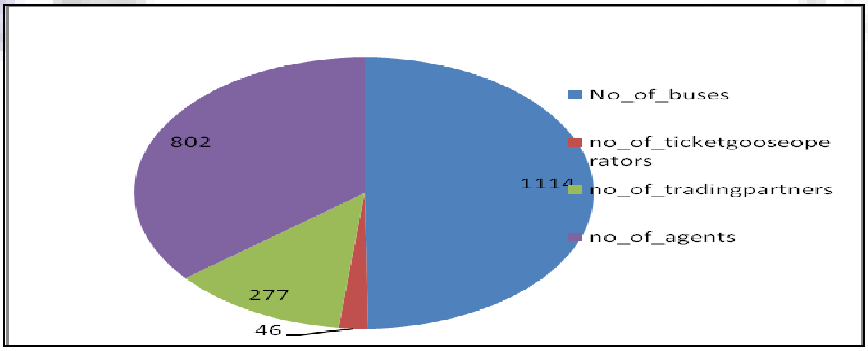


FIG f. NO. OF HITS PER YEAR

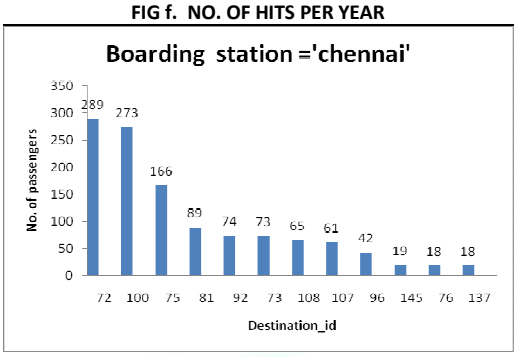


FIG g. NUMBER OF PERSONS BOARDING CHENNAI

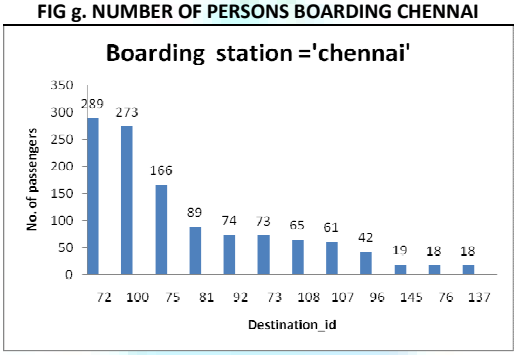


FIG h. HOW DID YOU NOW ABOUT TICKET GOOSE?

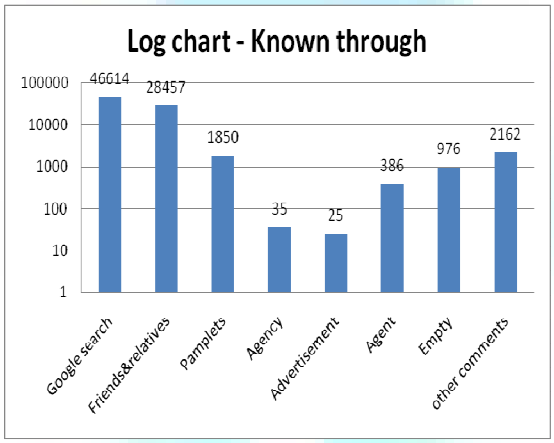
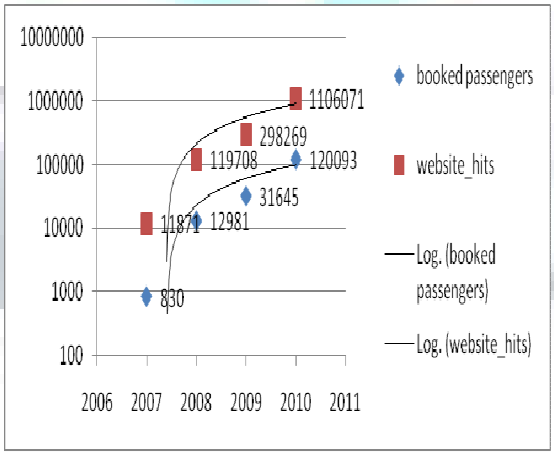


FIG i. NO. OF HITS VS PASSENGERS BOOKED



FUTURE DIRECTIONS

We had presented an automated mining framework which has been attempted by very few. The possibility of creating such a tool when non mining experts find it difficult to distinguish between of algorithms, resource utilization and optimized results for their business leverage. We plan to study further how an agent can recommend results based on user navigation of the tool.

ACKNOWLEGEMENTS

We would like to acknowledge Mr. Arunkumar Athiappan of ticketgoose, for provding us the dataset and accepting the results provided herewith.

REFERENCES

[1] Olszak, C & Ziemba, E. "Approach to Building and Implementing Business Intelligence Systems," Interdisciplinary Journal of Information, Knowledge, and Management, 2, 2007, 135-148.
[2] W. H. Inmon,"Buiildiing the Data Warehouse::Gettiing Started",2000 ,wiley & sons.
[3] H.Liu, L.Yu,"Toward Integrating Feature Selection Algorithms for Classification and Clustering", IEEE Transactions on knowledge and data engineering, vol. 17, no. 4, April 2005.
[4] S.Russell, P.Norvig, (1995.),"Artificial Intelligence: A Modern Approac", Prentice-Hal.
[5] Seydim A. Y. (1999)" Intelligent Agents: A Data Mining Perspective", Department of Computer Science and Engineering, Southern Methodist University, Dallas. http://engr.smu.edu/ yasemin/agentsdm.pdf,
[6] J.P.Bigus, J. Bigus(1998)," Constructing Intelligent Agents with Java - A Programmer's Guide to Smarter Application"s, John Wiley & Sons Inc.,
[7] Online ticket booking service, viewed on October 2011,http://www.ticketgoose.com/
[8] R. Kohavi and G.H. John, (1997) "Wrappers for Feature Subset Selection," Artificial Intelligence, vol. 97, nos. 1-2, pp. 273-324.
[9] A.L. Blum and R.L. Rivest, (1992),"Training a 3-Node Neural Networks is NP-Complete," Neural Networks, vol. 5, pp. 117-127,.
[10] Han, J. and M. Kamber, 2001. Data Mining: Concepts and Techniques, Morgan Kaufmann,USA.
[11] P-miner: Using user navigation patterns for personalizing topic directories, viewed on September 10, 2011, http://www.dblab.ntua.gr/~pbour/p-miner/

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mails i.e. **infoijrcm@gmail.com** or **info@ijrcm.org.in** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals