

# An Effective E-Commerce Management using Mining Techniques

Navneet Singh<sup>#1</sup>, Muhammad Rukunuddin Ghalib<sup>\*2</sup>

<sup>#\*</sup>School of Computing Science and Engineering  
VIT University, Vellore, India  
<sup>1</sup> navneet08singh@gmail.com  
<sup>2</sup> ruk.ghalib@vit.ac.in

**Abstract**— Web Mining is a technique to extract useful information through web, here we are using web usage mining approach, in this mining scheme some special features are there like analyse usage data, surfing information, and Marketing information. Here we are using web usage mining algorithms to find the useful information about the users, We are providing items to the users and as per the ratings we are collecting the ontology information about the users. Users can select and give the ratings for the items, suppose we are providing movie to the users so as per the users interest users can give the ratings, Organization can check the users information. Here we are using S-PLSA algorithms for the online reviews with the help of these algorithm we can get and analyse the user interest and user behaviour. Clustering is used for the same times of items and for the sentiments, we are preparing chart for the sales information where companies can check the sales information, and predict the sales information for future .We are gathering the overall ratings and we are using Collaborative filtering for the result which is given by the user through web mining technology, users can select the items as per own choice and through the payment portal they can pay the amount also. Organizations can check the review details and they can prepare better strategy for the future.

**Index Terms**- Web Mining, Clustering, S-PLSA Techniques, CRM, Collaborative Filtering.

## I. INTRODUCTION

Now a day, posting reviews online has become an increasingly popular way for people to express opinions and sentiments toward the products bought or services received. Analysing the large volume of online reviews available would produce useful actionable results for the better service[5][7][16]. In this paper, we are presenting E-Commerce features and deal with the problem of mining reviews for predicting product sales performance. Our analysis shows that both the sentiments expressed in the reviews and the quality of the reviews have a significant impact on the future sales performance of products in question[1][10][15]. For the sentiments we are using Sentiment PLSA (S-PLSA) approach, in which a review is considered as a document generated by a number of hidden sentiment factors, in order to capture the complex nature of sentiments[8][13]. Organization can check the final ratings about the movies for which one option is available they the filter the database and they can get the final ratings about the items. Web content mining and web usage mining we are using for the review as per the figure1 given below with various categories of web mining techniques.

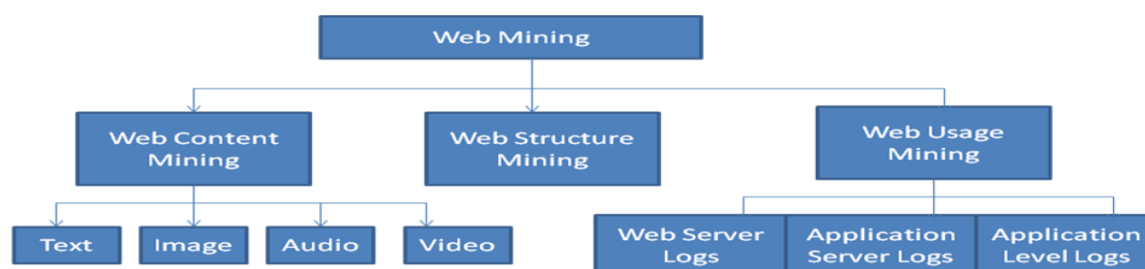


Fig.1. Categories of Web Mining (It shows the categories of the web mining and what are the uses of each web mining category)

## II. RELATED WORK

Since review mining is a part of text sentiment analysis, it is related with work of subjective classification and sentiment classification. Most existing work on sentiment mining focuses on determining the semantic orientations of documents[2][3][14]. Among them, some of the studies attempt to learn a better or bad results means like positive and negative class, In this paper we are using two types of web mining techniques first one is web content mining we can gather and mine the audio, video, text related things, then we are using web usage mining in which we can collect the information about the users[11][6], his session and

his whole record, with this concept we can give the ratings and own choice regarding any product, through clustering we can easily predict the same types of result and items. Web mining usage for different purpose we can use this concept for many purpose as follows: A) Subjective Classification B) Sentiment Classification C) Review Mining D) Clustering E) Users behavior [18]

### III. IMPLEMENTATION

The users enters to website and then register for the use after registration users can get his login details then with this users can access the website after the login users can select multiples items and as per the quality of the product he can give the ratings also, users can give the payments online [4][17]. He can select as per the category he need like suppose he wants to see the movies then he can select own types of movies and gives the ratings for that, another side admin can put the items and delete the items, he can check the response also about the movies, and at the admin side with graph representation admin can get the whole sell details profit loss, which is bad or good and the response of the customer and customers behaviour also [10][12]. At last if the admin filter the database which is stored in the weblog then he can get the final result like average, good or very good, bad like that.

#### A. Existing System

Consider the past sale performance of the same product, in the movie domain, past box office performance of the same items. We capture this effect through the use of an Autoregressive (AR) model, which is used for series of analysis problems, especially in econometric contexts [4][10]. Accuracy and Efficiency will be less. We don't analyse the feeling of people about a items. We can't analyse the future of one product. Vendors are considering only the volume of items sales. Users' information details not easy to get everytime, sales information for future very hard to predict, Customer relationship management techniques is tough to apply. It is difficult to achieve better result.

#### B. Proposed Model

We are proposing an approach which is called as a novel approach for sentiment mining based on Probabilistic Latent Semantic Analysis (PLSA), which we call Sentiment PLSA (S-PLSA). Different from the traditional PLSA, S-PLSA focuses on sentiments rather than topics [5]. Instead of considering all the words we focus primarily on the words that are sentiment related. We propose the S-PLSA model, which through the use of appraisal groups provides a probabilistic framework to analyse sentiments in reviews [7][13]. The sentiment-aware model is for predicting future product sales. We can get the users details, filtered results, ratings about the items [2].

#### C. Review Mining

With the rapid growth of online reviews, review mining has attracted a great deal of attention. Early work in this area was primarily focused on determining the semantic orientation of reviews. Among them, some of the studies attempt to learn a positive/negative classifier at the document level [6][10]. There are also studies that work at a finer level and use words as the classification subject. They classify words into two groups "good" and "bad" and then use certain functions to estimate the overall "goodness" or "badness" score for the documents [7][11].

#### D. Sentiment PLSA:

In this section, we propose a probabilistic approach to analyzing sentiments in reviews, which will serve as the basis for predicting sales information. We first consider the problem of feature selection, how to represent a given review as an input to the mining algorithms. We are using appraisal taxonomy [10][6]. Sentimental Analyzer is used to analyze the sentiment terms of people about one product. Based on Sentiments such as Type of Movie (Horror, Family Movie, Heart Touching) we are going to mine reviews of movie to improve sales performance [5][8]. In most of the studies cited above, the sentiments are captured by explicit rating indication such as the number of stars, few studies have attempted to exploit text mining strategies for sentiment classification. Our work is similar to in the sense that we also exploit the textual information to capture the underlying sentiments in the reviews. Compute the (relative) frequencies of various words in a given blog, Use the resulting multidimensional feature vector as the representation of the blog, sentiments are often multifaceted, differ from one another in a variety of ways, just classify the sentiments expressed in a blog as either positive or negative, too simplistic, a blog can be considered as being generated under the influence of a number of hidden sentiment factors, each hidden factor focusing on one specific aspect of the sentiments, accommodate the intricate nature of sentiments model sentiments and opinions as a mixture of hidden factors and use PLSA for sentiment mining [7][9]. With the sentiments we are applying clustering and we are also using appraisal words for the customers evaluation, We are using Probability Technique for the review mining and in which we are applying S-PLSA Algorithm. (Refer from <http://www.patentgenius.com/patent/7844449.html>) November, 2010.

Now we are formally present S-PLSA notations, suppose we are given a set of review  $\mathbf{F} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  and set of words (Appraisal words) from a vocabulary  $\mathbf{R} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ . The review data can be describes as a  $\mathbf{N} \times \mathbf{M}$  matrix  $\mathbf{H} = (c(\mathbf{e}_i, \mathbf{s}_j))_{i,j}$ , where  $c(\mathbf{e}_i, \mathbf{s}_j)$  is the number of times  $\mathbf{w}_j$  appears in review  $\mathbf{e}_i$ . each row in  $\mathbf{H}$  is then a frequency vector that corresponds to review [10].

#### Algorithm:

01. Pick a blog document  $b$  from  $B$  with probability  $P(e)$ ;
02. Choose a hidden sentiments factor  $z$  from  $Z$  with probability  $P(z|e)$
03. Choose a word from the set of appraisal words  $W$  with probability  $P(s|z)$

04. Result-

$$P(e,s) = P(e) P(s|e)$$

Where,  $P(s|e) = \sum_{z \in Z} P(z) P(e|z) P(s|z)$

05. Estimate mode Parameters:  $P(z), P(e|z), P(s|z)$

06. Maximize the following likelihood function:

$$L(F,R) = \sum_{e \in F} \sum_{s \in R} c(b,w) \log P(b,w)$$

**EM-Procedure:**

1. An Expectation step (E-step), where posterior probabilities for the latent variables (in our case, the variable  $Z$ ) are computed, based on the current estimates of the parameters.
2. A Maximization step (M-step), where estimates for the parameters are updated to maximize the complete data likelihood.

**In E-Step, we compute:**

$$Pr(z|e,s) = \frac{Pr(z)Pr(e|z)Pr(s|z)}{\sum_{z' \in Z} Pr(z')Pr(e|z')Pr(s|z')}$$

**In M-Step we compute:**

$$\text{Then, } Pr(e|z) = \frac{\sum_{e \in F} c(e,s) Pr(z|e,s)}{\sum_{e \in F} \sum_{s \in R} c(e,s) Pr(z|e,s)}$$

$$Pr(w|z) = \frac{\sum_{s \in S} c(e,s) Pr(z|e,s)}{\sum_{e \in F} \sum_{r \in R} c(e',s) Pr(z|e',s)}$$

$$\text{So, } Pr(z) = \frac{\sum_{e \in F} \sum_{s \in R} c(e,s) Pr(z|e,s)}{\sum_{e \in F} \sum_{s \in R} c(e,s)}$$

It can be shown that each iteration above monotonically increases the complete data likelihood, and the algorithm converges when a local optimal solution is achieved.

*E. System Structure:*

The system is depend upon the user side and admin side we can get the review result from the admin side and we can get the filtered result from the system after customers evaluation, the overall working is given below in the figure2.

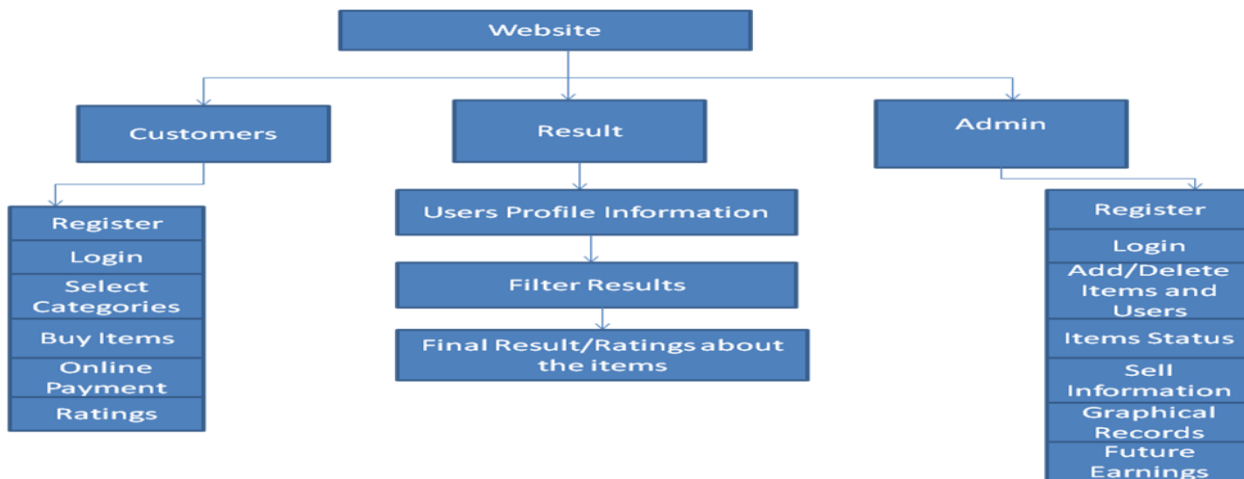


Fig.2. Architectural Diagram of the System(It shows the overall system architecture how the application works what are the modules are available, features of the application)

#### IV. RESULTS AND DISCUSSION

We are gathering the result of the current sale and future sale. users information we are fetching from the database with his session time payment details, ratings report and overall ratings for all the items means about the online marketing of an particular organization through customers with the help of this application. We are taking organization data means sales information regarding selling of items in months and year with the help of this we are preparing the review analysis with chart.(we are applying PLSA approach here) We are preferring PVR Cinema sales information for the year wise, monthly wise, week wise and day wise(Refer to <http://money.rediff.com/companies/PVR-Ltd/17040122/bse/day>) for that we are using the company dataset of the year 2012-2013 and we are comparing with the algorithm, how to achieve better revenue and how to create better relationship with the customers,so for the proposed algorithm we are analysing from the sales details of PVR Cinema and using own data sets with this we are preparing graph.We are predicting the future sales means it will be Bad or Good from the last sales. figure no.3 and Fig.6 shows the current details and future details about the sale and revenue of the organization.

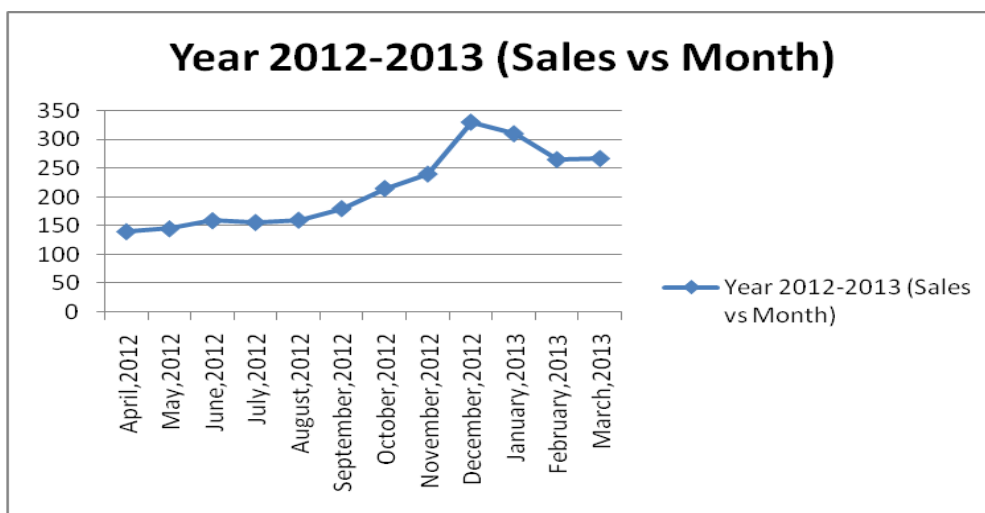


Fig.3. Refer to PVR sales Data (Year wise), It shows the year wise sales of the PVR Cinema, how it increase or decrease in each month.

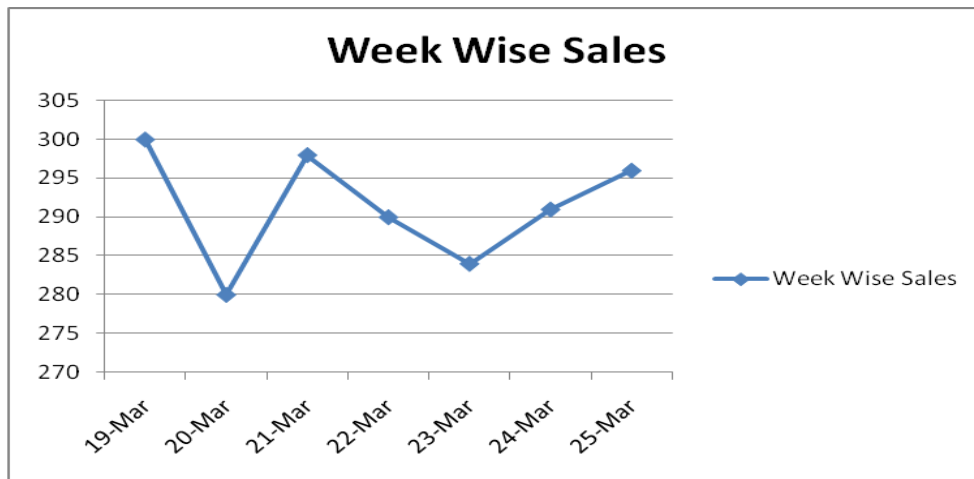


Fig.4. Refer to PVR Cinema Sales Data (Week Wise), It shows the Weekly Sales ,in each week how the sales performance are varying.

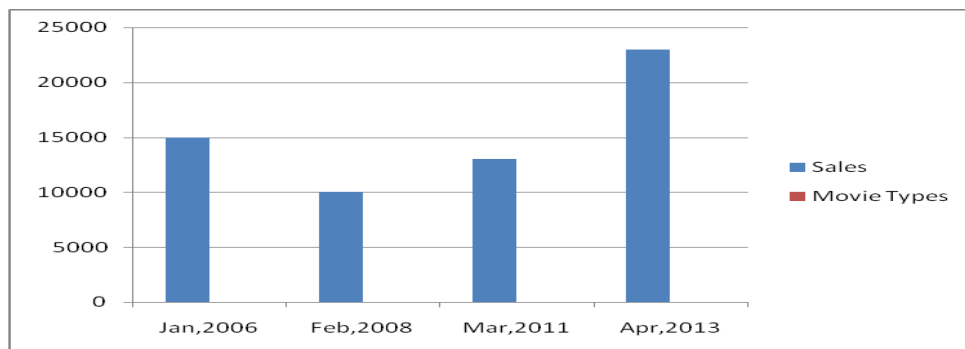


Fig.5. PLSA Result

Table.1. PLSA Analysis

Year/Month	Sales	Movie Type
Jan,2006	15,000	Comedy
Feb,2008	10,000	Romantic
Mar,2011	13,000	Horror
April,2013	23,000	Comedy

For the future sales we are gathering the databases of the organization and with this we are applying S-PLSA approach so with this we are preparing the chart for the future increment or decreament of sales and finding the ratio of sales in terms of months and year.

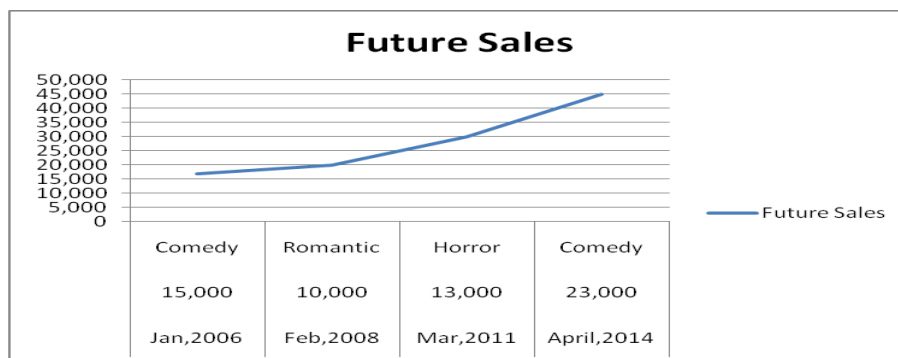


Fig.6. S-PLSA Result

Table.2. S-PLSA Analysis

Year/Month	Current Sales	Movie Type	Future Sales
Jan,2006	15,000	Comedy	17,000
Feb,2008	10,000	Romantic	20,000
Mar,2011	13,000	Horror	30,000
April,2014	23,000	Comedy	45,000

## V. CONCLUSION AND FUTURE WORK

The wide spread use of online reviews as a way of conveying views and comments has provided a unique opportunity to understand the general public's sentiments and derive business intelligence. In this paper, we have explored the predictive power of reviews using the movie domain as a case study, and studied the problem of predicting sales performance using sentiment information mined from reviews. We have approached this problem as a domain-driven task, and managed to synthesize human intelligence (e.g., identifying important characteristics of movie reviews), domain intelligence (e.g., the knowledge of the "seasonality" of box office revenues), and network intelligence (e.g., online reviews posted by moviegoers). The outcome of the proposed models leads to actionable knowledge that can be readily employed by decision makers. A center piece of our work is the proposal of S-PLSA, Using S-PLSA as a means of "summarizing" sentiment information from reviews, we have developed ARSA, a model for predicting sales performance based on the sentiment information and the product's past sales performance. We have further considered the role of review quality in sales performance prediction. With the analysis and from the dataset we are getting the sales performance and customers behavior with his full information, with company dataset we are getting his current sales details how it varies from 247-300 units or 105-350 units with respect to year, week and as per the algorithm the sales details are varies with the movie type with respect to year and month. So as per the current sales of the company we are presenting a better procedure to increase the sales in future and how to establish a better relation with the customers with his full details. For future work, we would like to explore its role in clustering and classification of reviews based on their sentiments. It would also be interesting to explore the use of S-PLSA as a tool to help track and monitor the changes and trends in sentiments expressed online.

## ACKNOWLEDGEMENT

We would like to thank VIT University for giving us the resources and the required infrastructure supports in accomplishing this research well in time, we are expressing our deep gratitude to the Management members of VIT University for their encouragement and extensive support in preparing and publishing of this paper and not the least our heartfelt thanks to PVR Cinema for making us available the required dataset online.

## REFERENCES

- [1]. Yang Liu, Xiangji Huang, Aijun An and Xiaohui Yu (2007), "ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs", Proc. of ACM SIGIR conf., pp 607-614.
- [2]. Olfa Nasraoui, Maha Soliman, and Antonio Badia (2008), "Mining and Tracking Evolving User Profiles and More – A Real Life Case Study", Proc. of IEEE Intl. J. of computer science, vol.20 Issue 2, pp 202-215 (doi.ieeecomputersociety.org/10.1109/TKDE.2007.190667).
- [3]. Mr. Dushyant Rathod, (2012), "A Review On Web Mining", Proc. of IJERT, vol.1 Issue 2.
- [4]. Olfa Nasraoui, Member, Maha Soliman, Esin Saka, Antonio Badia, Member, and Richard Germain (2012), "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", Proc. of IJCTT, vol.20, Issue 2.
- [5]. Li Zhuang, Feng Jing (2006), "Movie Review Mining and Summarization", Proc. of ACM Intl Conf., pp 43-50, doi-10.1145/1183614.1183625.
- [6]. Amsaveni, K., Vydehi. (2012), "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", Intl. J. of IJCTT, vol.3, Issue 4.
- [7]. M. Venkata Krishna, L. Raghavendar Raju, D. Jamuna, M. Gayathri (2012), "Analysis of Web Mining and Evolving of User Profiles", Intl. J. of IJEIT, vol.1, Issue 4.
- [8]. Tipawan Silwattananusarn and Dr. Kulthida Tuamsuk (2012), "Data Mining and Its Applications for Knowledge Management", Intl. J. of IJDKP, vol.2, Issue 5.
- [9]. Isinkaye, O. Folasade (2011), "Computational Intelligence in Data Mining and Prospects in Telecommunication Industry", Proc. of JETEAS, vol.2, Issue 4, pp 601-605.
- [10]. Xiaohui Yu, Jimmy Xiangji Huang (2012), "Mining Online Reviews for Predicting Sales Performance in the Movie Domain", Proc. of IEEE Transaction on Knowledge and Data Engineering, vol.24, Issue 4.
- [11]. N R Srinavasa Raghavan (2005), "Data mining in e-commerce: A survey", Proc of Shaghna, vol.30, Issue 2

- [12].Jyoti Yadav,Bhawna Mallick(2011),”Web Mining: Characteristics and Application in Ecommerce”,Intl.J.of IJECSE,vol.1,Issue 4
- [13].S.Yadav,K.Ahmad,J.Shekhar(2011),”Analysis of web mining applications and beneficial areas”,Intl.J.of IUM,vol.12,Issue 2.
- [14].Ketul B.Patel,Jignesh A.Chauhan,Jigar D.Patel(2011),” Web Mining in E-Commerce: Pattern Discovery, Issues and Applications”,Intl.J.of P2P Network Trends and Technology,vol.1,Issue3.
- [15].Yuan Chen Yuan Chen, Feifei Wang Feifei Wang(2009),” A Dynamic Pricing Model for E-commerce Based on Data Mining“,Intl.J. of Second International Symposium on Computational Intelligence and Design,vol.1,pp 363-366, DOI: 10.1109/ISCID.2009.99.
- [16].I Sumaiya Thaseen,Fazal E Sattar,Muhmmad Rukunuddin Ghalib(2011),”A Secure Adaptic Agent Model using Model Driven Architecture”,Intl.J.of Computer Information Systems,Vol.2,Issue 2,pp 56-62.
- [17].K Murugesan,Md Rukunuddin Ghalib,J.Geetanjali,J.Indumathi(2009),”A Pioneering Cryptic Random Projection Based Approach for Privacy Preserving Data Mining”,Proc.of IEEE IRI2009 Intl.conf. pp 437-439.
- [18].Prof. Ms. S. P. Shinde, Prof. Ms. S. S. Kapase, Prof. Ms. S. R. Mulik(2011),” E-Service Intelligence in Web Mining”,Intl J. of IJECBS,Vol.1,Issue 2.

#### AUTHORS

**First Author** – Navneet Singh, M.Sc Computer Science, VIT University,Vellore,Tamil Nadu,India  
E-mail – navneet08singh@gmail.com

**Second Author** – Muhammad Rukunuddin Ghalib,Assistant Professor(Senior),VIT University,Vellore  
Tamil Nadu,India,E-mail – ruk.ghalib@gmail.com