International Journal of Mechanical Engineering and Technology (IJMET)
Volume 9, Issue 3, March 2018, pp. 539–550, Article ID: IJMET_09_03_056
Available online at http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=9&IType=3
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

© **IAEME** Publication **Scopus** Indexed

AN APPROACH TO FINDING CUSTOMER-VALUE AND DEVISING MARKETING STRATEGIES BY USING DATA-MINING TECHNIQUES

Neha Dsouza
PG Scholar, Department of Computer Science, Christ University, Bengaluru, India

Joy Paulose
Professor, Department of Computer Science, Christ University, Bengaluru, India

ABSTRACT

The digital marketing space is an outcome of the advent of digital age and the booming network technology. The earlier methods of marketing have almost been eclipsed with this digital marketing, which could include any material, information, videos on the internet and social media, etc. The marked distinction between a digital and traditional marketing is that the former uses digital devices in the traditional space. People today consume innumerable products daily and getting to know the customer's interests, their transaction's history, identifying a loyal customer and estimating their lifetime value are all at the very heart of digital marketing. It is here, in this context, that data mining techniques play a key role. This unique blending of data mining techniques with digital marketing will help merchants increase their sales and revenue exponentially.

This research paper explores the possibility of finding the frequent item-set generated by a customer while making purchases, tracking their loyalty based on the number of purchases and their frequency and thereafter devising an appropriate marketing strategy for that particular product type. It is the analysis of the product based on the previous marketing experiences of the consumer. This system will boost a merchant's businesses by identifying their loyal customers (customer segmentation), best sold products (product recommendation) and reveal the different marketing strategies that can be adopted by them to reach out to their potential customers and thus increase their sales (Strategy selection/combination).

Keywords: Digital Marketing; Data Mining; RFM; Lifecycle grids; Apriori Algorithm; C4.5 algorithm; Marketing strategies.

Cite this Article: Neha Dsouza and Joy Paulose, An Approach to Finding Customer-Value and Devising Marketing Strategies By Using Data-Mining Techniques, International Journal of Mechanical Engineering and Technology, 9(3), 2018, pp. 539–550.
http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=3

1. INTRODUCTION

The top trending words in the last couple of years have been commercialization and digitization of markets. Every country that wants to be a world leader in the reckoning has been striving to provide a feasible platform for businesses to thrive and this is where international E-commerce giants like an Amazon, Alibaba, Uber, etc. have established and proven themselves alongside the indigenous performers like Snapdeal, Myntra, Ola etc. in India.

Most of the millennials (Gen Y) accustomed to digital marketing may be oblivious to the methods of traditional marketing as they have had technology available from early years. A traditional marketing generally involves printing and distributing of flyers, airing commercials on radio and TV, a-word-of-mouth marketing approach etc., while a digital marketing space involves the use of latest technologies like company websites and social media apps to attract customers [1]. The brighter side of Traditional marketing is that it has more personal touch associated with it, but its limitations are its costs, ambit of reach (as it could as much concentrate only on a specific localized area or people) and that it hardly allows for instantaneous responses and feedback.

The new-age digital marketing on the other hand has now become the norm. From groceries, clothes, equipment, cab rides to courses offered; one uses the different E-commerce sites and company websites that are available to make purchases. Interestingly, it even helps to create a wide customer-base spread across the world and also allows for instant branding. No geographical boundaries can restrict this kind of marketing.

The digital space allows for a two-way interaction and these responses/results can be measured in real time. The strength of a new-age digital marketing world is its ability to be a cost-effective and a timesaving platform. A digital platform unlike the traditional markets allows for a level-playing-field and opponents competing with one another end up offering the most-competitive price for customers. The audience gets to play the king-maker and the costs remain dynamic. The digital platform is intelligent enough to record the preferences of the customers and guide them on their future purchases. This is what is leading to now being called the big data or data mining which will determine the future course of our world trade and economy.

2. LITERATURE REVIEW

2.1. Customer Relationship Management

Interaction of business with customers can be termed as Customer Relationship Management (CRM). It essentially involves devising effective marketing strategies, offering premium services and using superior technology to attract and maintain loyal customers with an intention to increase business efficacy. Organizations that are able to executive seamless CRM will be the ones to survive in this competitive age, where currently it is the customer's loyalty defining the profitability of a company.

The Building Blocks of CRM:

- Maintain a database about regular customers
- Develop seamless methods to accumulate data
- Have a plan to scrutinize the information in the database and arrive at analysis
- Devise strategies to use the analysis to categorize and track potential and loyal customers

The architectural framework of CRM can be categorized into Operational CRM and Analytical CRM [2]. The computerization of business processes is known as Operational CRM, while the examining of customers uniqueness and actions in order to prop up an organization's customer management policy is known as Analytical CRM. It is the Analytical CRM that enables an organization to allocate its resources effectively for the benefit of customers. One of the most popular approaches to analyzing a customer's information in the Analytical CRM is the Data mining techniques/tools and the most suitable tools will help haul up the most important information about customers, thereby helping organizations make decisions impacting these very customers.

2.2. Customer Loyalty Analysis

A customer's loyalty is not easy to measure and repetitive purchases by a customer are no proof of his/her loyalty. Although, the previous purchase activities of a customer are indicative, but may not however hold good in all cases of tracking the patronizing intent of a customer. Author Bob Hayes has broken this loyalty analysis into 3 measuring segments: they are Retention, Advocacy and Purchasing [3].

Retention is the extent to which a customer is willing to continue with the products and services of a specific company and this sometimes helps gauge his/her loyalty. Advocacy is a kind of advertising of a product or service by the customer whose common belief and perception about products and services are very strong and positive. Purchasing is the willingness of the customer to invest his money on buying different products and services from the same company. All the three components i.e. Retention, Advocacy and purchasing can indicate the loyalty of a customer, increase the quantum of purchase and willingness to spend more.

2.3. Recency Frequency Monetary (RFM)

Who are paramount customers? Who is a devoted customer? Which customers should be retained? [4] RFM model that was proposed by Hughes in 1994, has been used in the field of marketing since a very long time. It is a method of data analysis that was defined to analyze consumer behavior [5]. This model categorizes a consumer's activities and divides them into three variables: Recency Value, Frequency and Monetary. Recency value is how recently has the customer purchased the product (how long before the transaction was made), Frequency is how often the customer has made transactions at a particular store, Monetary is how much amount was spent by the customer in a specified time.

Applications based on RFM concepts in data mining have been introduced in various fields such as Computer security, electronics industry and automobile industry. RFM variables induced for data mining techniques include neural network, rough set theory, genetic algorithm, sequential pattern mining and decision tree [2].

The blending both RFM analysis and Data-Mining techniques can yield constructive and useful information on existing and new customers. For example, the clustering in the RFM analysis offer better information on a customer's real marketing activities than just mere cluster analysis.

2.4. Life Cycle Grids

Life Cycle Grid is an idea based on the usage of thresholds or hurdles to understand a consumer's behavior than mere ranking (where the score of a consumer remains static although there are changes in the composition of data). The RFM or RF scores indicate the probability of a customer's response and loyalty for today, while the LifeCycle Grid indicates the possibility of a customer loyalty in future too. While Scores are short-term ideas, Grids are long-term ideas. When RF scores are tracked over a long period for a particular customer, it may almost accomplish the objective of Grids, with the exception of the truth that the scores of RF are altered even though the behavior of the customer hadn't changed albeit some data changes. As a result, it may end up prompting false indictors over a period of time.

Example, if someone deemed best buys a product while having the Recency score of 45 days and Frequency has 35 purchases is still best, no matter how many records you add to the database. If the cell boundaries are marked using a comparative grading/ranking, the mock change in grade/rank will not take place. These hurdles and thresholds indicate a customer's status and not his/her ranking. While RF scores indicate a customer's behavior at a point in time, the LifeCycle Grid is an indicator of the customer's behavior over a period of time.

2.5. Data Mining

The unified goal of this process (Knowledge discovery in databases- KDD) is to extract hidden/ useful information from large datasets [6]. In this process, raw data is keyed in, which is later stored in files, spreadsheets or large systems such as repositories. This data is then pre-processed i.e. data cleaning is done by removing the noise, missing values, duplicate entries after which data mining techniques are applied to extract relevant information from the data collected along the preprocessed data. Finally, data visualizations are employed to display the results.

The process of Data Mining involves the integration of conventional data analysis approach alongside the modern algorithms for processing of large databases. The goal is to understand a pattern in a customer's purchase behavior, track them and evolve new cost-effective strategies. Organizations that have large databases can use this technology. The applications where data mining is currently used are retail, medical services, finance and others. Data Mining as a subject involves multiple areas of study like Databases, Statistics and Machine Learning. The different techniques existing under data mining are [7]:

1. Association Rule Learning: - involves finding connection between items and generating association rules amid the many variables and is also termed as market basket analysis.
2. Clustering: - involves grouping of related data items and is also termed as unsupervised.

Classification

1. Classification: - involves categorizing data according to their class/group and is also termed as supervised classification.
2. Regression: - involves devising a function that will mirror the data with minimal slips.
3. Summarization: - involves presentation of reports and visualization.

To ascertain the association rules, this study involved using The Apriori algorithm and to perform the classification analysis C4.5 algorithm was activated.

2.6. Apriori Algorithm

Association is a known data mining technique where an Association, is a pattern established based on connections between items in a transaction. Association Rule involves locating the item-sets having support more than the minimum support and thereafter using large item-sets to produce the preferred instructions that have confidence far greater when compared to minimum confidence. An example of association rules is market basket analysis.

Apriori Algorithm is a one-way technique for market basket analysis. It is the most used algorithm for frequent item-sets mining and finding relevant association rules. It operates on the database containing a lot of transactions.

Association rule mining problem can be defined as: [8]

- I= i1, i2, ..., in, I denotes a set of items being purchased by the various customers.
- D= t1, t2, ..., tm, D is a set representing all transactions whereby every transaction is given a transaction id and the items purchased associated to it.
- Association rules *X => Y* which means that the transactions of the database which have X items have the items Y included also. Here X, Y are a subset of I and no two items can belong to both the sets.
- Support(X)= Support of the item-set is the count of occurrence an item appears in the whole database.
- Confidence (X =>Y) = is the likelihood of an item X being purchased, when any other Item is being purchased.

Apriori works on the idea of a seamless approach to using n-Item sets to locate (k+1)-Item sets. The count of each item in the set is collectively taken to find the first Item set. After this, the 1-Item sets are undertaken to find 2-Item sets and the subsequent ones till k-Item sets are exhausted; while all the items end in one last observation as noticeable in the final row of the table above. Each exploration can take a single scan of the entire dataset.

2.7. C4.5 Algorithm

C4.5 algorithm used to generate a decision tree was developed by Ross Quinlan [9] This tree can be used for classification which involves identifying a new observation belongs to which set of categories (sub-populations) on the grounds of a preparing a set of data comprising observations and instances whose group affiliation is identified. The C4.5 algorithm is a computer program to bring about classification rules as a decision tree from a group of specified instances. A decision tree is a forecast prototype using tree structure and is understood to be a very efficient technique to supervised learning. It basically divides the dataset into uniform sets to enable prediction of the variable. A set of ordered data is taken as the input which eventually gives an output resembling a tree, where the final points of the tree signifies a decision and the non-final points in the decision tree represents a test. C4.5 algorithm is most often used as it has a greater exactitude in decision making and it was preceded by ID3, where the parent and root are chosen based on both information gain and gain ratio. This is because parents are selected by firstly discovering the split information, negating the predisposition of Information gain and evolve a method of normalization to information gain by using a split information value.

Algorithm

- Let the set of training instances be represented by T.
- Next, select an attribute that distinguishes the samples in the training set.
- Since the algorithm uses Gain ratio, we determine.

GainRatio = Gain/splitInformation

The resulted information by dividing the set T into n number of partitions is represented by the split information value corresponding to the attribute.

- Later, construct a tree node based on the value Gain Ratio, where gain ratio is maximum.
- Next, child links are created from the node, whereby each of these links denotes a unique value.
- The values of these child links are further subdivided into sub-classes.

3. METHODOLOGY

Dataset used in this paper was from FoodMart2000. The dataset contained variety of food and non-consumable products at various stores in different cities. The data chosen for this paper was duplex supermarket in Tacoma. To find the most frequent item-set, recent and infrequent customers and the best strategy to promote a product, the proposed framework was divided into four stages.

1) Anatomy of data.
2) Product purchase Analysis.
3) Client Significance.
4) Devising Marketing Strategy.

1. Anatomy of Data

In the First stage, the dataset was analyzed and preprocessed. The required data such as product information, transaction history and marketing ways were noted.

- Product Information - such as name, category, product family, store details
- Transaction history - such as date of purchase, units, unit cost, customer name etc.
- Marketing behavior - promoting channel such as media type, promotion name, cost incurred etc.

The tools used in this paper are R statistical tool, Weka tool.

2. Product Purchase Analysis

Market Basket Analysis is a modeling system founded upon the supposition that when one purchases a definite set of items, he/she is more (or less) expected to purchase an additional cluster of items. The set of items purchased by the customer is taken as an item-set and the market basket analysis proposes to discover associations between purchases. Apriori algorithm was used to find the relationship between the products and find the item sets.

To find the frequency of items being purchased in a particular transaction, we needed to calculate the minimum support and minimum confidence for the data. For the data at the duplex store, Tacoma used the support was given 0.001 with confidence 0.8 and these parameters were used to exclude rules in the result that have a support or a confidence lower than the minimum support and minimum confidence respectively.

lhs		rhs	support	confidence	lift	count
{1083,717}	=>	{1436}	0.0012547051	1	140.3471	3
{1353,1389}	=>	{1436}	0.0012547051	1	140.3471	3
{1389,1436}	=>	{1356}	0.0012547051	1	199.2150	3
{1353,1416}	=>	{1389}	0.0012547051	1	183.9231	3
{1353,1389}	=>	{717}	0.0012547051	1	125.8421	3

Figure 1 Basket Rules.

Using the Support and confidence values there are 15 basket rules generated. From the figure 1 {1356,1389} => {1416} rule meant that along with the products 1356 ,1389 the customer was likely to buy 1416 also. The figure 2 shows the grouping of products for the rules generated.

Figure 2 Grouped matrix of 15 rules.

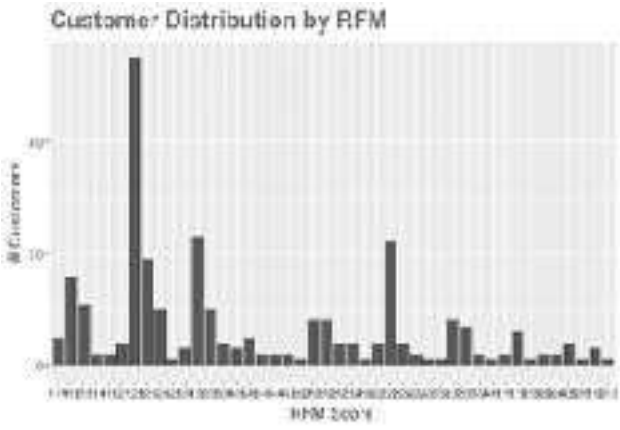
3. Client Significance

RFM algorithm is used to segment the customers based on their behavior. Recency is the time period between the recording of the last purchases by the customer and the time when the customer data is being analyzed. Frequency is a reference to the count of transactions within a definite time. Monetary value refers to the amount expended in a certain period. The ranges of recency, frequency, monetary were divided into 5 intervals each and then the values were assigned to the intervals they fit in. The RFM score was the combination of the R, F, M value.

customer_id	recency	frequency	avg_sum	rankR	rankF	rankM	groupRFM
79	209	17	5.279535	5	1	1	511
138	84	48	8.470083	1	3	2	132
235	84	32	8.111328	1	2	2	122
335	57	33	8.051229	1	2	2	122
341	84	63	8.296314	1	3	2	132

Figure 3 Scores for Recency, Frequency, Monetary.

From figure 3, the Rfm Score of 511 was denoted as recency score of 5, which meant that the customer had made a recent purchase, frequency score of 1 indicated that the number of purchases made was dismal, monetary score of 1 denoted that the average amount spent was low. The figure 4 describes the RFM scores against a given number of customers.



!

Figure 4 The distribution of customers under the various RFM scores.

K-Means Clustering with Recency, Frequency and Monetary

To calculate the loyalty of the customer, RFM scores generated above were considered individually and k means algorithm was applied on the dataset [Birant, Derya (2011)]. In this paper, cluster of 4 was considered and cluster labels such as very high, high, medium and low were assigned to the cluster. Figure 5 shows the scatter plot of the customers in their respective clusters.

By considering the output of the customers from the product purchase analysis step and then plotting them based on the clusters detected, from figure 6 it could be inferred that there are products that are purchased by both loyal and one-time customers on account of their everyday usage and hence it is essential that these products are stocked in more, so that a loyal customer never goes dissatisfied owing to shortage created by one-time or stray customer.

!

Figure 5 3D Scatter plot of the customer with loyalty very high, high, medium and low

!

Figure 6: Scatter plot of the very high loyal customers showing how frequent they buy the itemsets generated

RF with LifeCycle Grids

We may find dissimilarities between a customer purchasing a set of products just once (as a stray incident) and those who might end up purchasing those products 4 or 5 times, but it becomes increasingly challenging to differentiate between a customer who might have purchased 55 times as compared to another who would have purchased 60 times. Therefore, it is imperative to define limits from lower values to high.

Figure 7 Lifecycle Grid mapping the customers recency and frequency score

Figure 7 illustrates the Lifecycle Grids based on the recency and frequency score of the customers. For example, there are 67 customers who have recently purchased i.e. 41-80 days back and have also frequently purchased and hence titled recent frequent.

4. Devising Marketing Strategy

For a store, key challenge faced on an everyday basis is finding traffic of purchasers. In this stage, a selection and combination was done based on the media and promotion to hit upon out the best strategy, merchants could use to advertise their products.

There are various media types used in the world of digital marketing:

A. Social Networking Marketing

The concept Social Networking (SN) involves creation of a profile by a user by providing personal data, pictures etc. and then connecting with other similar users to share information through messaging, chats, etc. This happens either on mobile or computers through internet which does offer open as well as closed communication with others. The degree of access to the variety of information on different profiles can be restricted. It can optimally be used by individuals and organizations to generate interest in their products and services through Social Networking Marketing.

B. Email Marketing

Email Marketing as a concept involves sending advertisements to a customer's mailbox through the internet so as to publicize their products and services. However, the problem of plenty has now made it a cause of concern and most of these are tagged as spams. Here, it is important to have the consent of the customer and tweak the advertisement based on their preferences.

C. Mobile Marketing

The ubiquitous mobile phone has become the best source of communication and hence marketing on these devices has evolved and are in fact becoming creative by the day. These marketing techniques involve messaging, product-calls, etc.

D. Search Engine Marketing

Today's companies are tailoring their website to gain favor in the ranking order on Google search engine. If a company can identify a relevant keyword or a phrase, it can have large number of customers navigate to their website. There are even paid-advertising campaigns conducted to invite maximum traffic to their websites.

E. Content-based Website Marketing

The content marketing strategy relies on the quality of resources, the strong distribution space, having a relevant audience loyal to the content. These websites provide information on products and services aiding a customer's perception about a particular product or service. It involves using electronic newspapers and other professional messaging services to market their products by writing articles about their design, features, availability, uniqueness.

F. E-commerce portal site

The e-commerce sites are virtual markets where customers can navigate for booking their orders, making payments, interact for all other types of services, etc. It is a marketing and trading platform for both the seller and purchaser. Along with the media types, there are various other promotions that are used to attract customers

G. Freebie marketing

In order to promote alertness and responsiveness among their target customers, a freebie marketing strategy is used as an effective marketing technique to increase the company's sales. The strategy involves either giving an additional product at a low cost or even free with the main product. Example: the accessories that are provided free on purchase of a mobile phone.

H. Prize winning marketing

The prize-winning marketing strategy involves incentivising or rewarding the customers on participating in contests and competitions which are organised to raise awareness of a product by a company. This kind of strategy has over time evolved and there are takers for this on both sides of the market.

I. Pull marketing

The pull marketing strategy employed by the company attracts the customers to the company's web-site and its products, whereby the customer gets actively involved in searching and subscribing to information on the products and services. It can be customized to suit each individual customer's needs.

J. Push marketing

The push marketing strategy is just a one-way approach to receiving information, where it is intended to achieve product promotion in the form of e-mails, radio and television programs etc. It cannot be customized to each individual customer. To find out the best marketing strategy that can be used by the store to gain more sales. First, we calculate the sales individually for each media that has been made throughout the year. Weka tool is used to generate the decision tree based on the sales, media and the promotion name being used to find a marketing strategy where the merchant can use to increase his sales.

Figure 8 Decision tree performed over the sales earned through various platforms

Figure 8 confirms that when advertising promotions such as" two for one" are used on a content-based website platform (in other words a media platform), they eventually turn out to be the best strategies for promotion of products. Thus, the marketing strategy recommended was a freebie marketing on a content-based website platform. Based on the analysis as performed in the previous step of client significance, the highly loyal customers along with the promotion ids were considered. When crosschecked, it was inferred that the most loyal customers came from strategies such as E-mail Marketing platform with Super Duper Savers or SNS platform with Save-It Sale.

4. CONCLUSION

The digital environment is today witnessing many highs and lows. Although, there is a deluge of online customers, but the rapidly changing needs and loyalties of customers beckon a variety of marketing tactics. In a scenario such as this, it is incumbent to respond to the changing consumer needs and demands by proposing a highly structured digital marketing architecture with the unique combination of information technology and business management techniques so as to track and analyze customer data for the benefit of both customer and company.

Overall, the architecture of digital marketing first proposed analyzing the transactions made by the customer and selecting the best item-sets that were purchased which helped the merchants know their customer's requirements. Secondly, it analyzed the loyalty of the customer towards the firm by considering the attributes such as how recent was the purchase made by the customer, how frequently were the purchases made and how much on an average was being spent by the customer and all of those would finally help devise a marketing strategy by analyzing the sales made through the different media involved in promoting the items.

Additionally, it helps to identify products that are not being sold or which have fewer sales and assign it with the products that are always being sold, so has to increase their sales or otherwise discard the products which are infrequently brought. Further studies are recommended to find ways to retain the one-time, not-so-loyal and infrequent customers.

REFERENCES

[1] Parsons, M. Zeisser, R. Waitman. 1998, Organizing Today for the Digital Marketing of Tomorrow, Journal of Interactive Marketing, 12, pp. 31–53.

[2] Ngai, Eric WT and Xiu, Li and Chau, Dorothy CK, 2009, Application of data mining techniques in customer relationship management: A literature review and classification, Expert systems with applications, 36, pp. 2592–2602.

[3] Nyarko, Israel Kofi and Agyeman-Duah, Matthew Opoku and Asimah, Vincent, 2016, Measuring Customer Loyalty Using Retention, Advocacy and Patronages Key Denominators, International Journal of Scientific and Research Publications, 6, pp: 235–240.

[4] Liu, ChunNian and Zhu, XiaoWen, 2009, A study on CRM technology implementation and application practices, Computational Intelligence and Natural Computing, 2009, 2, pp. 367–370.

[5] Hughes, Arthur Middleton., 2005, Systematic approach for digital marketing strategy through data mining technology. Strategic database marketing. 4, pp. 172–181.

[6] R.J. Roiger, M.W. Geatz, 2003, Data Mining: A tutorial-based Primer

[7] Berry, Michael J and Linoff, Gordon, 1997, Data mining techniques: for marketing, sales, and customer support, John Wiley and Sons, New York.

[8] R. Agrawal, R. Srikant, 1994, Advances in knowledge discovery and data mining, Fast discovery of association rules, 12, pp. 307–328.

[9] Salzberg, Steven L, 1994, C4. 5: Programs for machine learning by j. Ross quinlan. morgan kaufmann publishers, inc.,1993. Machine Learning, 16, pp. 235–240.

[10] Birant, Derya., 2011, Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, 41, pp. 27–42.

[11] Umamaheswari R, Siva Purnima S and Dr. S. Saravana Mahesan, Customer Preservence for an Organisation using Data Mining. International Journal of Civil Engineering and Technology, 8(10), 2017, pp. 933–938.

[12] Mashael Saeed Alqhtani and M. Rizwan Jameel Qureshi, Data Mining Approach for Classifying Twitter's Users. International Journal of Computer Engineering & Technology, 8(5), 2017, pp. 42–53.

[13] Harshita Kotian, Divya Bhat, Charvi Kunder, Prof. Snehal Kulkarni, Data Mining Approach for Generation of Optimal Internet Telecom Plans, International Journal Of Computer Engineering & Technology (IJCET), Volume 5, Issue 6, June (2014), pp. 74-81