

ORIGINAL RESEARCH

A new web based data mining exploration and reporting tool for decision makers

Ahmet Selman Bozkir, Ebru Akcapinar Sezer

Department of Computer Engineering, Hacettepe University, Ankara, Turkey

Correspondence: Ahmet Selman Bozkir. Address: Hacettepe University, Department of Computer Engineering, 06800, Ankara, Turkey. Email: selman@cs.hacettepe.edu.tr.

Received: December 14, 2012

Accepted: March 28, 2013

Online Published: May 5, 2013

DOI: 10.5430/air.v2n3p70

URL: <http://dx.doi.org/10.5430/air.v2n3p70>

Abstract

The current DSS tools are generally built as “desktop applications” and designed for the use of data mining experts. In this paper, design and implementation of ASMINER, a new web-based data mining exploration and reporting tool, is introduced. ASMINER enables both decision makers and also knowledge workers, exploring and reporting with three data mining techniques (decision trees, clustering and association rules mining), by presenting a scalable, user-friendly and fully web-based thin client data mining tool. The approach and the tools of ASMINER have a significant potential at giving an opportunity to knowledge workers to participate in decision making phases.

Key words

DSS, Web-based data mining exploration, Decision trees, Clustering, Association rules

1 Introduction

During the last decade, use of data mining in organizations has expanded significantly since the data mining methods can be accepted commonly as a useful tool for decision-makers in predicting the future during the decision processes. This role of data mining makes it as the oracle of 21st century. Actually, data mining approaches generally take place in decision support systems as a part of business intelligence. Additionally, the data mining approaches have a great potential to solve forecasting problems encountered in all engineering fields, medical and applied sciences, etc. For this reason, it is not surprise that the expansion trend of the use of data mining approaches is continuing.

Marakas ^[1] gives the definition of decision support systems as “the system which is under the control of one of more decision makers that assists in the activity of decision making by providing and organized set of tools intended to impose structure on portions of the decision making situation and to improve the ultimate effectiveness of the decision outcome”. On the other hand web based technologies have been revolutionized the design, development and implementation stages of decision support systems ^[2,3]. Due to its nature, web based systems are distributed, multi-user oriented and platform independent. Nonetheless, Web has the chance to reduce the problems related hardware and software issues (e.g. versioning) ^[4]. For these reasons, Shim et al. consider the Web as a very important DSS development and delivery platform ^[5]. In the light of these consequences, it can be stated that Web is a brilliant platform for DSS usage and development. At this point, Web based DSSs and related technologies come into prominence. Although, according to Xie et al. ^[6], there is no general accepted description of Web-based DSS, Power ^[7] describes Web-based DSS in a wide

perspective by stating that the ideas which decision support systems services on the Web. Likewise, some researchers define it as the DSS implemented on Web such that decision makers use client browsers like Internet Explorer ^[62] for decision making operations ^[6]. On the other hand, this advance caused some important changes in DSS field which can be grouped in three aspects: user community, problem domain and technology underlying in it ^[8]. While DSS tools were luxury tools for managers and analysts in the past, current users of DSSs vary from product buyers to bosses now. Besides, according to Power ^[2], Web-based tools have begun to change the way of knowledge workers for data and information access, data analysis and retrieval of information from the decision support systems. Desktop or batch based DSS usage in past has gradually turned to rapid and platform independent Web based DSS usage. The key advantages of the web based tools when compared with the traditional batch-based or client-server approaches include ease of-use, universal access across different platforms and single minute response over dynamic and real-time data ^[9]. Furthermore, Web-based DSS has reduced technological barriers and let managers and staffs make benefit from easier and low cost decision making even in geographically distributed locations ^[5].

There have been different DSS usage profiles for various reasons. According to Alter ^[10] four DSS usage modes exist: (1) subscription mode, (2) terminal mode, (3) clerk mode and (4) intermediary mode lastly. In subscription mode, users of DSS get mails or reports periodically. In clerk mode, the arguments and requests are prepared off-line and submitted to system by using electronic batch processes ^[11]. Decision makers interact with DSS in an online, synchronous and real-time manner in only terminal mode. In other modes, there is a barrier between DSS and decision maker. Especially in intermediary mode, there exist one or more intermediary users between DSS and decision maker ^[11]. Although various benefits and limitations exist for all of these usage patterns, Marakas ^[1] suggests terminal mode by emphasizing the importance of speed and time consumed in decision making processes. Furthermore, Marakas ^[1] claims that tomorrow's decision makers must be skilled to interact with DSS directly for not to negatively affect from prohibitive effects of intermediary users.

By definition, data mining is the extraction information from raw data by utilizing machine learning, artificial intelligence, statistics and related methods in automatic or semi-automatic ways. In other words, data mining is a journey from raw data to knowledge. Although data mining and related methods are continuously expanding in many fields, it is still difficult for typical users including managers to learn and use; in addition, data mining solutions need significant theoretical background to work with. Data mining technologies and related applications are generally focused on data mining specialists. Romero et al. ^[11] stated that data mining tools are designed more for power and flexibility than for simplicity. Furthermore, a typical data mining based report generation scenario requires considerable much time involving lots of iterations between decision-makers and data mining specialists. These iterations slow down the decision making processes. However, a simple and user friendly web-based data mining tool can reduce the amount of time need for decision making phase and increase the productivity of decision makers since it requires only a web browser and general theoretical background of data mining. Especially for the firms, web-based data mining tools are becoming a backbone for managing the knowledge to gain a competitive advantage ^[9]. According to Power ^[12], the use of DSS tools and techniques in making critical decisions should be largely increased with the help of well-designed DSS's.

The international literature about the decision support systems contains a considerable amount of studies related to Web-based DSS in both theoretic and practical point of views. Bharati and Chaudhury ^[8] investigated the decision making satisfaction in web-based DSSs. Shim et al. ^[5] and Bhargava et al. ^[13] surveyed the status of web-based decision support systems. Heinrichs and Lim ^[9] evaluated the applicability of web-based data mining tools to the current business models. Likewise, Duan et al. ^[14] pointed out the lack of research in web based expert systems and argued the benefits and challenges of web based expert systems. In recent years, data mining and web-based decision support systems have been used in various areas such as healthcare industry ^[15, 16] manufacturing and marketing ^[4, 17-20], human resources ^[21], education ^[11] and network security ^[22]. Tsoumakas and Vlahavas ^[23] proposed a web application namely "CSF/DC" which enables the sharing of classification models on the web. Jorge et al. ^[24] developed a methodology for exploring large set of association rules on web pages. Lee and Chung ^[25] presented the discrimination between groups of DSSs and showed how

web based DSSs can be efficient in internet shopping by integrating them with virtual reality environments despite of classical visualization techniques. Similarly, Chen et al. ^[26] emphasized the importance and the efficiencies of group decision support systems and presented a tool called TeamSpirit for this purpose. Besides, Delen et al. ^[4] developed a data mining oriented and a domain specific web based DSS whose aim is predicting total income for a specific movie and helping Hollywood movie managers in giving decisions. Nonetheless, Cameron ^[27] proposed a prototype of an interactive web based DSS, which has potential usage in educative purposes for hazardous land-use planning. Cameron's study has also potential use in educative purposes. Additionally, Liu ^[28] presented a web based data mining decision support system for liability authentication of traffic crashes as another domain specific DSS application. However, the tools or applications given above are generally domain specific types of web based DSSs and actually do not present a general framework for non domain specific decision support purposes.

The purpose of the study is to save time in the exploration and reporting phases of data mining based decision making processes by developing a completely web based data mining exploration and reporting tool called ASMINER. Indeed, the name ASMINER is not an abbreviation and it constitutes the full name of the work. In this study, building a platform independent, user-friendly and easy to use data mining tool is targeted. With the developed tool, transforming the users of all other DSS usage modes to terminal mode is also aimed in the perspective of data mining centric DSS usage. ASMINER utilizes Microsoft SQL Server Analysis Services at behind the scene as the core data mining engine and it currently supports three data mining techniques such as decision trees, clustering and association rules. It presents web based model to explore and view powered by rich and easy to understand data visualization tools. Furthermore, it allows making basic classification and regression operations over decision tree models via its built-in web-based prediction tool. Additionally, in the clustering subsystem of ASMINER, a novel distribution based cluster dominancy exploration method is served. As a summary, ASMINER is a contributive tool for web based data mining oriented decision support systems because it changes the traditional decision making process. However readers should not confuse ASMINER with some visual data mining tools (e.g. Rapid Miner ^[64], SPSS Clementine ^[38]). The main goal of ASMINER is not to present a complete easy to use visual data mining environment which can be found at those kinds of products. Instead, it aims to separate the roles of data mining experts and the knowledge workers and rescue the knowledge workers and decision makers from dealing with hard to learn data mining theoretical background. In ASMINER approach, data mining experts load and manipulate the data on server. Nevertheless, decision makers deal with only model exploration, prediction and some limited model training options. Therefore, we put a seamless barrier between them and define their responsibilities. Furthermore, this approach is working on a platform independent web based environment.

On the other hand, the short version of this study had been presented ^[63]. This paper extends that study by explaining its features and discussing the underlying philosophy. Furthermore, comparison of the state-of-art data mining products with ASMINER is presented.

The rest of this paper includes six main sections such as a short overview on data mining, technologies used in system, the design requirements and architecture of ASMINER, the subsystems involved in ASMINER, use of the system with a real scenario, and conclusion.

2 A short overview on data mining and methodologies involved in ASMINER

As a simple definition, data mining is the process of automatically discovering useful information in large data repositories ^[29]. In past, data mining has been referred to as knowledge management or knowledge engineering. Recently, it has been an obscure an exotic technology which is discussed more theoreticians in the artificial intelligence fields ^[17]. Fayyad et al. ^[30] defined the data mining as a step in the KDD (Knowledge Discovery in Databases) process consisting of applying computational techniques that, under acceptable computational efficiency limitations, produce a particular enumeration of

patterns or models over the data^[17, 31]. It is evident that data mining tasks can be categorized into two groups such as predictive tasks and descriptive tasks. While the objective of predictive tasks is to forecast the value of an independent attribute based on the other attributes, the objective of descriptive tasks is to derive patterns (e.g. correlations) that summarizing the underlying relationships in data^[29]. Classification and regression techniques are take place in predictive methodologies. However, clustering, association rules mining, sequence analysis and summarization techniques are the instruments of descriptive methodologies.

Decision tree is a data mining approach that is often used for classification and prediction. Although other methodologies such as neural network and fuzzy c-means clustering etc. can also be used for classification, decision tree has the advantages of easy interpretation and understanding for the decision makers to compare with their domain knowledge for validation and justify their decisions^[21]. Therefore, it is selected as the classification method for ASMINER. The fundamental idea behind a decision tree is to split data recursively into subsets so that each subset contains more or less homogeneous states of selected target variable (predictable attribute). At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute^[32]. This stage is called as decision tree induction. Several algorithms such as ID3, CART^[33], CHAID^[34] and C4.5^[35] have been developed for decision tree induction^[29]. All of these algorithms use different mathematical approaches (e.g., Gini index and information gain) in construction of decision trees. According to Chien and Chan^[21], these algorithms can be distinguished in terms of different measurements, pruning methods, missing value decomposition, the number of branches at each node and the data type they could handle. Clustering is finding natural groups inside data when these groups are not obvious. In addition, Chiu et al.^[17] defined clustering as an exploratory method for helping to solve classification problems. Thus, clustering is also called unsupervised classification, where no predefined classes are assigned^[36]. According to Chiu et al.^[17], a good clustering method will produce high-quality clusters with high intraclass similarity low interclass similarity.

Association rule discovery is the process of unrevealing the unobservable and hidden relations among the cases in data. However, association rules mining is generally used for market basket analysis from the early days of its discovery. Likewise, it discovers relationships among attributes in databases, by generated if-then statements concerning attribute-values^[37]. Moreover, association algorithm is nothing more than a correlation counting engine^[32]. An $X \Rightarrow Y$ association rule expresses a close correlation between items (attribute-value) in a database with values of support and confidence. The confidence of the rule is the percentage of transactions that contains the consequence in transactions that contain the antecedent. The support of the rule is the percentage of transactions that contains both antecedent and consequence in all transactions in the database^[11]. Although there are many other types of data mining methodologies, these three types are briefly introduced due to reason that ASMINER currently supports only them.

3 ASMINER: A new web-based data mining tool

3.1 Problem description

Several software vendors and publications, such as Datamation (<http://www.datamation.com/dataw/>) affirm that all knowledge workers will become data miners in the future^[5]. However, there are numerous numbers of data mining tools and applications in the market that require professional data mining background and practice. Likewise, this makes data mining solutions as the software packages for only experts. Second, most of all commercial data mining solutions are implemented with non web-based approaches. (e.g., SPSS Clementine^[38], Microsoft Analysis Services^[39], SAS Enterprise Miner^[40]). In other words, although those applications are created to be used in worldwide companies, they are not designed for geographically distributed decision making activities so far. Also these types of tools need significant data mining theory and practical experience to be operated. These reasons complicate decision making processes and create a bottleneck of knowledge workers and decision makers to operate DSS directly. In every cycle of report generation all the participants of the process should be involved in the scenario. Therefore huge amount of time is consumed even for a single report building. Moreover, current data mining tools does not contribute to this problem and they still require

exhaustive and time-consuming processes. Although, these tools offer rich options, majority of the decision makers generally need standard and quickly buildable reports. On the other hand, many companies have various numbers of offices in distributed locations. With non-web based solutions collective decision making requires much more effort due to the geographic distances.

However, difficulties of those users in this field can be minimized by offering a basic data mining tool including only the most required capabilities. For these reasons, ASMINER targets the knowledge workers and decision makers to help them in the process of becoming data miners and to achieve this, it presents easy to understand, user friendly and perspicuous user interfaces in exploring three types of mining models. In addition, it is aimed to reduce the time required in report generation processes and gain speed in the model re-building phases. Another useful point is that migration to ASMINER can be easily carried out due to the high penetration of Microsoft SQL Server product (<http://mediaproducts.gartner.com/reprints/microsoft/vol3/article7/article7.html>).

3.2 General solution

After the statement of the problem and according to requirements analyzes, it is believed that ASMINER should be a purely web-based application. On the other hand, for scalability, fastness, reliability and integration to other data warehouses and third party algorithms, Microsoft SQL Server Analysis Services is located at the backend of the system. At implementation, ASP.NET 2.0 is employed. The main reason for choice of ASP.NET is that, it has a native support for using Analysis Services API. Nonetheless, it is a popular web based application development platform with reliable and rich components serving practical graphical user interface objects. Besides, Internet Information Services 7 employed as the server infrastructure. As a visualization requirement, Adobe Flash, Microsoft Silverlight and Sun Java technologies are employed in client side. Thus, in client mediums existence of these packages is essential. However, ASMINER is not only planned to use these packages, also it has been foresaw to have pure HTML controls and visualization components. Currently, HTML5 controls are being developed to replace the visualization components and make ASMINER full platform independent. However, currently ASMINER is compliant with Mozilla Firefox, Chrome and Internet Explorer browsers.

3.2.1 Design principles

In their empirical study, Bharati and Chaudhury^[8] investigated the factors of affecting satisfaction in decision-making period and evaluated this subject in three aspects: system quality, information quality and information presentation. In analyze and design stages of this study, the findings^[8] are highly considered. System quality covers system reliability, ease of use, flexibility and convenient to access which all reflect the overall performance^[41]. Likewise, information quality refers to accuracy, completeness, relevance and timeliness^[41-43]. Finally, Information presentation covers efficient navigation and appropriate color, graphics and style of presentation^[8, 44, 45]. In Bharati and Chaudhury's work^[8], it is clearly revealed and stated that system quality and information quality are the most important and effective factors in efficient decision making and the satisfaction in decision making periods. On the other hand, it is pointed out that quality of information presentation is not as important as the system and information quality. Therefore, at designing and building process of ASMINER performance, security and reliability issues are mostly undertaken. As a consequence of this decision, instead of giving much effort to create new non-standard visualization tools, existing and well used data mining model viewers of Analysis Services are migrated and rebuilt again for web environment.

In order to develop an easy to use, user friendly and powerful web based data mining exploration and reporting tool, the following characteristics had been taken into consideration

Performance: To overcome legendary bandwidth problem of internet based applications two approaches are undertaken in design of ASMINER. First of all, amount of data transported was reduced by using efficient web components and AJAX (Asynchronous JavaScript and XML) architecture. This leads fewer page refresh and reduces the amount of data transferring over network. Second, small sized interactive visualization components which were implemented in Flash and

Java were employed when required. Due to fact that, these tools are cached in browser, only the model data is sent to client. This approach avoids us, re-loading of visualization components in multiple times.

Accessibility: As stated before, to create a data mining platform aim of gathering geographically distributed decision makers, a web based approach has been chosen. Furthermore, browser compatibility was noted. Due to the rapid development of smart phones and devices (e.g. Tablet PC) the client side of the scripting was re-coded regarding the latest HTML5 compliant mobile browsers like Dolphin Browser^[46]. On the other hand the Java and Flash technologies which ASMINER visualization components based-on, are worldwide and standard technologies which almost all browsers have support for. As a conclusion a highly accessible and platform independent tool was tried to be provided.

Scalability: In today's business environment, as the expected amount of data may increase to terabytes, a system to be developed must be capable of handling such large volumes of data. To be scalable and fast enough, Microsoft Analysis Services is employed as the core data mining engine. Analysis Services has the capability of handling many requests from clients and even processing data mining models at the same time on server side by benefiting from multi core processors and even cluster computing. Therefore ASMINER is guaranteeing to serve even in highly task-loaded times. With this approach, ODBC connectable databases can be used through Microsoft SQL Server OLTP engine.

Security: In today's Web arena, security and data security in particular plays a key role. As the proposed tool is a DSS and the knowledge stored in the system is potentially valuable, security mechanism was mainly considered. All objects such as models are protected by access rules defined in the central relational database. Instead of ASP.NET internal membership mechanism, a customized user-role-module authentication system was developed for this study. With this mechanism model viewing, model training and prediction privileges have been totally customizable and manageable for system administrators.

User-friendly environment: Majority of the computer users get used to use Windows and similar operating systems' GUI controls. Thus, the graphical user interface of the developed system is expected to be perspicuous, easy to use and meet the standard Windows GUI norms. Furthermore, it should benefit the facilities of Web 2.0 and a new web programming model named AJAX. As the findings^[8] concludes the information presentation has relatively low impact at decision making satisfaction, a full RIA (Rich Internet Application) based tool development was not evaluated as a feasible choice due to its inconstancy to true platform independence. Instead, AJAX powered HTML based development was preferred due to accordance to platform independency.

Ease of use: As the most of the data mining solutions include complex tools for analysis and reporting, ASMINER targets ease of use in whole model exploration phases. To achieve this goal, simplistic and well designed user interfaces are presented to users by considering aesthetics rules. Additionally, AJAX based controls are employed in GUI elements to let users feel like using a desktop application on web environment.

3.2.2 Microsoft analysis services

Microsoft Analysis Services^[39] has been the business intelligence component of Microsoft SQL Server software since 2000. It provides OLAP and data mining features for decision makers. In addition, Analysis Services serves its own decision trees, clustering, association rules mining, naïve Bayes classification, time series analysis and neural network algorithms^[47]. Moreover, it has the feature which allows algorithm developers to code and integrate their algorithms to Analysis Services.

In decision tree algorithm platform, Microsoft invented it is own decision tree algorithm namely "Microsoft Decision Trees". This algorithm can handle both categorical and continuous variables as well as in CART and CHAID methods. In addition, it supports entropy and Bayesian score as the splitting strategy and unlike the other famous algorithms, it offers no pruning phase. In Analysis Services, with the decision tree model creation, a corresponding dependency network is also built. Tang and MacLennan^[32] state clearly that dependency networks display the relationships among the attributes

derived from decision tree model’s content. Hence, dependency networks are useful to explore correlations and influences of attributes in a dataset. In clustering models, Microsoft Analysis Services offers two types of clustering algorithms such as K-Means and EM (Expectation-Maximization) with scalable and non-scalable versions. On the other hand, well known Apriori algorithm is also employed in association rules mining.

Current version of Microsoft Analysis Services provides some standards for the mining architecture such as XMLA, DMX (Data Mining Extension to SQL for data mining operations), ADOMD.NET (ActiveX Data Objects for Data Mining – provides access to Analysis Services data objects from managed languages such as C#, VB.NET) and AMO (Analysis Management Objects – the interface to manage the Analysis Services). Although Analysis Services works with simple client-server architecture allowing clients to connect either through LAN TCP/IP connection or through IIS using the Internet, hence it does not provide a web based exploration environment. However, it enables clients to connect and retrieve the data mining model contents and even manage everything on server via Visual Studio IDE. Due to existence of its application programming interface (API) in .NET [48] and high penetration of the product in industry, Analysis Services is selected as the backend data mining system in this study.

3.2.3 System architecture and general considerations

System architecture of ASMINER is presented in Figure 1. As can be seen from Figure 1, ASMINER is built on three-tier architecture. The first data layer includes Microsoft SQL Server relational database engine and data mining services. In the second layer, the application server hosts the application logic of ASMINER. However, Internet Information Server (IIS) is chosen as the web server in the third layer. Whole application is developed using C#.NET language [49] with ASP.NET 2.0 platform on Windows Vista™. In the application layer, visualizations were constituted via HTML, Adobe Flash [50], Microsoft Silverlight [51] and Java Applet [52] technologies. In design and client side scripting, browser compatibility should be considered. Internet Explorer, Mozilla Firefox and Google Chrome are selected as the target browsers that ASMINER must be compatible with.

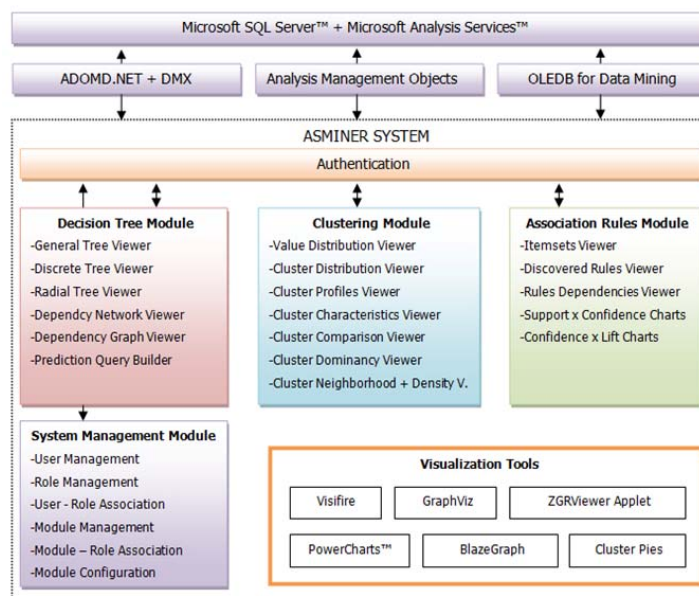


Figure 1. Detailed modules and components chart of ASMiner

To accomplish its goal, ASMINER employs client connectivity interfaces of Microsoft SQL Server at both OLTP and data mining aspects. ADOMD.NET and AMO has been used as the entry point to Analysis Services. ADOMD.NET is mainly focused on retrieving mining models’ meta-data. Besides, AMO provides management facilities on server objects at Analysis Services. Thus, model training/processing operations and model settings can only be handled via AMO. Domain

experts can load, create and manage data mining models on Analysis Services by using Visual Studio shipped with Microsoft SQL Server. With creation of data mining model in Analysis Services, it is saved with its metadata and this metadata can be retrieved by ADOMD.NET. Cooperating with AMO and ADOMD.NET, ASMINER accesses data mining models' metadata.

ASMINER consists of five main modules such as authentication mechanism, decision tree subsystem, clustering subsystem, association rules subsystem and management tools (see Figure 2). Each of these modules were designed and implemented as independent components. Authentication subsystem authorizes every request and validate if the user has access right to requested page and operation. A session based ticket is assigned to users after login step. According to ticket's privileges, users can view only the allowed modules and mining models. Furthermore, only authorized ticket owners can train and make prediction queries on related mining models. Decision tree, clustering and association rules mining subsystems have their specific type of mining model viewers. In these viewers some third party open source charting and visualization components (e.g., Visifire^[53]) are either used or self-developed in this study. These companion tools and more detailed system module chart are given in Figure 2.

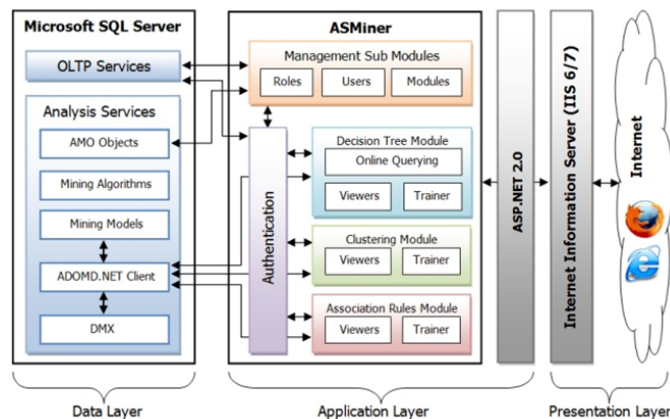


Figure 2. The general system architecture of ASMiner

ASMINER also provides a management tool for the system administrators that users, roles, modules and the relationships among them can be managed. Moreover, detailed mining model training/processing operations can be performed via this helpful tool. However, all raw data loading, ETL and model building phases are carried out by data mining experts on Microsoft SQL Server product by using Visual Studio^[65] product.

3.2.4 Decision tree module

Decision tree subsystem has five different types of viewers and an online prediction query tool. Three of the viewers are specialized to explore the nodes of decision trees and details of node contents such as distributions. In these tree viewers, default tree depth is chosen as five and in the case of deeper trees, users can expand it. Therefore, time in process of tree construction is reduced and browser resources are saved by this approach.

3.2.5 Decision tree exploration

ASMINER decision tree module has three types of tree viewer such as general tree viewer, discrete tree viewer and radial tree viewer. General tree viewer has a capacity to draw both regression trees and classification trees. To gain speed and interactivity, Javascript client side scripting technology is employed at layouting of nodes on browser window. As is known, Javascript is a de-facto client scripting language supported by all of the browsers. Besides, Walker tree drawing algorithm is employed to build perspicuous and aesthetic trees (see Figure 3). Users can navigate on trees by expanding or

closing the nodes by clicking appropriate buttons on nodes. Visifire^[53] charting solutions were used in the node histogram display for experiments.

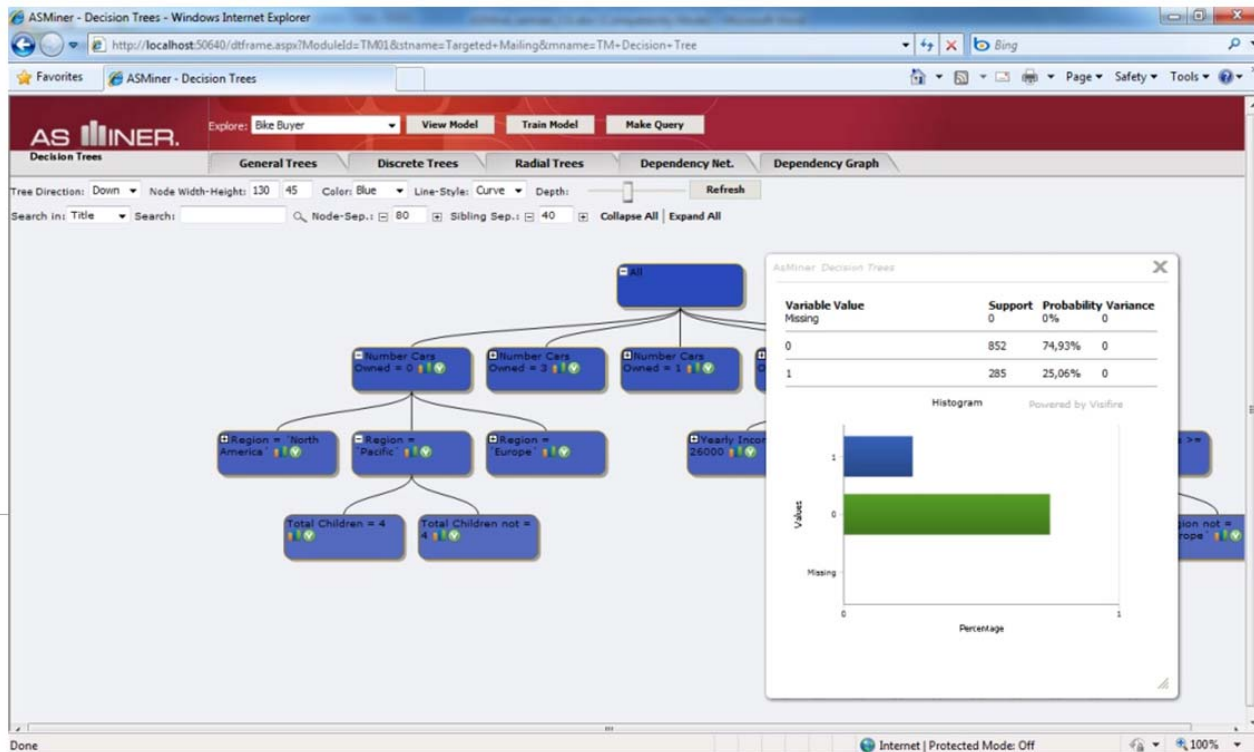


Figure 3. ASMiner general tree viewer.

However as it is against to platform independence, it is being planned to replace it with a full HTML5 compliant solution as a future work One of the other important features of general tree viewer is to have a drill-through support. The drill-through refers the technique of retrieving the raw data which owns the pattern. Drill-through data can then be saved as CSV or Excel file.

On the other hand, classification trees which have discrete values can be shown and analyzed by classification tree (discrete tree) viewer. This viewer owns some special features. First of all, unlike the first viewer, the nodes have small histogram bars which users can easily take an idea.

Second, users can select a categorical value of a predictable variable and all nodes are then be gradually colored to show the portion of the selected value in the histogram distribution of the related node. To give the users an opportunity of viewing tree structure in a different point of view, radial tree viewer is empirically integrated. In this viewer, PowerCharts™ which is one product of FusionCharts^[54] was employed. Radial view has the advantage of showing every node proportionally in one shot.

For the correlation exploration, dependency network graphs were presented to users. ASMINER has two types of dependency network viewers. In the first one, network graph is created by GraphViz and then it is shown via applet version of ZGRViewer^[55]. Graphviz is an open source and platform independent graph visualization software developed by AT&T Labs. Furthermore, it has the capability of generating directed and undirected graphs with rich drawing options. It provides its own graph description language .dot which allows users to define their graphs in textual base. Thus, Grapviz renders the graph with the help of built-in graph layout modules based on supplied .dot file. Rendered graphs can be saved in various formats such as SVG, JPG or PNG^[56]. However, ZGRViewer is a graph visualization tool implemented in Java

that is capable of panning, zooming, searching etc. Thus, users can easily navigate on the overall graph and explore the content. A sample dependency network graph displayed with ZGRViewer is presented in Figure 4. The strength of the links between the attributes shows the power of correlations between them. Therefore, a slider was located at the right side of the page to filter and see the most powerful correlations.

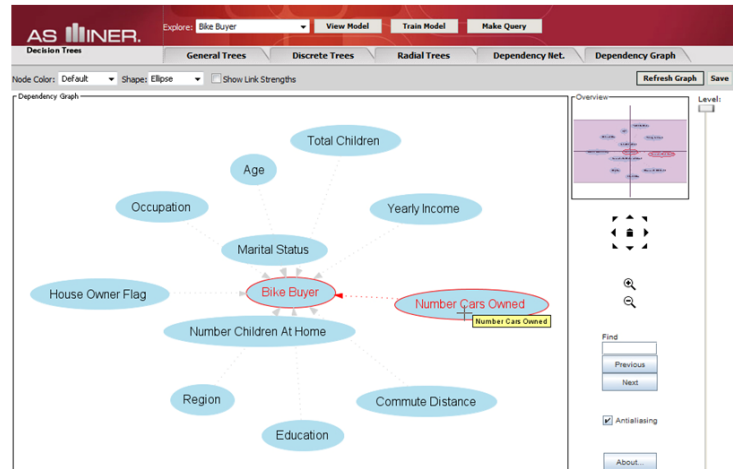


Figure 4. A sample dependency network viewer powered with ZGRViewer.

The other dependency network graph viewer is based on Flash technology and it serves different options for the users. In this viewer, Blazegraph^[57] which is a graph layout engine implemented in Flash and ActionScript is employed. Blazegraph employs a force directed algorithm to draw entire graph. In fact, the purpose of force-directed algorithms is to position nodes in two or three dimensional space so that all the edges are of roughly equal length and minimum numbers of cross edges exist.

In the lack of Java Runtime, this Flash based viewer is thought to give service to users. On the other hand, this viewer is capable of highlighting and showing the most nearest neighbors of selected nodes beside the features like zooming, rotating and unique coloring of nodes (see Figure 5).

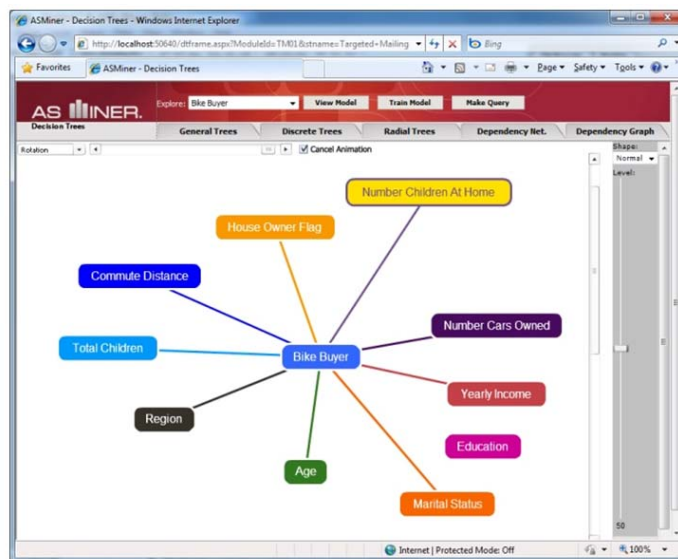


Figure 5. Dependency network viewer in ASMiner.

3.2.6 Online prediction

To complete the aim of decision tree based decision making, ASMINER presents a web-based online prediction tool for this purpose. In fact, decision tree based prediction is no more than hoping on the decision nodes with appropriate directions. At the last step of this recursive process, the value of target variable (attribute) becomes clear. In the case of regression trees, the value of target variable is calculated by the formula at each node on tree.

Two types of prediction schema such as batch and singleton queries exist in Analysis Services. While, a batch query has multiple input cases, a singleton query has only one input case that is constructed on the fly [32]. Currently, ASMINER supports only singleton queries. On the other hand, predict(), predict-with-histogram(), predict-deviation(), predict-variance() and predict-support() are the built-in functions that both Analysis Services and ASMINER currently supported by ASMINER.

Online prediction is an iterative process that can be repeated many times to achieve best decision. In the prediction scenario, user(s) may need to change the input variable(s) and select different predictable variables. Therefore, the developed tool must have ease in use, quick response time and flexibility to the needs of users (see Figure 7). In Figure 6, the steps that involved in ASMINER web-based prediction tool are presented.

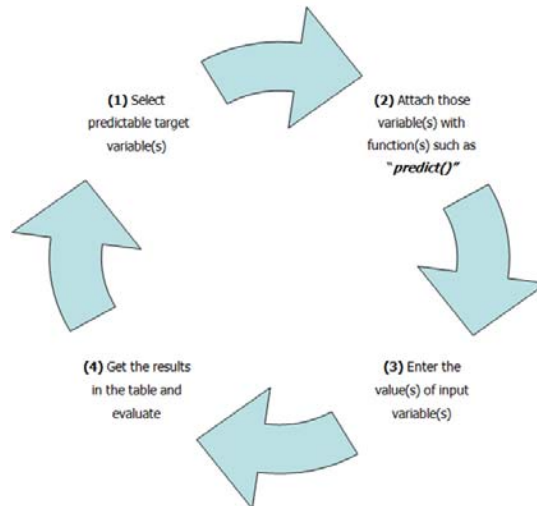


Figure 6. The steps involved in web-based prediction of ASMiner.

Query Tool

<p>Predictable Variables</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Bike Buyer 	<p>Functions:</p> <ul style="list-style-type: none"> Predict() Predict-Histogram() Predict-Probability() <li style="background-color: #e0e0e0;">Predict-Deviation() Predict-Variance() Predict-Support() <p style="text-align: center;">Add to List</p>	<table style="width: 100%; border-collapse: collapse;"> <tr><td>Age</td><td>18</td></tr> <tr><td>Bike Buyer</td><td></td></tr> <tr><td>Commute Distance</td><td>1-2 Miles</td></tr> <tr><td>Customer Key</td><td></td></tr> <tr><td>Education</td><td>Partial High School</td></tr> <tr><td>Gender</td><td>F</td></tr> <tr><td>House Owner Flag</td><td></td></tr> <tr><td>Marital Status</td><td>M</td></tr> <tr><td>Number Cars Owned</td><td>0</td></tr> <tr><td>Number Children At Home</td><td></td></tr> <tr><td>Occupation</td><td>Manual</td></tr> </table>	Age	18	Bike Buyer		Commute Distance	1-2 Miles	Customer Key		Education	Partial High School	Gender	F	House Owner Flag		Marital Status	M	Number Cars Owned	0	Number Children At Home		Occupation	Manual
Age	18																							
Bike Buyer																								
Commute Distance	1-2 Miles																							
Customer Key																								
Education	Partial High School																							
Gender	F																							
House Owner Flag																								
Marital Status	M																							
Number Cars Owned	0																							
Number Children At Home																								
Occupation	Manual																							
<p>Added Variables:</p> <ul style="list-style-type: none"> Bike Buyer Predict() Bike Buyer Predict-Probability() Bike Buyer Predict-Support() <p style="text-align: center;">Delete from list Clear all</p>		<p>Query Now!</p>																						

Query Result	Bike Buyer-Predict	Bike Buyer-Predict-Probability	Bike Buyer-Predict-Support
1		0,633941204213174	2687

Figure 7. Online prediction tool of ASMiner

(1) – (2): In the first and second step, decision maker selects the predictable variable(s) and attach them with a required member of predict function family. Pure predict() function results the value of target variable. On the other hand, predict-support() function returns the support value of the predicted target variable.

(3): To make a prediction, decision maker must enter the case that will be predicted. Thus, in the third step, the input variables are entered.

(4): In the last step, results are gathered on the fly and evaluated by decision maker. If it is required, predictable or input variables may be changed with different combinations and overall scenario repeats itself till the decision maker satisfies.

3.2.7 Clustering module

As the nature of clustering is describing the data itself, ASMINER clustering subsystem has focused on describing and introducing discovered clusters in different point of views. On the other hand, well introduction of clusters means well-informed decision makers. Therefore, most of the viewers that exist in ASMINER clustering subsystem, target to inform the users about characteristics, statistical differences and discriminations of clusters. Furthermore, a distribution based cluster dominancy exploration method and viewer is empirically developed to gain insight on that which clusters are highly dominant or recessive at the intersection of values of discrete variables in two dimensional spaces.

3.2.8 Understanding and interpreting the characteristics of clusters

To understand the clusters to be discovered, ASMINER presents six different types of clustering viewer. These are value distribution viewer, cluster distribution viewer, general cluster profiles viewer, specific cluster characteristics viewer, cluster comparison viewer and lastly cluster neighborhood + distribution density viewer.

Although a well-used general cluster profiles viewer is useful at reporting distribution of clusters at a glance, to be more precise and informative, ASMINER presents two new viewers. As the first one is value-variable distribution over a specific cluster viewer then the second one is that cluster distribution over a value of one specific variable viewer. With these viewers decision makers equipped with the opportunity of having statistical insights of clusters. Visifire charts are used in the visualizations of these viewers by presenting different charting options (e.g., charting types, colors etc.). However, the users may want to obtain information about both clusters and variables at a glance. For this purpose, ASMINER offers cluster profiles viewer. The approach and design of the viewer is inspired from original Analysis Services “Cluster Profiles”. The difference is that, ASMINER’s cluster profile is running on web environment and rendered with pure HTML. Thus, the page can be saved to local drives or copied to clipboard. In cluster profile viewer, discrete valued variables are shown with a colored histogram and the numbers of items in the histogram can be increased. However, unlike the discrete variables, numerical variables were indicated by “box-plot” technique that has the ability of showing mean and standard-deviation of the variable.

In some situations, identification of characteristics of the clusters’ dominant properties may be required. For that reason, ASMINER presents cluster properties viewer. In this viewer, the properties of a picked cluster are listed in descending support value order. In addition, this viewer has the capability of listing the general properties of overall population. However, in order to develop an efficient decision support system, a cluster comparison viewer is also implemented. The goal of comparison viewer is to list the discriminations between two clusters or one cluster and its complement. Both of the viewers were also implemented by inspiring from Microsoft Analysis Services. In addition to those viewers, a cluster neighborhood and distribution density viewer is implemented by using ZGRViewer and GraphViz. The first purpose of this viewer is to show the similarities or dissimilarities among the discovered clusters and relative strength of them. The second is to highlight the values of variables according to their dense in their distribution over clusters. As can be seen in Fig. 8, clusters are gradually blue colored due to the discrete value distribution and the links between them indicates the strength of the similarity among them. By using this viewer and the cluster distribution viewer, a typical decision maker of a market can answer the questions such as “Where are my customers who have at most one child?”

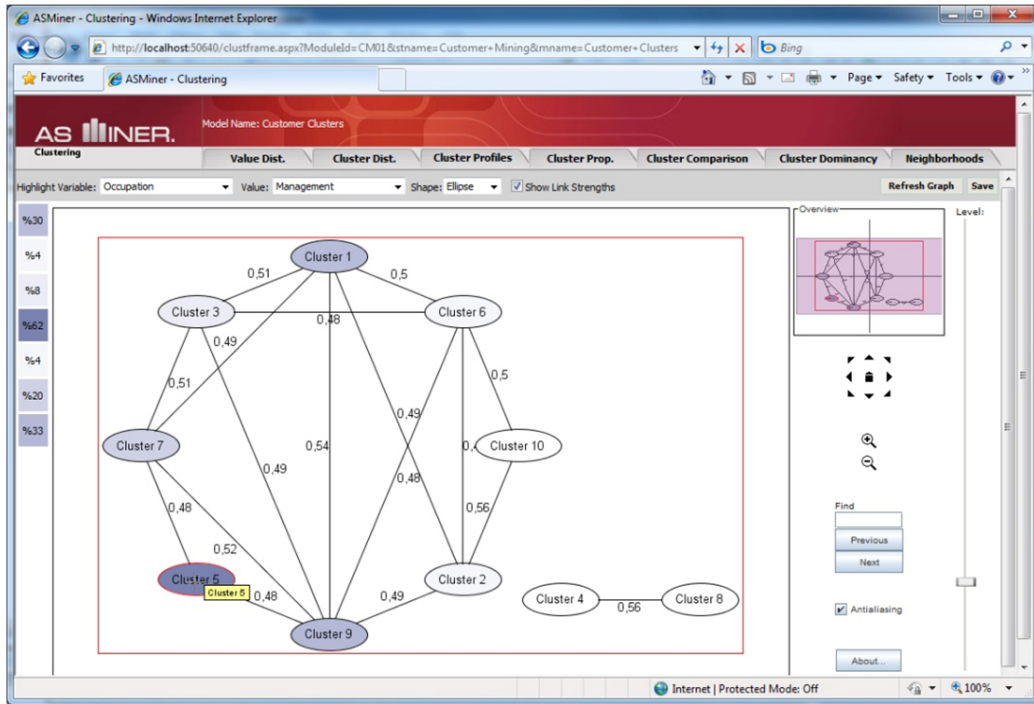


Figure 8. Cluster neighborhood viewer.

3.2.9 Deep diving at cluster investigation

For better understanding of clusters and the cases that exist in, an empirical study called cluster distribution and dominancy exploration viewer is designed and implemented. The main idea behind of this viewer is that deep diving to the clusters and examining them in the case level. Furthermore, this viewer is specially designed to investigate specific cluster(s). If the decision maker is conscious and determined the target cluster(s) to examine, the approach of this module is used. Cluster distribution and dominancy viewer display the cluster distributions over the intersections of values of two selected variables and work only for discrete or discretized variables. In fact, the final result diagram is no more than a grid in two dimensional spaces. However as a pre-requirement subjected cluster mining model must be created in order to have drill-through support. This feature is essential due to the fact that, the viewer must retrieve the real cases of clustering model by drill-through feature. Therefore, the system operator or the data mining expert must enable this feature in process phase of model. At the second stage, a pie graph is constructed for every value intersection on the X-Y axis based on the coverage of clusters in means of percentage. An example screenshot is given in Figure 9. If the picked cluster is the most dominant one for a value pair then the contour of the related circle is being drawn with a bold line style. For better understanding, the viewer presents a more detailed view of the pie graph which is under concern. As a result, decision makers have an opportunity to discover the unseen properties of the cluster(s) that they investigate and have an insight about which clusters are dominant or recessive in which value pairs. Furthermore, as the discrete variables do not have a natural order. Therefore if they are logically ordered (e.g., education level or age ranges), decision makers can find a chance to discover some dense regions which a cluster dominantly distributed. This is sampled in Figure 9. As can be seen on Fig. 9, the pie graphs of cluster distributions are depicted. If a picked density of picked attribute is maximal or minimal at a cluster distribution then that pie graph is marked with bold outer line. Nevertheless, if the centre points of these circles are linked to build closed convex hulls, then some decision region(s) can be formed on 2D graph. Furthermore, those regions may contain some hidden strategic knowledge. However, in order to benefit from this viewer, decision makers should run it for many times to explore different points of views. This property is left to next version as a future work.

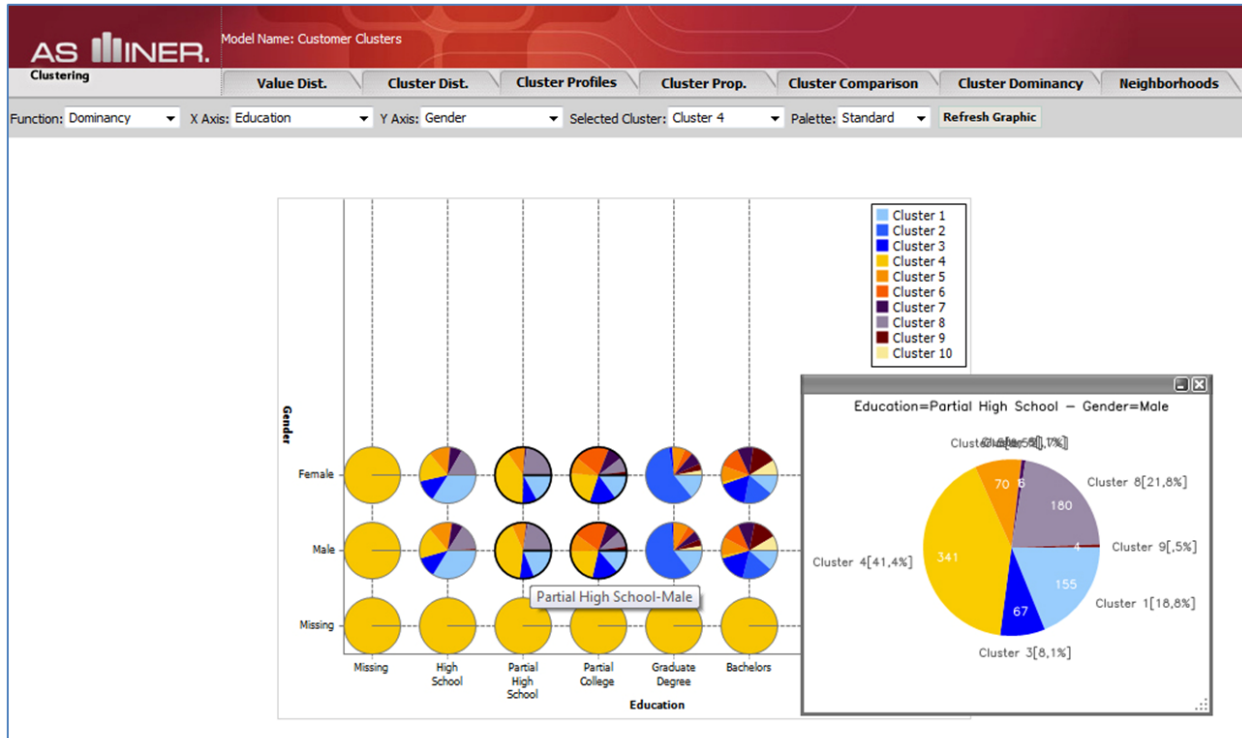


Figure 9. Investigation of a selected cluster with cluster dominance viewer.

3.2.10 Association rules mining module

Association rules mining is probably one of the most important data mining analyze for marketing sector. Although, it is a well used technique in market-basket analysis, association rules discovery has been used in various fields. As it is specialized for revealing hidden associations and describes the data in this way, it deserves a special attention. For these reasons, ASMINER is equipped with association rules subsystem.

Inspired from Analysis Services, three viewers are designed for this module. Itemset viewer, rules viewer and rules dependency network viewer constitute the association rules module of ASMINER. As stated before, Analysis Services employs the algorithm Apriori for association rules mining. Therefore, itemset features are very important as they constitute basis of first pass in the algorithm. For this reason, itemset viewer was equipped with fundamental features such as filtering and sorting.

On the other hand, ASMINER presents a comprehensive rules viewer (see Figure 10). Unlike the other implementations of Apriori algorithm, Analysis Services focuses on the Importance (namely lift) score for measuring the usefulness of the rule^[48]. Rule importance score ranges between -1 to 1. As the positive scored rules have potential advantages, ASMINER focuses on the ways of filtering and saving the important rules that decision makers require. Therefore, the rule viewer is equipped with a minimum importance, minimum confidence and textual search features. In addition, result page sizes can be set and users have the ability to save them in Excel or CSV file formats. However, ASMINER offers two new viewers such as support x confidence and confidence x lift charts which are not available in original Analysis Services application. By using these viewers, decision makers gain the opportunity to view and analyze the rules in different perspectives. The fundamental operation in these viewer is to plot the rules on a two dimensional chart by their values on the related axes. Then, plotted points can be studied and checked for further analyzes. Finally, correlations between the antecedents and the predicates of the rules are shown by associative dependency network viewer. Blazegraph component was employed in this viewer. As the association rules analysis is a major area in marketing, a comprehensive association rule exploration system is aimed for easy and effective decision making. In addition, quick report generation is another objective for this study.

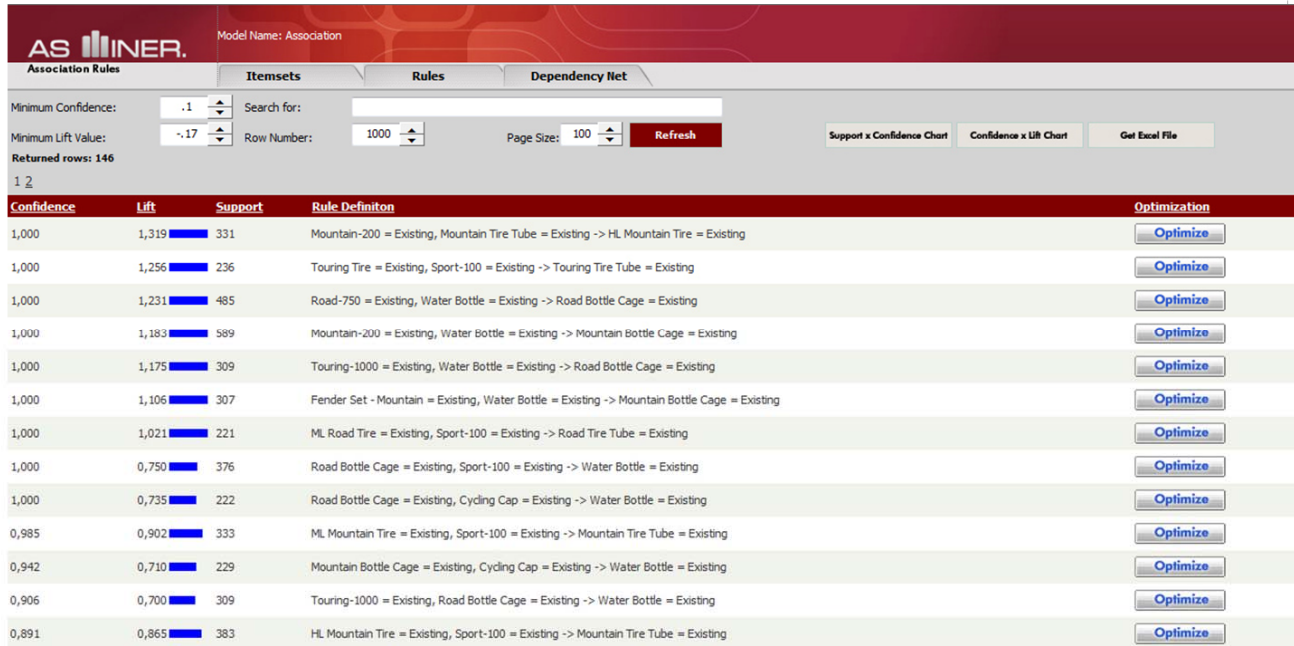


Figure 10. Association rules viewer module of ASMINER

3.2.11 Management facilities of ASMINER

As the claim of ASMINER is “geographic freedom at decision making”, a web-based system management utility for system administrators was also developed to support this claim. With this tool, users, roles active mining models in terms of modules and the relationships among them can be managed easily. In addition, data mining models can be retrained or reprocessed with the parameters given by administrators. With the help of Analysis Services’ AMO programming interfaces, model information can be retrieved and the updates are directly reflected. Another advantage of this utility is that system administrator can allow or ban the user(s) to explore models. Moreover, even the user has right to access the model, he/she can be forbidden of training or making predictions over it. To sum up, the management utility of ASMINER aims to establish easy to manage and flexible tool for the system administrators.

3.2.12 Example usage scenarios

At present, data mining oriented decision making processes in companies take too much time due to the barriers between decision makers and data mining experts. Every minor variation in construction of reports requires alternation in data mining models. Therefore, these iterations consume considerable amount of time as it starts a new loop between the decision makers and data mining experts. Yet, this bottleneck can be reduced by a bit participation of decision makers in maintaining the mining models. Current approach in data mining software packages left this responsibility only to data mining experts. However, according to the idea of this study, decision makers can do minor changes in mining models and generate reports without the help of data mining experts. Moreover, they can do predictions by themselves. By this little contribution, the amount of time required in decision making can be reduced. In the light of these ideas, knowledge workers and decision makers also start to have the opportunity of becoming lightweight data miners. To support this claim, exploration and reporting phases of knowledge discovery were enabled for decision makers and knowledge workers by presenting a platform independent web-based data mining tool.

There exist many case studies which could be performed with ASMINER. To give cases in point, officers of banks and insurance companies can use ASMINER without encountering a problem. Decision tree based risk assessment for all incoming requests can be carried out by officers themselves on web browser by the help of proposed online prediction tool.

On the other hand, a market manager can analyze the customer segmentations even at vacation. Moreover, researchers who have data mining models to be shared can also use ASMINER and introduce their models to the others.

4 Discussion and comparison

As stated previously, ASMINER has been designed and implemented for data mining oriented decision making activities for those who are even not familiar data mining. Especially, enabling the knowledge workers and let them participate in decision making periods is mainly targeted. However, ASMINER has some advantages and limitations. Although the fact that Web based DSSs are very domain specific and application centric, ASMINER provides a generic data mining analysis platform for decision makers. This generic approach is a plus for handling unstructured analyze tasks for various business processes. However, as a result of this, decision makers should be aware of the data that they work on. On the other hand, ASMINER is a useful tool that can help to decision makers or model builders to share their models to other people throughout the world. In lack of such kind of tool, this process requires excessive work and often faces misunderstandings. Besides, it enables data mining experts to manipulate models with limited numbers of facilities. However, although ASMINER is built for online decision makers, it is not purely designed for group decision making activities. Therefore ASMINER cannot be considered as group DSS. Though it has multiple user supports, it does not serve any collaborative approach such as exchanging and sharing information, discussion boards etc... As a consequence, ASMINER can be enhanced by adding group DSS paradigms and facilities leading ASMINER to become a pure group DSS.

On the other hand, management of the mining models is limited in current version of ASMINER. As stated before, knowledge workers have either no or very limited information about data mining methodologies. Therefore, current version is based on separation of responsibilities of each type of user. Detailed model building phase is left to data mining experts on server side and model exploration interfaces are served for knowledge workers. In some recent web based data mining tools such as in Data-Applied^[58] it can be seen that whole data mining process is left to all kinds of users. However these types of tools are designed for data mining experts and do not share the isolation idea of ASMINER. As stated before, ASMINER aims to serve an easy to use data mining platform for knowledge worker by isolating them from complex operations. To do so, model training parameters can be adjusted by authorized users, the responsibility of selection of variables and suitable algorithm is left to data mining experts. Furthermore, in current situation, data transformation and loading processes are all handled in Microsoft SQL Server environment. However, full management capabilities including data loading, transformations and model construction phases for experts are planned to be handled in ASMINER in next version. With this upgrade, data mining experts will also be available to operate on web environment which means full platform and location independence. In literature, some studies like^[4, 23] which include these steps exist; however, those tools are either domain centric or application specific. On the other hand, building a domain and platform independent, general purpose data mining oriented DSS have many technical difficulties which must be resolved. First of all, the nature of Web's client server architecture does not well suit for transporting huge amount of data over HTTP protocol. Transporting the data which are in the range of megabyte sizes generally causes no problems; however sending data which are in gigabyte sizes is a remarkable problem due to delays and connection cut failures with server. Furthermore, GUIs of a full featured web based DSS needs highly talented user controls which is used to simulate conventional desktop based applications. Therefore, rich internet application (RIA) frameworks come to into prominence and Microsoft Silverlight or Adobe Flex based development come up. By this means, some of the facilities (e.g. video conferencing) which belong to group DSS approach can then be implemented and integrated but these development technologies have the platform independence problem such as browser incompatibility issues. Thus, a hybrid approach which combines HTML and RIA can be considered to be the best solution for available browsers. However not all the RIA options are platform independent. Currently a tendency to tablet usage exists. In these tablets different operation systems are being employed. While Adobe Flash is not enabled in some products, Silverlight does not exist in many tablets. Thus, true platform independency has not been fully provided in current tablet systems and even in desktop operating system. As a solution for providing more dynamic Web, HTML 5^[59] has been proposed. Nevertheless, due to client side scripting limitations which prohibits data layer access, HTML5 is limited to be a candidate platform for online web based data

mining solutions. However, for visualization purposes, its employment stands the most reliable solution because of being a worldwide standard. In the end, currently employment of RIA technologies instead of server side scripting methods (e.g. PHP^[60], ASP.NET) constitutes a question mark for true platform independence. Nonetheless, if platform independence is not the main goal, RIA services stand as the best choice due to very rich opportunities in terms of fast interaction and high quality data visualization. In the Table 1 below, features of ASMINER are compared with some of the commercial on the shelf products and academia. On this comparison table, readers can find the aspects in which we are comparing ASMINER with others. On this table two kinds of data mining products are listed: (1) desktop based and (2) web based. In web based solutions ASMINER and Data-Applied^[58] come to prominence. As a pure web tool, Data-Applied has many useful features. However on two main features ASMINER discriminates: (1) achieving true platform independence and (2) defining different responsibility roles between decision makers and data mining experts. From this point of view, ASMINER and Data-Applied discriminate. As the product Data-Applied™ is fully based on RIA (Silverlight), ASMINER is fully based on HTML. In some visualization portions Flash, Silverlight and JavaScript is used partially. However, with the development of HTML5, it is being planned to move HTML5 for sophisticated graphs. ASMINER currently supports three data mining method which are shipped with MAS. However, MAS provides other types of methods such as time series with ARIMA and ARTxp algorithms and sequence clustering method. In next version, it is planned to implement viewers of these methods and enable users to benefit from these methods. At this point, one benefit of MAS comes up. Microsoft Analysis Services supports third party algorithm implementation on C# or C++ languages and integration of these third party implemented algorithms is straightforward. Therefore, although ASMINER is highly coupled with MAS, with this open algorithm development feature, it's not limited to Microsoft's choices. One another good thing is that Microsoft SQL Server supports all open database connectivity ODBC databases. So, data located in many database systems including Oracle can be imported to SQL Server and then be analyzed via MAS and ASMINER. These two features make ASMINER free to access various data sources. In ASMINER, many viewers are designed inspiring form the original Microsoft Analysis Services (MAS) viewers. The main reason behind this is to provide the compatibility with MAS and ease of orientation for the users who are accustomed to it. From this point of view, this study could be evaluated as the web based shell of MAS. However, the viewers of ASMINER have more or less differences comparing to original MAS viewers. First of all, while MAS viewers are purely Windows controls, ones in ASMINER are built for Web browsers. Thus, they are either based on dynamic HTML or more dynamic rich internet application components like Flash. On the other hand, Web based applications have another well known bandwidth problem that must be also addressed in ASMINER. For solution, amount of data transferred between Web browser and server are tried to be kept at minimum level. At this point RIA based viewers contributed as they save data in cache at client side. Furthermore, available data blocks such as "Viewstate" data in ASP.NET are compressed before sending. On the other hand ASMINER has some viewers which do not exist in Analysis Services (e.g., cluster dominancy and radial tree viewers). Besides, the main contribution of ASMINER arises from adding one tier to the data mining architecture of MAS. Although, MAS presents some features for connection via HTTP protocol, ASMINER offers a complete three-tiered web-based data mining platform. In addition, to let users the feel of desktop applications, AJAX based techniques and controls are employed. Thus, the performance and user interaction capabilities enhanced.

5 Conclusion

In this study, combining data mining oriented decision making and the power of Web by developing a web-based DSS namely ASMINER is aimed. Furthermore, it is designed and implemented to take full advantages of ultimate technologies in Internet and decision support systems. Moreover, independency from location in decision making process is emphasized. In the availability of an Internet connection, online decision making is delivered to the decision makers whenever and wherever they may need it. Likewise, as emphasized in^[14], these kinds of systems can be accessed globally; the location of decision makers is irrelevant and they do not need any installation by offering a web-based DSS.

At the background, main goal of this study is to enable the decision makers to participate in management of knowledge base instead of viewing the results only. However this approach includes some risks that must be addressed. Intervention

to mining models needs data mining knowledge and business analyst skills. Therefore, at first stage, this intervention is kept at minimal level by not to add specific features. However, with the use of ASMINER the trends should be analyzed by business analysts and level of intervention should be checked periodically and set.

Table 1. Comparison table for ASMINER and the some other products

	Analysis Services ^[39]	ASMINER	SPSS ^{Clementine [38]}	Data-Applied ^[58]	Weka ^[61]
System Architecture	2 Tier	3 Tier	Native App	3 Tier	Native App
Platform	Windows™	Platform Independent	Windows™	Windows™, Linux (Silverlight™ req.)	Windows™, Linux, MacOSX
Server Requirements	Windows™, .NET 3.0+, Analysis Server, SQL Server	Windows, .NET 3.5, GraphViz, Analysis Server, SQL Server, IIS	No server need	Windows™, IIS, SQL Server or MySQL	No server need
Client Requirements	Windows™, .NET 3.0+, Analysis Services, Visual Studio.NET IDE	Web Browser, Silverlight, Java Runtime & Flash plug-in	Windows™	Web Browser, Silverlight plugin	Windows™, Linux, MacOSX, Java Runtime
Web Based	No	Yes	No	Yes	No
Multiuser Support	Intranet	Internet and Intranet	No	Internet and Intranet	No
Sophisticated Decision Tree (DT) Viewer	Yes	Yes	Yes	Yes	Yes
Dependency Net on DT	Yes	Yes	No	Yes	No
Radial DT Viewer	No	Yes	No	No	No
Drill Down on DT Viewer	Only listing	Listing and Excel file save	No	Listing and text file save	No
Saving Query Results	DB File	Excel & CSV	Yes	Yes	Yes
Cluster Profiler Viewer	Yes	Yes	Yes	Yes	Limited
Cluster Attribute & Value Distribution Viewers	Limited	Detailed	Detailed	Yes	No
Cluster Graph	Yes	Yes	No	Yes	No
Cluster Density Viewer	No	Yes	No	Yes	Yes
Cluster Discrimination Viewer	Yes	Yes	No	No	No
Assoc. Rules Itemset Viewer	Yes	Yes	Yes	No	Yes
Assoc. Rules Rule Viewer	Yes	Yes	Yes	Yes	Yes
Assoc. Rules Filtering	Yes	Yes	Yes	Yes (Limited)	No
Rule Ordering Optimization	No	Yes	No	No	No
Support vs. Confidence Graphs on Assoc. Rules	No	Yes	No	No	No
Assoc. Rules Dependency Network	Yes	Yes	No	No	No
Interaction Speed on DT	Fast at few number of nodes, moderate on many nodes	Fast at few nodes and slow on many nodes	Relatively fast	Fast	Relatively fast
Model Training	Data manipulation and parameter selection is possible	Parameter selection is possible	Data manipulation and parameter selection is possible	Data manipulation and parameter selection is possible	Data manipulation and parameter selection is possible
Saving Graphs	Copy to clipboard	PNG Image File	Possible	Yes (Screenshot)	Yes
System Extensibility	New algorithms can be developed and integrated with C++ ve C#	New viewers can be developed and integrated with C# and VB.NET	Connection to Analysis Services™ is possible	Batch processing or manipulation via XML Web API	New algorithms can be developed via WEKA API

On the other hand, extending the management capabilities of data mining models and integrating new methods to the system is planned. As these further research directions are integrated into the existing implementation of ASMINER, the system would be fully automated and more comprehensive web-based data mining DSS for both decision makers and data mining experts.

References

- [1] G. M. Marakas, Decision Support Systems In the 21st Century (Prentice Hall, New Jersey, 2007).

- [2] S. Ba and A.B. Kalakota, Whinston, Executable Documents DSS, in Proc. of 3rd International Conference on DSS. Hong Kong. 1995: 22-23.
- [3] H.K. Bhargava, D.J. Power, Decision support systems and web technologies: a status report, AMCIS 2001 Proceedings, 2001.
- [4] D. Delen, R. Sharda, P. Kumar, Movie forecast Guru: A Web-based DSS for Hollywood managers, Decision Support Systems. 2007; 43: 1151-1170. <http://dx.doi.org/10.1016/j.dss.2005.07.005>
- [5] J.P. Shim, M. Warkentin, J.F. Courtney, D.J. Power, R. Sharda, C. Carlsson, Past, present, and future of decision support technology. Decision Support Systems. 2002; 33(2): 111-126. [http://dx.doi.org/10.1016/S0167-9236\(01\)00139-7](http://dx.doi.org/10.1016/S0167-9236(01)00139-7)
- [6] Y. Xie, H. Wang, J. Efstathiou, A research framework for Web-based open decision support systems, Knowledge-Based Systems. 2005; 18: 309-319. <http://dx.doi.org/10.1016/j.knsys.2004.12.001>
- [7] D. J. Power, Web-based and model-driven decision support systems: concepts and issues. 'Model-driven and web-based decision support systems' Mini track, AMCIS 2000, Americas Conference on Information Systems, (Long Beach, CA, August, 2000).
- [8] P. Bharati and A. Chaudhury, An empirical investigation of decision-making satisfaction in web-based decision support systems, Decision Support Systems 2004; 37: 187-197. [http://dx.doi.org/10.1016/S0167-9236\(03\)00006-X](http://dx.doi.org/10.1016/S0167-9236(03)00006-X)
- [9] J.H. Heinrichs, J. Lim, Integrating web-based data mining tools with business models for knowledge management. Decision Support Systems. 2003; 35(1): 103-112. [http://dx.doi.org/10.1016/S0167-9236\(02\)00098-2](http://dx.doi.org/10.1016/S0167-9236(02)00098-2)
- [10] S. L. Alter, Decision Support Systems: Current Practices and Continuing Challenges (Adison Wesley, 1980).
- [11] C. Romero, S. Ventura, E. Garcia, Data mining in course management systems: Moodle case study and tutorial, Computers & Education. 2008; 51(1): 368-384. <http://dx.doi.org/10.1016/j.compedu.2007.05.016>
- [12] D.J. Power, Decision Support Systems: Concepts and Resources for Managers (Quorum Books, Westport, CT, 2002).
- [13] H.K. Bhargava, D.J. Power, D. Sun, Progress in Web-based decision support technologies, Decision Support Systems. 2007; 43: 1083-1095. <http://dx.doi.org/10.1016/j.dss.2005.07.002>
- [14] Y. Duan, J. S. Edwards, M-X. Xu, Web-based expert systems: benefits and challenges, Information & Management. 2005; 42(6): 799-811. <http://dx.doi.org/10.1016/j.im.2004.08.005>
- [15] F. Zetian, X. Feng, Z. Yun, Z. XiaoShuan, Pig-wet: a web-based expert system for disease diagnosis, Expert Systems with Applications. 2005; 29(1): 93-103. <http://dx.doi.org/10.1016/j.eswa.2005.01.011>
- [16] R. Kohli, F. Piontek, T. Ellington, T. VanOsdol, M. Shepard, G. Brazel, Managing customer relationships through e-business decision support applications: a case of hospital-physician collaboration, Decision Support Systems. 2001; 32(2): 171-179. [http://dx.doi.org/10.1016/S0167-9236\(01\)00109-9](http://dx.doi.org/10.1016/S0167-9236(01)00109-9)
- [17] C. Chiu, Y. Chen, I. Kuo, H.C. Ku. An intelligent market segmentation system using k-means and particle swarm optimization, Expert Systems with Applications. 2009; 36(3): 4558-4565. <http://dx.doi.org/10.1016/j.eswa.2008.05.029>
- [18] T. Arora, Web-based analytics for the plant, Manufacturing Systems. 2000; 18(9): 76 - 83.
- [19] S. Liao, J. Chen, T. Hsu, Ontology-based data mining approach implemented for sport marketing. Expert Systems with Applications. 2009; 36(8): 11045-11056. <http://dx.doi.org/10.1016/j.eswa.2009.02.087>
- [20] L.M. Liu, S. Bhattacharyya, S.L. Sclove, R. Chen, W. J. Lattyak, Data mining on time series: an illustration using fast-food restaurant franchise data, Computational Statistics & Data Analysis. 2001; 37(4): 455-476. [http://dx.doi.org/10.1016/S0167-9473\(01\)00014-7](http://dx.doi.org/10.1016/S0167-9473(01)00014-7)
- [21] C. Chien, L. Chen, Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, Expert Systems with Applications. 2008; 34(1): 280-290. <http://dx.doi.org/10.1016/j.eswa.2006.09.003>
- [22] N.Y. Jan, S.C. Lin, S.S. Tseng, N.P. Lin, A decision support system for constructing an alert classification model, Expert Systems with Applications. 2009; 36(8): 11145-11155. <http://dx.doi.org/10.1016/j.eswa.2009.02.097>
- [23] G. Tsoumakas, I. Vlahavas, An interoperable and scalable Web-based system for classifier sharing and fusion. Expert Systems with Applications. 2007; 33(3): 716-724. <http://dx.doi.org/10.1016/j.eswa.2006.06.021>
- [24] Jorge, J. Poças, P.J. Azevedo, A Methodology for Exploring Association Models, In Visual Data Mining: theory, Techniques and Tools For Visual Analytics, S. J. Simoff et al., Eds. Lecture Notes In Computer Science, vol. 4404. (Springer-Verlag, Berlin, Heidelberg. 2008: 46-59, 2008. http://dx.doi.org/10.1007/978-3-540-71080-6_4
- [25] K.C. Lee, N. Chung, A web DSS approach to build an intelligent internet shopping mall by integrating virtual reality and avatar, Expert Systems with Applications. 2005; 28(2): 333-346. <http://dx.doi.org/10.1016/j.eswa.2004.10.013>
- [26] M. Chen, Y. Liu, C-W. Wang, Y-W. Fan, Y-P. J. Chi, Team-Spirit: Design, implementation, and evaluation of a Web based group decision support system, Decision Support System. 2007; 43(4): 1186-1202. <http://dx.doi.org/10.1016/j.dss.2005.07.008>
- [27] T. Cameron, An interactive web based decision support system for hazardous industry land-use planning, Computers & Chemical Engineering. 2000; 24(2-7): 1057-1062.
- [28] P. Liu, A self-organizing feature maps and data mining based decision support system for liability authentication of traffic crashes, Neurocomputing. 2009; 72(13-15): 2902-2908. <http://dx.doi.org/10.1016/j.neucom.2008.06.032>

- [29] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, (Pearson Education Inc., Boston, 2006).
- [30] U. Fayyad, G. Piatetsky-Shapiro, P. Smith. From data mining to knowledge discovery in database. American Association for Artificial Intelligence. 1996 August: 37-54.
- [31] H.M. Chung, P. Gary, Special section: Data mining. Journal of Management Information Systems. 1999; 16(1): 11-16.
- [32] Z. Tang, J. Maclennan, Data Mining with SQL Server 2005 (Wiley Publishing, Indiana Polis, 2005).
- [33] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees (Wadsworth & Brooks/Cole Advanced Books and Software, Monterey, CA, 1984).
- [34] G.V. Kass, An exploratory technique for investigating large quantities of categorical data, Applied Statistics. 1980; 29(2): 119-127. <http://dx.doi.org/10.2307/2986296>
- [35] J. Quinlan, C4.5: Programs for Machine Learning, (Morgan Kaufmann, San Mateo, CA, 1993).
- [36] J.T. Tou, R.C. Gonzalez. Pattern recognition principles, (Addison-Wesley, London, 1974).
- [37] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases. in Proc. of the ACM SIGMOD international conference on management of data. 1993: 207-216.
- [38] SPSS Clementine. Available from: <http://www.spss.com/software/modeling/modeler/>
- [39] Microsoft Analysis Service. Available from: <http://www.microsoft.com/sqlserver/2008/en/us/Analysis-Services.aspx>
- [40] SAS Enterprise Miner. Available from: <http://www.sas.com/technologies/analytics/datamining/miner/>
- [41] Srinivasan, Alternative measures of system effectiveness: associations and implications, MIS Quarterly. 1985; 9(3): 243-253. <http://dx.doi.org/10.2307/248951>
- [42] J. Bailey, S. Pearson, Development of a toll for measuring and analyzing user satisfaction, Management Science. 1983; 29(5): 530-545. <http://dx.doi.org/10.1287/mnsc.29.5.530>
- [43] W.R. King, B.J. Epstein, Assessing Information System Value: An Experimental Study, Decision Sciences. 1982; 13(4): 34-45.
- [44] Vessey, The effect of information presentation on decision making: a cost-benefit analysis, Information and Management. 1994; 27(2): 103-119. [http://dx.doi.org/10.1016/0378-7206\(94\)90010-8](http://dx.doi.org/10.1016/0378-7206(94)90010-8)
- [45] E.B. Swanson, A note of informatics, Journal of Management Information Systems. 1985-1986: 2(3): 86-91.
- [46] Dolphin Browser, <http://home.dolphin-browser.com/tunny/Home.htm>
- [47] Z. Tang, J. Maclennan, P.P. Kim, Building data mining solutions with OLE DB for DM and XML for analysis, ACM SIGMOD Record. 2005; 34(2): 80-85. <http://dx.doi.org/10.1145/1083784.1083805>
- [48] J. Maclennan, Z. Tang, B. Crivat, Data Mining with SQL Server 2008 (Wiley, Indiana Polis, 2008).
- [49] H. Schildt, C#, A Beginner's Guide (Osborne/McGraw-Hill, Berkely, CA, 2001).
- [50] Adobe Flash, Available from: <http://www.adobe.com/products/flash/>
- [51] Silverlight, Available from: <http://www.microsoft.com/SILVERLIGHT/>
- [52] Java, Available from: <http://java.sun.com>
- [53] Visifire, Available from: <http://www.visifire.com>
- [54] Fusion Charts, Available from: <http://www.fusioncharts.com/>
- [55] E. Pietriga, A Toolkit for Addressing HCI Issues in Visual Language Environments. IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05). 2005: 145-152. <http://dx.doi.org/10.1109/VLHCC.2005.11>
- [56] GraphViz. Available from: <http://www.graphviz.org>
- [57] Blaze Graph, Available from: <http://sourceforge.net/projects/blazegraph/>
- [58] Data-Applied, Available from: <http://www.data-applied.com>
- [59] HTML5, Available from: dev.w3.org/html5/spec/
- [60] PHP, Available from: <http://php.net/>
- [61] Weka, Available from: <http://www.cs.waikato.ac.nz/ml/weka/>
- [62] Internet Explorer: Available from: <http://windows.microsoft.com/tr-tr/internet-explorer/download-ie>
- [63] S. Bozkir, E. A. Sezer. A Three-Tiered Web-based Exploration and Reporting Tool for Data Mining, IADIS Data Mining Conference, (Freiburg, Germany, 2010). PMID:20202047
- [64] Rapid Miner, Available from: <http://rapid-i.com/content/view/181/190/>
- [65] Microsoft Visual Studio, Available from: <http://www.microsoft.com/visualstudio/eng>