

## Data Mining: Giving E-Tailers What They Want

Maurice Mulvenna, MINEit Software Limited, 5a, Edgewater Business Park, Edgewater Road, Belfast, BT3 9JQ, Northern Ireland, UK Tel: +44 (0) 28 90368875 Fax: +44 (0) 28 90366068 Email: maurice@mineit.com

Alex Büchner, MINEit Software Limited, 5a, Edgewater Business Park, Edgewater Road, Belfast, BT3 9JQ, Northern Ireland, UK Tel: +44 (0) 28 90368394 Fax: +44 (0) 28 90366068 Email: alex@mineit.com

Louis Moussy, @t Internet, Centre Lafayette, Av. Kennedy, 85, France, Tel: +33-5-5792 3434 Fax: +33-5-5634 1763 Email: lmoussy@dial.oleane.com

Dave Slater, AdNet Ltd, 14 Merrion Square, Dublin 2, Ireland, Tel: +353 (0) 1 6764262 Fax: +353 (0) 1 6764268 Email: dslater@adnet.ie

### Abstract

This paper looks at data mining applied to Internet retailing from the perspective of a user partner in the MIMIC research project. It discusses the methods used to identify user requirements, and outlines these user requirements. It also examines the nature of the business process within which data mining should be implemented. In particular, the paper focuses upon the relationship between Internet retailers and service providers who provide the retailers with specialist services. It also provides an overview of the main results to date of the MIMIC project.

## 1 Introduction

The paper introduces the work carried out under the ESPRIT project 'Mining the Internet for Marketing Intelligence' - MIMIC<sup>1</sup> [1]; in particular the user requirements capture and associated business processes. As described in more detail in the section below, MIMIC is specifically concerned with the analysis of data and information relating to Internet retailing – although it may be assumed that many of the techniques and methods identified and developed will be applicable in business-to-business and other areas of Internet operation.

The reader should be made aware at the outset that this paper is written from the perspective of a non-technical potential user of data mining (and related techniques) in the area of Internet-based retailing – e-tailing [2]. Both Adnet and @t Internet are what is now commonly known as Internet design and hosting companies, or service providers. Service providers make available the facilities whereby e-tailers can initiate, develop and - critically in the context of the MIMIC project - monitor and assess their Internet sales presence. Adnet and @t Internet are user partners in MIMIC, and are responsible for user requirements identification and elaboration in the project. Readers interested in delving into the more technical aspects of MIMIC should refer to [3-6].

---

<sup>1</sup> MIMIC acknowledges funding by the EU Commission

## 2 Internet Retailing Business Model

The majority of e-tailers enter the Internet marketplace by contracting a service provider, or other agent offering similar services. In the vast majority of cases, the retailers' primary area of expertise is retailing, rather than any technical or information technology area. For them, the Internet is usually another sales vehicle, which will operate in parallel with their conventional outlets. However, it is normally one with which they are less familiar and may sometimes even be perceived by them as a threat. This mindset is slowly changing, as the Internet becomes more widely used and understood. However, in the experience of the MIMIC user partners, such change is occurring more slowly than many Internet commentators suggest and most retailers still approach the Internet with a certain degree of trepidation. Exceptions include large companies operating in high technology sectors, such as computer equipment suppliers, and companies who are specifically and solely Internet retailers (i.e., they have no conventional retailing operations). These organisations often have developed the skills required to implement relatively complex technical solutions in-house, or Internet retailing is so core to their business activities, and contributes so substantially to their revenue streams, as to make investment in acquiring those skills financially justifiable.

The prospective e-tailers will initially discuss their business objectives and expectations, and develop a 'road map' for their migration to Internet retail. Human resources and technology needs will be identified, the site will be developed and promotion and development strategies drawn up. On an ongoing basis, and on foot of sales, traffic and other feedback data, these strategies will be refined. The service provider usually offers a range of services such as graphic and Internet specific design skills, software development and business advice. Often, the e-tailer becomes dependent upon the provider, and relies on the service provider for specialist market intelligence and support regarding their Internet retailing site on a long-term basis.

Data mining, then, would become one of a wider range of value-adding services offered by the service provider to the e-tailer. It is within the context of this broader client-specialist business model that the user requirements and specific business model described below should be seen. Thus the Internet business model nearest to data mining services is, according to Timmers [7], the *Value Chain Integrator*.

## 3 Background to the MIMIC Project

MIMIC is a project funded by ESPRIT IV Framework Thematic Programme for Electronic Commerce. The MIMIC Consortium consists of four partners, represented by the University of Ulster, and three small-to-medium enterprises (SMEs). One of these, formerly Integral Solutions Limited and now owned by SPSS, is a specialist in data mining and produces the 'Clementine' data mining tool. The other two, Adnet and @t Internet are Internet service providers as outlined in the introduction above.

The technical objective of the MIMIC project is to develop a data mining toolkit that will make possible the mining of the data generated by e-tailing companies [8]. The aim is to develop those algorithms and other analytical tools for application in a data warehouse containing Internet server log data, product, customer and transaction data. This data can then be mined to provide high level marketing intelligence for the Internet retailer – within the context of the business model described above. This marketing intelligence may be channelled for use by 1) the Internet retail site; 2) the technical staff in charge of the site;

and 3) marketing experts seeking the knowledge necessary in order to provide strategic marketing intelligence.

With these project objectives in mind, MIMIC initially faced a number of core questions to address over its project life cycle. These were:

1. What data is normally available to an e-tailer? How greatly does this data vary from site to site?
2. Is that data adequate to allow for credible data mining? What measures must be taken to ensure adequacy where it is not?
3. How can the user requirements best be identified?
4. What are those user requirements?
5. What is the business process model within which data mining will take place?
6. When data mining is undertaken, are the results actually useful to e-tailers?
7. Can concrete action be taken on the basis of those results, and by what means?
8. Even where results are useful and actionable, to what extent do they meet the fuller range of user requirements identified? What gaps remain, and what future interventions are needed?

Within the MIMIC project, we have now addressed – or are in the process of addressing - all but the last two of these questions. The first two, which of a more technical nature, are explored in detail in various MIMIC documents, deliverables and through the concrete output of MIMIC activities. In this paper, we will now briefly discuss the more non-technical, user-orientated, questions that the project has addressed to date.

### **3.1 Current Context of MIMIC**

At this time, a large number of companies world wide are developing e-CRM (electronic Customer Relationship Management) software, which is intended to provide multi-channel CRM support across the multiple customer touch points, i.e., call centres, Internet, email, telephone, mail. MIMIC is targeted at provision of *visitor* relationship management, and the maximisation of the rate of conversion of visitors to customers. As such, the philosophy behind MIMIC is still widely applicable. Although the current applied artificial intelligence focus of e-CRM is on the use of data mining for *customer* segmentation and profiling, the authors believe that it is in the provision of marketing intelligence for *visitor* information through the application of data mining – or web intelligence – that results in major benefits for e-tailers. Consider that it cost much more to acquire new customers than to cross- or up-sell to existing customers. Clearly, any technology that assists in making visitors more responsive to retail offers etc, through personalisation supported by web intelligence, then there must be a substantial market in e-tailing for this technology.

The technology developed in MIMIC is licenced by MINEit to SPSS in the CapRI algorithm, and is embedded in the Easyminer product from MINEit.

## 4 User Requirements Capture Process

e-tailing is relatively new – and especially for many of the retailers themselves. While data mining is a comparatively established technique, it has yet to be applied successfully within Internet retailing, at least in the context of the overall MIMIC objectives. Therefore, when approaching the task of identifying potential user requirements – especially those that are technically feasible and of real benefit to e-tailers – we were entering a largely unexplored field [9].

In order to simplify our task, we separated user requirements into two broad categories:

- ‘Strategy Led’ user requirements. Generally, these arise from a strategic or business goal that can be identified among the constituency of e-tailers. They may also arise as responses to common types of problems encountered when migrating to or operating via the Internet. Examples might include combating competition, diversifying the customer base or internationalising. When examined more closely and de-constructed into core and supporting functions and tasks, ways in which data mining can assist were developed. The user requirement, however, ultimately originates from and is driven by the strategic objective identified by the user.
- ‘Opportunity Led’ user requirements. New functions, services or innovations are identified which may be of relevance or useful to e-tailers. Through their implementation, they lead to a commercial or business advantage for users. However, they are ultimately led by the innovation itself.

In general, the capture of requirements from a ‘strategic led’ perspective is more robust and accurate in the longer-term than requirements derived from an ‘opportunity led’ perspective, which is more sensitive to technological change. It is also a truer reflection of business reality to suggest that strategic goals dictate functional aspects of an organisation – except perhaps during the very initial start-up phase and especially, it must be recognised, in the high-technology sector.

On the basis of these classifications, a two pronged approach to user requirements identification was adopted. First, a rigorous review of current literature and activity was undertaken in order to, as fully as possible, identify ‘opportunity led’ requirements and the innovations, processes or trends driving them. On the basis of this, an initial set of requirements was drawn up.

Secondly, a number of in-depth interviews were conducted with e-tailers and others with informed opinion on Internet retailing. These interviews were used mainly to elucidate the more important, and common, strategy led user requirements.

Additionally, once strategic issues had been discussed, those ‘opportunity led’ requirements identified through desk and other research were explained to the users and their relevance in their business context validated. This was more difficult for a number of reasons. Experience of interviewing retailers currently using the Internet has shown that while they are very familiar with their own sector, its products and their traditional customer base, they are poorly equipped to fully understand the concept of data mining or to visualise concretely its use in their operations. In the main, they appear to be seeking advice, from their service provider or other Internet consultant, as to the development and implementation of their site and its functions rather than driving change themselves.

## 5 User Requirements Identified

As we shall see below, the business process facilitating data mining is not entirely simple. It may well vary from retailer to retailer, and will form one part of the overall business relationship between e-tailers and their service providers. Those providers who will be ones using any data mining tools developed, will have their own very specific set of requirements, many relating to the technical functions and characteristics, and ancillary supports, relating to those tools. It is interesting to note that many providers envisage using data mining tools, that analyse behavioural patterns of site visitors, to add value and generate continuing business in site redesign and optimisation.

Here, however, we are discussing purely the high-level strategic requirements of the e-tailers themselves. These requirements are under constant review in the project, and it will be particularly interesting to see how they change, develop or are refined as we begin to work with real users when equipped with a prototype.

It may be apposite to summarise, at this point, that most of the user requirements expressed by e-tailers refer to the provision of the type of information available to them in their conventional operations, which are key to many of their operational decisions and overall understanding of their business. Who comes into their shop, when, what type of people are they? What do they do when they are in the shop, where do visitors go or not go? What type of people buy what products? What ongoing changes can be observed? This is the type of information, previously taken for granted, which allows them to apply their knowledge and experience of retailing successfully. Much of this is presently unavailable when retailing on the Internet. By providing such information, data mining should provide a key business and competitive advantage to e-tailers.

The core user requirements [10], then, are to be able to:

- Identify clusters of user types. This would seem as the most important use of data mining, as this type of information is useful in a myriad of other planning and development tasks – and in fact forms a key component of many of the further areas of application outlined here. Retailers are keenly aware that this type of information is key to success. They desperately need to know who is using their web site. Even large, high-technology companies often have no accurate idea as to who is using their site.
- Assist in understanding and analysing the behaviour of users on their site. This includes issues such as path followed, differences in behaviour between non-buying and buying visitors and between members of the different clusters identified, particular barriers that users seem to have in using the system etc. This type of information is presently even more difficult for e-tailers (and webmasters generally) to obtain than, for example, customer cluster analyses.
- Provide input to the product selection and development process. This may require, for example, analysis of links between the clusters identified and products bought etc. Such a facility was particularly highly regarded by retailers interviewed.
- Assist in developing the look and feel of the site, by being able to perform controlled tests when changes are introduced.
- Provide input to the process of promotion, via the Internet and by other means, of their site. This could entail advertising, direct marketing or other interventions.

These are the direct objectives when applying data mining to Internet data – though of course they normally form a key element of a larger process that fulfills a broader business objective.

## 6 The Business Process

It is useful first to describe the peculiarities of data mining when applied within an e-tailing organisation toward meeting some or all of the objectives outlined in the section above. Usually, the data mining model implicitly presented in the existing literature is one where data mining is used by an organisation whose *primary function is not necessarily the provision of information or where the generation and use of information is not a core business activity* [11]. Information regarding their tasks and activities is generated and stored, and then mined in order that non-intuitive patterns or secondary information can be derived.

The data mining process provides an input to production and planning decisions, but is separate from the actual production process itself. It is not data which is being produced (essentially a web-site is a data generating engine, including pages presented to the customer, their activity on the site, sales data or server or other log data) but an ancillary service.

Therefore, data mining is a discreet task outside what management may well consider core business, and certainly production, activity – although it may well have an important contribution to make. This leads to the linear model of a business process, which is comparable to most forms of consultancy activity. Under these circumstances, data mining is used to provide businesses with useful *information* that is of relevance to and use in their core business activity. However, in the realm of e-commerce data mining is an important component in a suite of specialist services that provides the end-user with a core business *function* which is integrated into other activity.

For the e-tailer, data mining and other specialist services form an integral part of the ongoing business process whereby they ensure that their site remains successful and responsive to user and customer needs. This is illustrated in Figure 1 below.

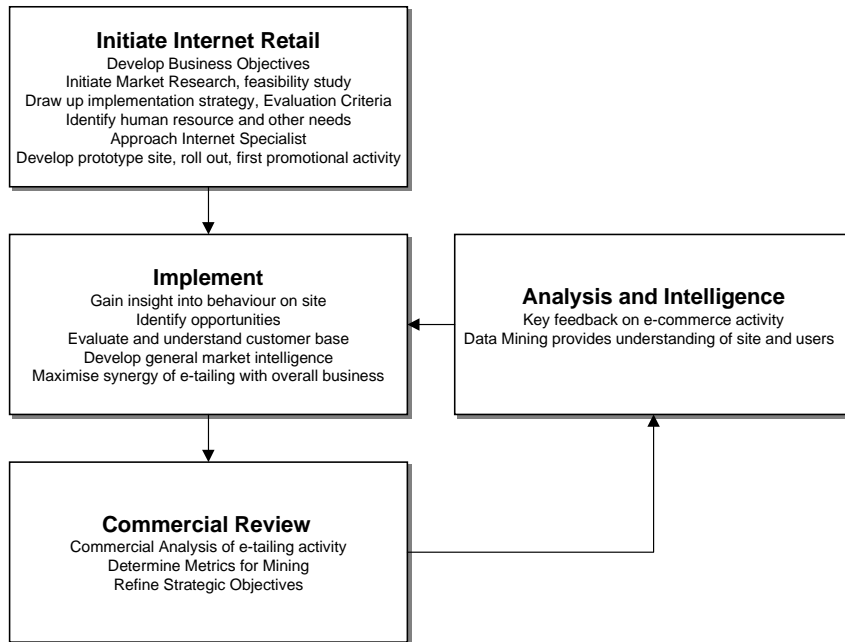


Figure 1. Data Mining as a Core Business Function

There is a complex interplay between the service provider and the e-tailer. The provider gives key inputs in all stages, with the possible exception of financial and commercial review of retail operations. An organisation offering only data mining services is unlikely to be an attractive option for the vast majority of e-tailers – except perhaps very large ones who manage all other technical aspects themselves – although they may obtain market share in providing such services to the service provider.

The core service requirements of the e-tailer, with regard specifically to data mining, are:

- Consultancy in order to determine their data mining needs, and to plan how the results of data mining can be implemented concretely in order to improve their Internet site (using an agreed metric: sale volume, value, etc);
- Mining of the data. This will be done in conjunction with the consultant. The e-tailer is heavily involved in this task, as their knowledge of their sector and organisation is crucial;
- Assistance in interpreting the results of the analysis and in implementing any changes or the addition of new functions to the e-tailing site (e.g., personalisation and/or customisation);

## 7 Results

The MIMIC project is nearing completion, and a number of core achievements have been made – both technical and non-technical, and first impressions of eventual usefulness can be given. These include:

- Internet-specific data mining algorithms have been seamlessly integrated into the data-mining tool (Clementine). The core functional platform required is now available.

Crucially in terms of meeting some of the main user requirements, a sequence analyser (CaprI) has been developed;

- Some of the core technology is being exploited for the University of Ulster by its spin-off, MINEit Software, and the Easyminer product;
- It has been demonstrated that raw web log data can be transformed into coherent data representations used by new developed data mining technologies. Detailed analyses have been performed on such data. In short, we have shown that it is possible to mine Internet retail data.
- It has been shown that pre-processed data is even more useful than raw data, and pre-processing techniques have been developed.
- Visualisation methods, including an innovative method for visualising Internet activity sequences, have been agreed between technology suppliers and users.

In terms of the user requirements' collection process, the approach adopted – that of interviewing key people in the Internet retail and related fields was found to be generally useful and successful. However, it is important to bear in mind that the approach adopted in MIMIC did not constitute a survey, with a representative sample or standard response fields. Such an approach was beyond the resources available to MIMIC, and may in any case be inappropriate in the context of determining user requirements where the product is based on a relatively innovative, and therefore unfamiliar, concept.

The next steps in MIMIC are clear. Feedback from users of CaprI and MINEit's Easyminer is being used to develop a coherent architecture for the application of data mining. This was one of the key objectives of MIMIC. A schematic of the components is shown in Figure 2, and updated information can be found at [www.mineit.com](http://www.mineit.com).

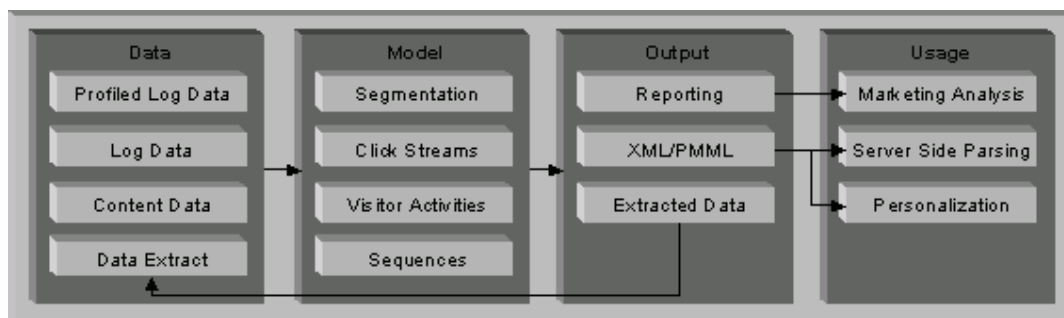


Figure 2. Schematic for Marketing Intelligence (MIMIC)

To date, MIMIC has overcome most of the main hurdles and obstacles that were anticipated or feared when the project commenced. On the technology side, substantial research has developed successful new mining technologies. On the user side, it is equally clear that data mining has a role to play, and an important contribution to make, in e-tailing. In terms of exploitation of results, a successful new spin-off company, MINEit, has been created. The remaining barrier to success lies in making the results of the mining process usable to web servers, thus providing automated marketing functions such as personalisation [12].

## References



- [1] EU Commission ESPRIT, Project No. 26749, Mining the Internet for Marketing Intelligence (MIMIC).
- [2] Markham, J.E., *The Future of Shopping: Traditional Patterns and Net Effects*, MacMillan Business, London, 1998.
- [3] Büchner, A.G., Mulvenna, M.D. Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining, *ACM SIGMOD Record*, ISSN 0163-5808, 27(4): 54-61, 1998.
- [4] Mulvenna, M.D., Norwood, M.T. & Büchner, A.G. Data-driven Marketing, *Int'l Journal of Electronic Commerce and Business Media*, 8(3):32-35, 1998.
- [5] Büchner, A.G., Mulvenna, M.D., Anand, S.S. & Hughes, J.G. An Internet-enabled Knowledge Discovery Process, *Proc. 9th Int'l. Database Conf.*, pp. 13-27, 1999.
- [6] Baumgarten, M., Büchner, A.G., Anand, S.S., Mulvenna, M.D., Hughes, J.G., *Navigation Pattern Discovery from Internet Data*, In: Masand, B., Spiliopoulou, M. (eds.) *Advances in Web Usage Analysis and User Profiling*, *Lecturer Notes in Computer Science*, Springer-Verlag, 2000.
- [7] Timmers, P., "Business Models for Electronic Markets", *International Journal of Electronic Markets*, 8(2), <http://www.electronicmarkets.com>
- [8] Mulvenna, M.D., Büchner, A.G., Norwood, M.T., Grant, C., "The 'Soft-Push': Mining Internet Data for Marketing Intelligence", In: *Working Conference: Electronic Commerce in the Framework of Mediterranean Countries Development*, Ioannina, Greece, October 1997.
- [9] Adnet, *State-of-the-Art in Internet Retailing*, MIMIC Deliverable 2.01, June 1998.
- [10] Adnet & At Internet, *MIMIC User Requirements*, MIMIC Deliverable 2.02, September 1998.
- [11] Mulvenna, M.D., Büchner, A.G., Hughes, J.G., "Re-engineering Business Processes to Facilitate Data Mining", *Proceedings 1st International Conference on Practical Aspects of Knowledge Management*, Basel, Switzerland, Vol. 1, 1996.
- [12] Mulvenna, M.D., Anand, S.S., Büchner, A.G., Böhm, R., Baumgarten, M., "Applying Web Intelligence", *International Conference on Electronic Commerce*, Seoul, Korea, 2000, Submitted, (Pre-print available upon request).

**Keywords:**

Data Mining, Marketing Intelligence, E-tailers, User Requirements

### **Biographies & Photographs:**

Maurice Mulvenna, BSc, M.Phil., PGCUT, is the Managing Director of MINEit. He has over 9 years experience working on and managing projects in the areas of artificial intelligence, electronic commerce and database systems. He was the manager of NIKEL, the University of Ulster's technology transfer facility for Informatics, from 1993-97. This post involved the management of many local, national and international projects, involving research staff, and budgets. He is on the editorial board of the International Journal of e-Business Strategy Management; a member of the editorial review board of the Web Net Journal; a current or past committee member of the International Programme Committees for EMMSEC, PAIS-2000 at ECAI and Web-KDD '00.

Alex Büchner MSc. is the Product Development Director of MINEit. He has 4 years experience in Data Mining technologies and has been involved in several related projects in Europe and the Far East. He has more than 10 years experience in developing procedural and object-oriented software, and has extensive experience in developing database applications on various platforms. He has written more than thirty international research publications and has co-authored a data-mining book, published by Financial Times Pitman Publishers. He is a committee member of Web KDD '00.

Louis Moussy brings first-hand experience in practical electronic commerce development and application. In 1997 he joined the Electronic Commerce Europe association, and was elected a member of the Board, and in 1998 General Secretary. Already in 1995 Louis developed, in association with ATInternet, a start-up ISP, one of the first-ever electronic commerce malls in Europe, selling Bordeaux wines and gourmet products to the world at large. This followed the successful launch, in 1987 of his own consultancy company, LanceXport, which specialises in marketing and sales support to companies engaged in the import/export of high-tech products and systems. Previously, Louis spent 10 years with IBM in engineering and 20 years with Thomson, where he was head of unit in the marketing and sale of electronic high-tech products and systems to the Space and Scientific Communities.

Dave Slater is a freelance writer and researcher based in Dublin, Ireland. He works in a number of fields, including the Internet and broader science and technology issues. His work on the MIMIC project has been undertaken for Adnet Ltd, an Internet services company based in Ireland.

