

# Data Mining, Internet Marketing and Web Mining

Vishesh S<sup>1</sup>, Manu Srinath<sup>1</sup>, Supriya Ramarao Prasanna<sup>2</sup>, Supriya Yadati Narasimhulu<sup>2</sup>

B.E, Department of Telecommunication Engineering, BNM Institute of Technology, Bangalore, India<sup>1</sup>

B.E, Department of Information Science Engineering, BNM Institute of Technology, Bangalore, India<sup>2</sup>

**Abstract:** Data mining is the computational process of discovering patterns in large scale data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. [1] In this paper, we will draw limelight on one of the applications of data mining- 'Internet Marketing'. For internet marketing, data mining is helpful in attaining increased number of page views per session, increased number of average profit per check, increased number of referred customers, increased number of visitors as well as retention rate and increased brand awareness. Data mining acts as a tool which practices like collection, integration, analysis and presentation of business intelligence data in an internetwork. We would like to discuss about web mining and how it is different from data mining.

**Keywords:** Data mining, artificial intelligence, machine learning, statistics, database systems, Internet Marketing.

## I. INTRODUCTION

Data mining automates the process of finding predictive information in large databases. Data mining tools sweep through databases and identify previously hidden patterns in one step. An example for pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together, i.e., when one purchases an item, say laptop from an e-commerce site, the software/commodity would suggest other products such as Operating System, antivirus software, etc which may not be linked to the laptop but have been previously purchased by customers historically. This would maximise business and also advertise other commodities along with the purchase of the product. The probability of the customer purchasing a particular brand can be predicted easily using data mining and analytics. Figure 1 shows the black box that makes predictions about the future based on information in the past and present.

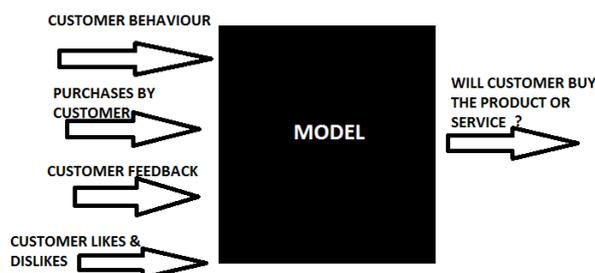


Figure 1 Black box that makes predictions about the future based on the information in the past and the present

## II. INTERNET MARKETING

Internet is a global computer network [2] providing a variety of information and communication facilities,

consisting of interconnected networks using standardised communication protocols. [3] This definition is so true that in the 21<sup>st</sup> century world that almost everyone on this planet are connected, use or are aware of the internet. It is a global phenomenon. The internet or internetwork is simply the interconnection of networks globally. Figure 2 shows the internet population of 2007 vs. 2012, a 2 times increase in 5 years. [4]

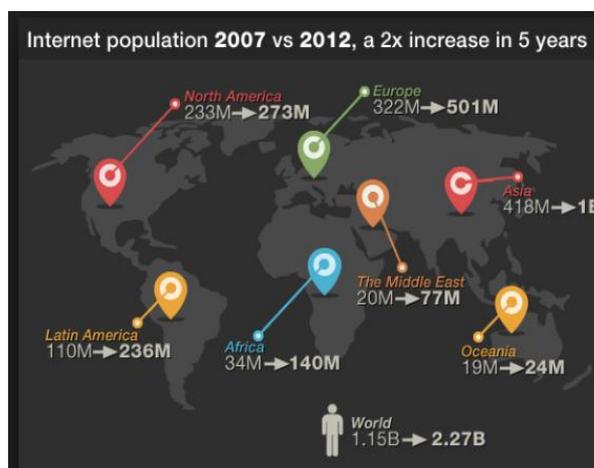


Figure 2 Increase in Internet population of 2007 vs. 2012

With the everlasting, expanding global internetwork, all businesses are targeting clients over the internet and hence internet marketing is expanding and is in a boom. Internet marketing efforts are done solely on the internet. This type of marketing uses various online advertisements to drive traffic to an advertiser's website. Banner advertisements, Pay Per Click (PPC) advertisements and targeted email lists are often used methods in internet marketing. Data



mining plays a crucial role in internet marketing, a few examples are:

#### A. Financial Banking

With computerised banking and interconnection of networks, huge amounts of data are supposed to be generated with new transactions. Data mining can contribute to solving business problems in banking and finance by finding patterns, causalities and correlations in business information and market prices that are not immediately apparent to managers because the volume of data being mined is huge. The managers may find this information for better segmentation, targeting, acquiring, retaining and maintaining a profitable customer.

#### B. Corporate Surveillance

Corporate surveillance is the monitoring of the behaviour of a person or a group by a corporation. The data collected is most often used for marketing purposes or sold to other corporations. It can be used by the businesses to tailor their products desirable by their customers. The data can be used for direct marketing purposes, such as targeted advertisements on web pages, where ads are targeted to the users of search engines or web pages by analysing their search history, emails and sessions.

### III. WEB MINING

Web mining [5] is the use of data mining techniques to automatically discover and extract the information from the web documents and services. There are three general classes of information that can be discovered by web mining:

- Web activity from server logs and browser activity tracking.
- Web graph, from links between pages, people and other data.
- Web content for the data found on web pages and inside documents.

There are four steps in content web mining [6]:

1. Collect - fetch the content from the web.
2. Parse – extract usable data from the formatted data (HTML, PDF, etc.)
3. Analyse – tokenize, rate, classify, cluster, filtering, sorting, etc.
4. Produce – turn the results of analysis into something useful (report, search index, etc.)

### IV. DATA MINING AND WEB MINING

There are three main differences to be considered between data mining and web mining. They are:

#### A. Scale:

In traditional data mining, processing 1 million records from database would be a large job. In web mining, even 10 million pages wouldn't be a big number.

#### B. Access:

When performing data mining for corporate information, the data is private and often requires access rights to read them. For web mining, the data is public and rarely requires access rights. But web mining has additional constraints, due to the implicit agreement with the web masters regarding automated (non-user) access to this data. This implicit agreement is that a webmaster allows the crawlers access to useful data on the website, and in return the crawlers promise not to overload the website, and has the potential to drive more traffic into the website once the search index is published. With web mining, there is often no such index, which means the crawler has to be extra careful/polite during the crawling process to avoid causing any problems to the webmaster.

#### C. Structure:

A traditional data mining task gets information from a database, which provides some level of explicit structure. A typical web mining task is processing unstructured or semi-structured data from web pages.

### V. WEB PERSONALISATION

Website personalisation can be defined as the process of customising the content and structure of a website to the specific and individual needs of each user taking advantage of user's navigational behaviour. The overall process of usage-based web personalisation consists of 5 modules, which correspond to each step of the process:

#### A. User Profiling

In the web domain, user profiling is the process of gathering information specific to each visitor, either explicitly or implicitly. A user profile includes demographic information about the user such as interests and behaviour while browsing a website. This information is exploited in order to customise the content and structure of a website to the visitor's specific and individual needs.

#### B. Log Analysis and Web Usage Mining

This is the procedure where the information stored in the web server logs is processed by applying data mining techniques in order to:

- Extract statistical information and discover interesting usage patterns.
- Cluster users into groups according to their navigational behaviour.
- Discover potential correlation between web pages and user groups.

#### C. Content Management and Website Publishing

Content management is the process of classifying the content of a website in semantic categories in order to make information retrieval and presentation easier for the users.

Website publishing – A publishing mechanism is used in order to present the content stored locally in a web server

and/or some information retrieved from other web resources in a uniform way to the end user.

#### D. Information Acquisition and Searching

In many cases, information provided by a website is not physically stored in the website's server. In case of a web portal, users are interested in information from various web sources. It remains to the website editors to search the web for content of interest that should consequently be classified into thematic categories. Searching and relevance ranking techniques must be employed both in the process of acquisition of relevant information and in publishing of the appropriate data to each group user.

### VI. CONCLUSION

Data mining along with the internet have been the major platform for many businesses in the 21<sup>st</sup> century world. Any commodity or service can be purchased or sold within minutes, even from/to remote locations with the help of Internet marketing. Also, Narendra Modi's idea of Digital India has enabled more businesses online. Automated and machine learning has opened up new possibilities and a prospective approach to businesses rather than the traditional retrospective approach. [1]

### REFERENCES

- [1] Data Mining and Analytics: A Proactive Model - <http://www.ijarcce.com/upload/2017/february-17/IJARCCCE%20117.pdf>
- [2] Harasim, L. (Ed.) - Global Networks: Computers and International Networks, 1993
- [3] Protocol, Encyclopædia Britannica, retrieved 2012-09-24
- [4] <http://royal.pingdom.com/2012/04/19/world-internet-population-has-doubled-in-the-last-5-years/>
- [5] A Review on Basics in Web Mining, L.Shobby Neeta Fancy 1, A.Rajamurugan2, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 11, November 2014
- [6] Web Mining: Creating, Enhancing, Mining and Acting on Web Data, by Jesus Mena, DECEMBER 18, 2008

### BIOGRAPHIES



**VISHESH S** born on 13<sup>th</sup> June 1992, hails from Bangalore (Karnataka) and has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka in 2015. He also worked as an intern under Dr Shivananju BN, Department of Instrumentation, IISc, Bangalore. His research interests include Embedded Systems, Wireless Communication and Medical Electronics. He is also the Founder and Managing Director of the company Konigtronics Private Limited. He has guided over a hundred students/lecturers/interns/professionals in their research works and projects. He is also the co-author of many International Research Papers. Presently Konigtronics Private Limited has extended its services in

the field of Real Estate, Webpage Designing and Entrepreneurship.



**MANU SRINATH** hails from Bangalore (Karnataka); he has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka. His research interests include networking, image processing and cryptography. He is the Executive Officer at the company Konigtronics (OPC) Pvt. Ltd.



**SUPRIYA YADATI NARASIMHULU** hails from Bangalore, Karnataka. She has completed Bachelor of Engineering (B.E) in Information Science and Engineering from Visvesvaraya Technological University(VTU), Belgaum, Karnataka. Her research interests include - Data Sciences and Analytics, Image Processing, Artificial Intelligence and Machine Learning.



**SUPRIYA RAMARAO PRASANNA** hails from Bangalore, B.E in Information Science at BNMIT. Research interest includes image processing and data analysis.