

BRINGING BUSINESS INTELLIGENCE INTO FINANCE CURRICULUM

Stevan Mrdalj, Eastern Michigan University, smrdalj@emich.edu
Alahassane Diallo, Eastern Michigan University, adiallo@emich.edu

ABSTRACT

The quantity of information around us is rising wherever we look. Companies are using Business Intelligence techniques to explore data to get a complete view of their operations which allows them to discover trends, improve their forecasting and operate more efficiently. The banking industry perceives Business Intelligence as a major tool for determining the consumer creditworthiness. For that we need new kinds of business intelligence professions. At the same time, Business Intelligence has not yet become a mainstream field at Business Schools and the shortage of Business Intelligence related professions will continue to rise. The main goal of this paper is to report on the development of an applied Business Intelligence (BI) module for the finance curriculum. The structure and components of this module will be presented as a BI life cycle that can be incorporated in a banking course in the finance curriculum. Such initiatives may be extended to other courses and disciplines, thus helping to promote Business Intelligence to become a widely recognized applied field of study in Business Schools.

Keywords: Business Intelligence, Finance, Consumer Creditworthiness, Data Mining, SQL Server, Curriculum

INTRODUCTION

Today all areas of life are affected by the way that information is managed. Economic production used to be based in the factory, now Business Intelligence (BI) mines the companies' abundant data for new ideas. For example, the availability of business intelligence enables businesses to customize to small niche markets around the world.

With today's highly dynamic markets and extremely aggressive competitors, "guessing" and "trial and error" are not effective managing strategies. All levels of business analysts need quick-and-easy access to information, to make timely and accurate decisions. The BI is a systematic approach to improve high-volume operational and managerial decisions using analytical tools. Hence, it promotes a shift from speculative guessing to informed and fact-based decision making [13].

The Economist reports in its special edition on BI [5] that "A few industries have led the way in their ability to gather and exploit data. Credit-card companies monitor every purchase and can identify fraudulent ones with a high degree of accuracy, using rules derived by crunching through billions of transactions. They learn that stolen credit cards are more likely to be used to buy hard liquor than wine, for example, because it is easier to fence. Insurance firms are also good at combining clues to spot suspicious claims: fraudulent claims are more likely to be made on a Monday than a Tuesday, since policyholders who stage accidents tend to assemble friends as false witnesses over the weekend".

Before the BI software has matured and the price of computing and storage has declined, BI was available only to the large companies. Now that BI has moved into the mainstream, companies use Business Intelligence techniques to get a complete view of their operations which allows them to discover trends, improve their forecasting and operate more efficiently.

The literature and the above examples make BI appear simple: Just apply computerized algorithms created by great minds (such as decision trees, neural networks, genetic algorithms, etc.) and "poof", the results appear. The reality is quite different. BI is an iterative learning process that takes long-term hard work by highly trained people. BI requires new kinds of professionals that combine the skills of statisticians, business analysts and storytellers/artists to extract the nuggets of gold hidden under mountains of data. As Hal Varian, Google's chief economist, predicts [5] "the job of business intelligence will become the "sexiest" around [2]. Data, he explains, are widely available; what is scarce is the ability to extract wisdom from them". At the same time, Business Intelligence has not yet become a mainstream field at Business Schools and the shortage of Business Intelligence related professions will continue to rise.

The following literature review provided by Mrdalj [13] is also applicable for development of BI modules. Some of the first experimental undergraduate data mining course development is reported by Lu and Bettine [10]. A paper by Banks et

al. [1] describes teaching data mining courses as “an exciting addition to the curriculum at the senior/graduate level” that provides the opportunity to apply computer related education to various domains and applications. A very good tutorial on data mining that can be used by educators as an introduction to business intelligence is presented by Roiger [15]. The typical structure of the data warehousing and data mining course is well described by Fang and Tuladhar [7], as well as by Watson [16]. One of the teaching approaches by reading research papers and developing algorithms is reported by Musicant [14].

This paper presents the structure and components of a BI module that is designed to appeal to students of the Finance program as well as the MBA students selecting the Finance concentration. The BI module presented here is built around consumer creditworthiness as one of the common topic in banking courses dealing with consumer lending. Specifically, this module focuses on “business aspects” of BI without a very intensive study of mining algorithms. The main focus is on having students learn how to:

- Define the business issue
- Build a data mining model
- Interpret results
- Deploy the results

We use hands-on in-class lab activities to enable students to immediately apply presented concepts and to reinforce their understanding. Our desire that students identify with real-world problems and BI applications is achieved with business oriented assignments. The closest approach to the one in this paper is reported by Mrdalj [13] and Chawla [4]. That course is also motivated from the “real-world” usage of data mining and Mrdalj reports that his course received positive student feedback on the course structure and teaching methods very similar to ours.

The BI module presented in this paper is created to utilize the Microsoft Excel with Microsoft SQL Server [11] Data Mining Add-ins that provides a comprehensive set of business intelligence tools for reporting, query and analytics. There are many other tools that could be used for such a module like SAS Enterprise Miner [6], IBM/Cognos, Oracle/Hyperion, SAP BW/Business Objects, Micro Strategy, QlikViw, Tableau, etc. Our decision to use Microsoft Excel Data Mining Add-ins is based on the wide availability of this tool and the ease of its use for the business students through the Excel front end. The

comprehensive comparison of the available BI tools is outside of the scope of this paper.

CONSUMER CREDITWORTHINESS

One of the paper’s objectives is to illustrate the use of BI software in solving business problems. The field of banking relies heavily on BI, especially in credit assessment. This paper narrows the illustrative choice to consumer creditworthiness.

Consumer lending is a high volume activity that may entail large credit losses. One objective for value maximization is to control the credit risk inherent with granting credit. The underwriting of consumer credit typically uses various risk management systems that can be classified according to the applied methods as follows:

- lending officer’s *Judgment*,
- *credit scoring*,
 - *credit approval*
 - *credit line adjustment*
- *default/delinquency prediction*.

In general, the underwriting of smaller consumer loans has been automated while judgment has been applied to higher amounts. Very often, however, companies will use computer-based systems in conjunction with a judgmental system, with judgment overriding the computer’s decision if necessary as reported by Caouette et al. [3].

Judgmental Systems

In most instances these systems assess the applicants’ creditworthiness by analyzing characteristics such as income, job type, homeownership, and credit worthiness. For example, the lender may set the *criteria* along the following lines similar to the criteria listed in [3]:

1. At least *one year of residency* at current address; *proof of residency* required; *phone number* required,
2. At least *one year of current employment*;
For borderline cases, *three years* of verifiable *proof of employment history*; employment length *waived* for recent *college graduates*.
3. *Minimum monthly salary* of \$1,700 with *proof of income* for employees; several years worth of *copies of 1040 forms* for self-employed.
4. Maximum *personal debt ratio* of 50 percent including all debt obligation payments, and insurance premiums.
5. Clean credit bureau rating.

Credit Scoring Models

There are two types of credit scoring models:

- *credit approval* models and
- *behavioral scoring* models.

Credit approval models help determine whether to make a loan or not. They focus on variables very similar to the one used in the judgmental systems. Some of the variables used in these models are: rent or own, years at current address, income, marital status, number of dependents, profession, credit cards carried, credit bureau inquiries, number of inquiries in the past 6 months, number of adverse remarks on credit history, telephone number, and steady income level.

The *Behavioral credit scoring models*' purpose is to improve the profitability of accounts by adjusting the offered credit line or the collection methods for instance. They use criteria such as: the frequency of past due payments in the last year (30, 60, 90 days past due), number of negative items in the credit report, volume of transactions, and average line utilization.

These models use discriminant analysis or logic regression analysis to generate a *credit score* by assigning *statistically predetermined weights* to the applicant's attributes. The applicant is approved if the score compares favorably to some *benchmark value*. The underlying assumption of these models is that there exists some benchmark that can separate *good credit* and *bad credit* into *two distinct distributions*.

Default/delinquency Prediction Models

In these models, the applicants' characteristics are broken down sequentially from the most important to the least important by using the decision tree

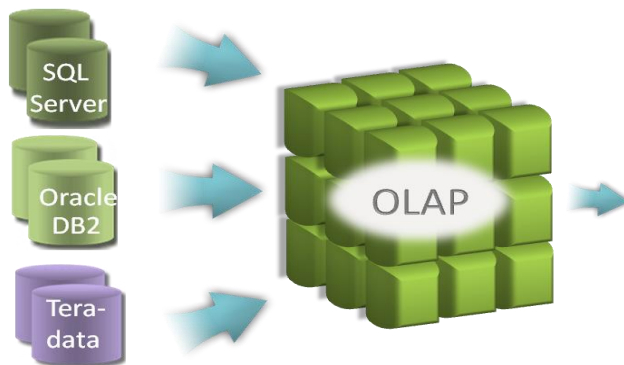


Figure 1. End-user access to OLAP Cube

- Predictive capabilities readily available for business users in Excel

approach. For example, the applicant population could be split into

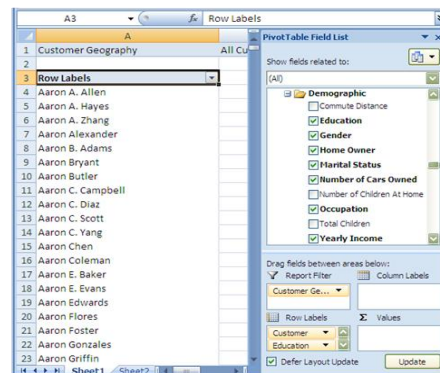
- *renters* and *owners*. These groups could be divided into:
 - *different income levels*, income levels could be subdivided into:
 - *indebtedness ratio levels*, that could be subdivided even into:
 - *number of years at the current address*.

The *end objective* is to make a decision based on the *probability of default/delinquency at each of the stages*. As tree branches are created the probability of selecting a bad or a good applicant is moved toward 0 or 1, thus increasing the chances of making the right choice.

ENABLING TECHNOLOGY

Advancements in technology such as Online Analytical Processing (OLAP) provide a platform for end-user based utilization of data mining [9]. *OLAP* tools, based on multi-dimensional views (such as loan, account, time, geography) give users flexible and fast access to measures (such as balance, payments, income) both summarized and detailed. Data representation is the form of an OLAP Cube which can be accessed by the end-user using tools such as the Microsoft Excel PivotTable [8] as illustrated in Figure 1.

The Microsoft SQL Server Data Mining Add-ins [11] enables Microsoft Excel to be the analytical client to the OLAP analytical servers. It enables easy to use predictive analysis at every desktop for every information worker through:



- Data mining client for building data mining models in Excel

- Data mining templates for project visualization in Visio

The BI module proposed in this paper uses the Data Mining Client for Excel whose ribbon shown in Figure 2 depicts the variety of available tools. It is the front end tool which uses the power of the SQL Analysis Services to perform all analytical processing.

Setting up a proper lab environment to use the Data Mining Client for Excel is a multi step process. First it requires installing the SQL Server. For the simplicity of the authentication process and performance purposes, we opted to install standalone copies on each lab computer. The next step is to install a sample data warehouse including OLAP Cubes. Unfortunately, deployment of OLAP Cubes requires the Microsoft Visual Studio project. Thus the pre-installation of the Visual Studio is needed as well as some familiarity of the SQL Server Business Intelligence Development studio. The last step is to

download and install Microsoft SQL Server Data Mining Add-ins. The last task of this process is to connect Excel to the SQL Server Analysis Service and to configure an analysis database to be used by Excel to perform data mining tasks.

Since the proposed BI module is intended for students of the Finance program as well as the MBA students selecting the Finance concentration, the current major challenge is to enable them to perform the above described installations on their personal laptops.

We hope that cloud-based database servers offering database as a service, like CloudOLAP and Microsoft SQL Azure (in the near future), would be willing to provide their service to the universities teaching BI courses. It would enable instructors to upload their example data warehouses and OLAP Cubes to be available to students wherever Internet access is available.

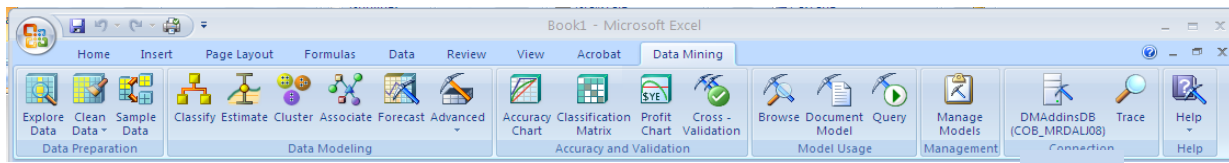


Figure 2. Data Mining Client Ribbon

BI MODULE OUTLINE

The BI module is structured around the following six steps of the BI life cycle:

1. Define the business issue
2. Determine the data mining task
3. Data collection, cleaning and transformation
4. Building the data mining model
 - a. Create mining structure
 - b. Add algorithm (model) to structure
 - c. Train model
 - d. Validate model
5. Interpret the results
6. Deploy the results (use model)

For illustration purposes we will use the following example from the Salford Systems consulting experience [2]:

A mortgage lender wants to be able to predict which of its customers would refinance their loans at any point in the future, because this operation usually leads the borrower to seeking and moving to another lender. With advance information the bank could attempt to keep

desirable borrowers by offering a reduced interest rate. The bank's data warehouse contains monthly information on each borrower such as employment, real estate, current interest rate, original interest rate and detailed credit bureau scores.

The first step concentrates on the business problem formation. It tries to answer questions like:

- What are the problems you are trying to solve? Possibilities based on the given example:
 - Give each borrower a refinance probability score.
 - Classify each borrower as desirable or undesirable.
 - Which borrowers' characteristics would predict his/her departure?
- How do you know that you need data mining?
- Do you have appropriate data to solve your problem?

The second step is one of the most important decisions that students need to learn to make. Based on the business problem, students need to determine

the appropriate type of the data mining task. Data mining tasks are usually categorized as:

- Descriptive
 - Clustering (example: customer segmentation)
 - Association (example: market basket analysis)
- Predictive
 - Classification (examples: credit risk scoring, churn prediction)
 - Prediction (examples: sales prediction, coupon redemption rate prediction)
 - Forecasting (example: future stock value)

The SQL Server provides nine algorithms for data mining, each suited for a different type of task [12]. Based on the Microsoft suggestions [12], Table 1 shows their possible usage based on the data mining tasks. For our example, if students would like to find out which borrowers' characteristics would predict his/her departure, they should select to use Decision Trees and Neural Network algorithms.

The third step obtains needed data by creating a Pivot Table as illustrated in Figure 1. In order to create a data mining model, students need to explore and understand obtained data. Exploration techniques include but are not limited to calculating, minimum and maximum values, mean and standard deviations, and looking at the distribution of the data. In addition, the Data Mining Client provides three tools inside the Data Preparation tab (see Figure 2):

- Explore Data
- Clean Data, and
- Sample Data.

Students are encouraged to use those tools to explore the data in the light of their understanding of the business problem and the data mining task. If students discover that the dataset contains defective data, they can formulate a strategy for correcting the problematic data.

Table 1. Data Mining Algorithm Selection

Task	Algorithms to use
Predicting a discrete attribute. For example, predict whether the recipient of a targeted mailing campaign will buy a product.	Decision Trees Algorithm Naive Bayes Algorithm Clustering Algorithm Neural Network Algorithm
Predicting a continuous attribute. For example, forecast next year's sales.	Decision Trees Algorithm Time Series Algorithm
Predicting a sequence. For example, perform a clickstream analysis of a company's Web site.	Sequence Clustering Algorithm
Finding groups of common items in transactions. For example, use market basket analysis to suggest additional products to a customer for purchase.	Association Algorithm Decision Trees Algorithm
Finding groups of similar items. For example, segment demographic data into groups to better understand the relationships between attributes.	Clustering Algorithm Sequence Clustering Algorithm

The building of the data mining model step starts by creating a mining structure which defines which data will be used. Once the mining structure is defined, one or many data mining algorithms needs to be added to it to determine the type of processing (see Figure 4). Training of the model refers to the process of applying selected data mining algorithms to the data and to the mining structure in order to extract patterns. Students may create different results by adjusting parameters to each algorithm and by applying filters to the training data to use just a subset of the data. Before students proceed, they need to test how well the model performs. Typically,

students are asked to create multiple models with different configurations and to test them to see which ones provide the best results for their problem and selected data as illustrated in Figure 3. The SQL Analysis Services provides under Mining Accuracy tab the following automated testing tools:

- Accuracy
 - Lift and Profit Charts
 - Scatter Plots
 - Classification Matrix
- Reliability
 - Cross-validation

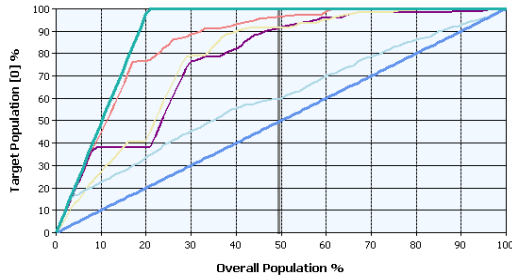


Figure 3. Sample Lift Chart

The fifth step requires students to interpret the results by answering questions like:

- Does it make sense?
- Does it provide insight or only obvious trivialities?
- Can it add value?

Several excellent model viewers/browsers are available under the Model Usage tab in Figure 2. In this example we illustrate this with a discrimination viewer shown in Figure 4. Students are especially cautioned to pay attention to the possibilities of discovering things that are not true and those that are true but not useful.



Figure 4. Sample tutorial slide

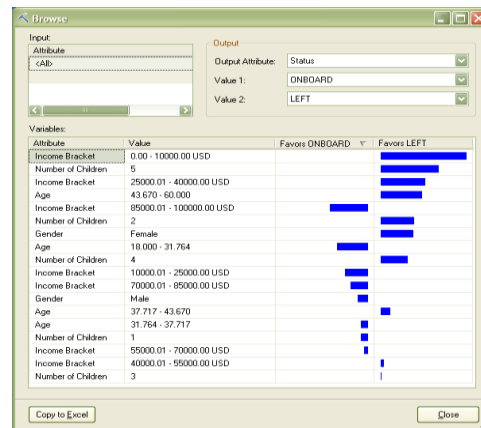
In-class Lab Activities: BI modules should be conducted in a computer classroom with access to the SQL Server and Excel with Data Mining Add-ins as described in the Enabling Technology section. Lab activities are designed to help students gain hands-on skills in using the Data Mining Client. Given the complexity of the tools, many concepts and procedures may not be immediately clear to the students. Therefore, we recommend that detailed in-class lab tutorials be developed. An example of the tutorial slides is shown in Figure 4. Using such slides, the step-by-step procedure on how to solve a given

The last step in this process is to select the model that performs the best and to deploy it into the production environment. The exact way of using the model depends on the tasks it is intended for and the production environment.

TEACHING METHODS

The structure of the proposed BI module blends the theory and practice based on the “real-world” usage of the data mining and data warehousing. The teaching methods that we suggest are based on the ones proposed by Mrdalj [13] and adopted to be used in this module as follows:

Lectures: Important material from the text and outside sources should be covered during lectures using PowerPoint slides. The slides should be available online as lecture notes. Students are encouraged to take careful notes as not all material can be found in the texts or readings. All concepts covered during the lecture portion should be reinforced by using in-class lab activities.



problem is explained first, and then the students should be given a business problem to solve. Such activities should help the students to understand the underlining concepts and how to apply them, as well as, to develop analytical and problem solving skills

Homework Assignments: Periodical individual homework assignments should be assigned to further reinforce the BI discovery process and to encourage the use of the supplemental material outside of the text. Such assignments may include experimentation using various available demo OLAP cubes.

Class Project: A group project which involves building a business intelligence application is gradually assigned in parallel as the BI techniques are discussed. The students are asked to divide into groups that consist of 2-3 members. Each group needs to define, develop and implement its own business case that includes the following:

- Identify the business opportunity and define the problem
- Browse available OLAP cube and evaluate the data
- Select OLAP analysis and/or data mining techniques
- Extract and preprocess the data
- Evaluate the data
- Perform OLAP analysis and/or build data mining model and perform data mining
- Interpret and translate the mining results into “actionable business strategies”.

The students are required to use several data mining techniques in creating their projects. Several sample project strategies are presented in class and fully discussed with students. Each group is required to submit a 1-2 page project proposal with their business problem, project strategy and what potential results they expect to accomplish. This allows the instructor to provide early feedback to the students and in some cases to discover major flaws in the proposed project strategy. Beside the final project report, the students are required to do in-class project presentations. These presentations are followed by a peer-review discussion among students and the instructor.

CONCLUSIONS

In this paper we report on the development of a BI module to be used in Finance courses dealing with customer creditworthiness. It is based on the ability of students, as end-users, to access the data warehouse using Microsoft Excel as a front-end to the OLAP Analytic Server. Such technology brings high power data mining tools to the fingertips of students, thus providing them some valuable, but currently lacking, experience in the use of BI tools and their applications.

Based upon the experience and lessons learned from teaching a finance course using the proposed BI module, we plan on developing a comprehensive set of similar modules for other courses whenever appropriate. At the same time, we would like to point out that the preparation of such BI modules is extremely time consuming.

REFERENCES

1. Banks, D. L., Dong, G., Liu, H., and Mandvikar, A. (2004). Teaching Undergraduates Data Mining Engineering Programs. *Proceedings of the 34th ASEE/IEEE Frontiers in Education Conference*
2. Berry, M.J.A. & Linoff, G.S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd Edition, John Wiley & Sons, Inc.
3. Caouette, J.B., Altman, E.I. & Narayanan, P. (1998). *Managing Credit Risk: The Next Great Financial Challenge*. John Wiley & Sons, Inc.
4. Chawla, N. V. (2005). Teaching Data Mining by Coalescing Theory and Applications. *Proceedings of the 35th ASEE/IEEE Frontiers in Education Conference*, Indianapolis, IN, S1J-17-23.
5. Cukier, K. (2010). Data, data everywhere: A special report on managing information, *The Economist*, 394(8671), February 27th 2010.
6. Dickey, D. A. (2005). Teaching Data Mining in a University Environment, *Proceedings of the SUGI 30 Conference*, Philadelphia, Pennsylvania, 1-10.
7. Fang, R. & Tuladhar, S. (2006). Teaching data warehousing and data mining in a graduate program of information technology, *Journal of Computing Sciences in Colleges*, 21(5), 137-144.
8. Harts, D. (2008). *Microsoft Office 2007 Business Intelligence: Reporting, Analysis, and Measurement from the Desktop*. McGraw-Hill/Osborne
9. Larson, B., (2009). *Delivering Business Intelligence with Microsoft SQL Server 2008*. McGraw-Hill/Osborne
10. Lu, Y, & Bettine, J, (2003). Data Mining: An Experimental Undergraduate Course. *Journal of Computing Sciences in Colleges*, 18(3), 81-86
11. MacLennan, J., Tang, Z.H. & Crivat, B. (2009). *Data Mining with SQL Server 2008*. Wiley Publishing, Inc.
12. Microsoft TechNet (2009). *SQL Server 2008 Books Online*. <http://technet.microsoft.com/en-us/library/ms130214.aspx>
13. Mrdalj, S (2007). Teaching an Applied Business Intelligence Course. *Issues in Information Systems, Volume, VIII(1)*, 134-138
14. Musicant, D. R. (2006). A data mining course for computer science. *ACM SIGCSE Bulletin*, 38(1), 538-542.
15. Roiger, R. J. (2005). Teaching an introductory course in data mining. *ACM SIGCSE Bulletin*, 37(3). 415-421.

16. Watson, H. J. (2006). BI and Data Warehousing in Universities. *Business Intelligence Journal*, 11(3), 4-6.