Training Business Intelligence Applications: From Pedagogical Reasoning to Business Practices

Dimitar Christozov, American University in Bulgaria, dgc@aubg.edu
Alexandr Bronasco, American University in Bulgaria, avb110@aubg.edu

Abstract

In this paper, we share the experience obtained in offering a pair of courses related to Application of Business Intelligence (BI) Techniques. The first course addresses mostly technological aspects of building BI applications such as Data Cube modeling, ETL process and Data Visualization, while the objectives of the second one were in developing understanding regarding complexity in implementing BI solutions in a particular institution. Possessing domain knowledge is essential for both courses and selection of the registration processes of the university as a training case is the natural one. Exploration of this case in class resulted in a well-defined and justified approach of how to make some of the decisions through this process rational and data driven by using well-known analytical techniques.

Keywords: Business Intelligence, Business Analytics, Enterprise Information Management Maturity Model (EIMMM), Course Registration, Customer Relationship Management (CRM)

Introduction

Business Intelligence (BI) or Business Analytics (BA) are relatively recently established areas of computer and information support of managerial practices. Usually, such tools are applicable to resolve the problems of managing by utilizing accumulated large amounts of heterogeneous data - Big Data. Introducing students to theory, technologies and practices associated with the process of designing and implementing Business Intelligence applications requires selecting a case for investigation, which must allow applying methods relevant to the variety of activities associated with this process. The major requirement, and also the major constrain, is the need of domain knowledge. Under those conditions, the course registration process is the natural choice as a class project. This business process involves several categories of stakeholders: administration, faculty (individuals and departments) and students. Administration and faculties are making decisions in the processes of selecting and offering courses for a given semester, and students in selecting which courses to register to graduate on time by following a smooth road-map over the four years of study according to their preferences, priorities and attitudes. The process itself simulates market relationships of demand and supply. Offering courses that don't attract enough customers (students to choose them) and that may be dropped, creates problems to sellers (the faculty who offer them) and to customers (students who have already registered to them). For the institution, as a whole, better utilization of available resources (facilities and faculties) is of critical importance for effectiveness.

BI is a combination of technical tools, data processing algorithms and business practices. Institutional readiness is the critical success factor in implementing BI services. The American University in Bulgaria (AUBG) is applying the liberal-arts model of education. It is a small, campus-based, institution and the common assumption, that people know everything, dominates
on all units and all levels across administration. This assumption prevents considering BI or BA tools as worthy to implement in a way to support decision-making. Hence, AUBG is a typical candidate for a real institution with significant resistance toward adopting BI, because administration has to admit that decisions made are not entirely rational or based on objective facts. This results in selecting an appropriate methodology to investigate the level of institutional readiness and to adopt realistic implementation strategy.

Looking on BI from Knowledge Management perspective allows distinguishing three categories of applications:

1. Learning from the past/history allows avoiding mistakes and preserving knowledge;
2. Adequate reaction on current state – minimizing the effect of hazards in real-time behavior; and
3. Prediction – evaluate the exposure of potential events in a way to adopt measures to avoid negative development.

Intuition or accumulated expertise provides unstructured knowledge regarding demand and supply of courses and it is a valuable resource, but BI application may allow reducing uncertainty in several ways:

1. Exploring the historical data enables structuring the accumulated knowledge in a way to make it actionable. BI application will allow preserving the knowledge and disseminating it among participants in the process of offering and registration, and will exclude dependence on “knowledgeable people”.
2. Exploring the available data describing the current state (structure of students’ body) will allow adequate reaction (supply) to actual needs (demand) and will:
   - increase students’ satisfaction by facilitating a smoother education process;
   - improve utilizing the available resources;
   - reduce cost.
3. Assessment of tendencies will enable predicting the direction of changes in students’ interests and will allow reacting on time to them.

Within this framework, the objectives of the course were set up in the following way:

1. To investigate institutional readiness and, according to the findings, to develop strategy of how to implement BI services.
2. To define a collection of services by specifying:
   a. Objectives, including stakeholders addressed, and especially stakeholders’ benefits;
   b. Data sources: these include investigation of available information resources and what information is needed to respond to the needs of a given service. We distinguished between "reactive" services, to which existing operational systems provide the needed data; and "proactive" - services that need data, which currently is not collected or not recorded in a structural form.
   c. The ETL process for every data source
3. To structure the Data warehouse and needed OLAP cubes;
4. To define user interfaces
5. To choose technology for implementing the service
The paper is organized in several sections. The first one describes the business processes, roles of participants, and problems, which they are facing. Naturally, the emphasis is given to the group of students. Second section presents the selected methodology and organization of the project's development. Further, we present the findings in investigating the cases. The proposed solutions are presented as a business case and implementation of several experimental applications to illustrate the benefits of applying BI in these processes. Conclusion summarizes experience as pedagogy, but also discusses how the findings in this course can serve as a starting point in development of BI applications.

**Setting the stage: Business process and stakeholders**

Two components define an educational model as "liberal" - significant component of general education and free elective courses, and liberty in choosing individual path in taking courses in a way to obtain the needed credits to graduate in a desired major. Support of the students' choice of courses is what we emphasized in these BI courses. The three categories of users - students, faculty and administration - play different roles in this process and need different support. The process can be summarized as the following steps:

1. Faculty departments suggest courses to be offered in the next semester, including the faculty to teach them, preferred time slots, and other details - whether the course has to be offered in a regular classroom or a computer laboratory, maximal number of places, etc.;
2. The registrar's office reviews the proposals, summarizes the offers, resolves the conflicts in using facilities in cooperation with the departments, and develops the draft schedule.
3. The schedule is reviewed by the Dean of Faculty, and by students via Students Government.
4. Students submit requests for changes to the Dean, and the Dean, after consultations with the departments, finalizes the schedule.
5. The list of courses are included into registration systems, which also include all conditions to register as prerequisites, standing, etc.
6. A student has to decide which courses to register by considering also the time the course is offered and also whether there will be place in the desired section of the course. Selection order depends on students' standing, more senior students choose earlier. When places of a given section are fully occupied, a waiting list is organized.

To help students in their decisions, the institution of Faculty advisor is launched. Advisors are faculties who meet students, discuss their road-maps and help them in resolving problems and conflicts in the selection of courses. Despite this, the majority of students, as our research shows, use and rely more on peers' advises, which is typical for a campus-based institution where all students are living together.

This process simulates real market relationships in course registration, driven by demand - students’ interests, and supply - the courses offered. Applying principles of CRM may lead to improvement in the effectiveness of the process and the entire institution as well. The objectives of improvement are, first of all, students' satisfaction, via offering a smooth way toward graduation, but also efficient use of resources and creating efficient training conditions. Therefore, the objective of the CRM in the registration process is to offer courses to satisfy in the
best possible way students’ road-maps by clustering demand to fit the boundaries of organizational effectiveness - the number of students in class. The right number of students in class, according to pedagogical reasons, is considered to be the main goal. To administration, this means that resources - facilities and faculties - are used in the best way and that may result in reducing the cost for the institution. For faculty, the right number of students allows effective training and efficient application of the selected pedagogical approach - too many students create problems with grading and directing individual advancement, too few students create a risk that the course will be dropped, but also make some of the pedagogical techniques inapplicable. The benefits for students include training effectiveness, but more important, the possibility of the real threat that there will be not enough places, or the course will be dropped, because of few registered, will be minimized.

**Methodology**

The course started with a brief introduction to the basic components of BI, its building blocks, benefits to stakeholders. In general, this is a quite abstract and theoretical part. Special attention was given to different models of maturity of the institutions from the BI adoption point of view. In general, we chose to follow the Gartner's EIM MM (see Newman D., Logan D., 2008). Students were asked to assess AUBG maturity level. The majority assessed AUBG on stage "Reactive" and few "Aware", without being able to argue their judgment. Instead of discussing arguments, students were directed to collect first-hand evidences via interviewing different categories of stakeholders.

The class was divided into three groups, which interviewed the three categories of stakeholders:

- administration, including the President, the Provost, and staff from the offices engaged in registration - registrars and admission offices;
- faculty, including members of different departments and also department chairs;
- students with different standing.

The objectives of the interviews were twofold - to assess the level of awareness, following Gartner's criteria; and to highlight the problems faced by different categories of stakeholders in relation to offering and registering courses.

After performing the interviews, the overall assessment of the maturity level became "Unaware". This result defines the objectives of the class project - developing materials for informal training to promote BI by illustrating the benefits for different categories of users and to develop scenarios for solving the identified problems.

To illustrate the relevance of the proposed solution, we used historical data - registration and grades for the first 12 years of the university from Fall 1991 till Spring 2003, having in mind that there was not a full range of courses during the first three academic years. These data, of course, were anonymous - students’ names were removed and even ID# were modified, but still allow showing what a given student has done.

In the next sections, we discuss the identified problems and the solutions.
Findings

The second goal of interviews was to identify those of the problems in the registration process that are suitable for BI support. Formally, we separated the addressed problems in two categories:

- reactive - problems that can be solved by exploring available data sources;
- proactive - problems, solution of which needs additional data.

There are some problems that allow exploring firstly the "reactive" approach, to improve decision-making by use of available information, and then further extend the decision support by including newly launched information resources.

The following two problems were addressed:

A. Support faculty in offering courses via prediction of the demand. There two separate problems:

1. Which courses, and number of sections, to offer in the semester? The intuitive knowledge and experience accumulated in the departments allows solving this problem quite successfully, especially for upper-level courses with established student groups that passed prerequisites. For example, 20 students have passed Discrete Structures course in Fall 2014, which means that for Spring 2015 it is expected that about 20 students will register for the Fundamental Data Structures course, thus, offering one section with 25 seats will satisfy the demand, even if there are some candidates who have postponed taking the course from the previous year, and also it is not likely that the majority of those 20 students will decide to postpone it.

2. In which time slot to offer a particular section? Within a department there is understanding which courses are targeted by the same groups of students, but there is no way to manage time conflicts for courses offered by different departments. Resolving time conflicts has to be completed by the Dean, who needs support in assessing the exposure of this risk, based on historical data adjusted to the current structure of the student body.

B. Support students in selecting courses to register. Many problems were highlighted by students, but the following were recognized by the class as really important:

1. Lack of sufficient information about the course in advance. The information from the academic catalog provides only official information. Students cannot assess the real amount of efforts and time they have to dedicate to pass the course with the desired grade. They make their judgment based on the opinion of peers, which is not always objective and not always reflects the given student’s potential, interests and personality.

2. Time conflicts. Individual student's road-map requires that a student registers to a particular set of courses on this particular semester, not necessarily in the same major,
or offered by the same department. The time conflict appears when two of the desired courses are offered in the same time slot of the weekly schedule. In this case, the student has to choose only one of them, which may lead to disturbance in acquiring all the desired credits and even not graduating on time.

3. Availability of courses. The course, the student plans to take in a given semester, may not be offered; or there are no places for the given student to register for the course. The system of "waiting-lists" exists, but the probability that the university will be able to react to unexpected demand is really low.

4. Composition of the class. Psychological compatibility of students in class is among the most important factors for establishing an atmosphere for effective training. This process is naturally driven by exploring informal methods, but sometimes it is important for a student to know that somebody from his personal "black list" has already registered for the course.

5. Road-map projection. Registering for a course requires adjustment of student's individual road-map. Often, students do not follow a well-defined sequence of courses to take, but such information is actually beneficial since it shows them the effect of the decisions they made.

**Solutions**

In approaching problems, described in the previous section, we considered firstly the available sources of information. We distinguished three categories of information needed to resolve the problems:

- well-structured and maintained databases provided reliable data;
- available, but not well-structured and organized data;
- unavailable data.

Based on this analysis, we separated problems in two categories:

- Reactive: problems that can be solved only by the available well-structured data sources, with a relatively clear ETL process.
- Proactive: problems that need either new operational information system to gather the needed data, or a sophisticated preprocessing to make the available data usable.

For some problems, as forecasting of demand, there is a reactive solution based on historical data, but there are also a couple of proactive approaches aimed at improving the precision of the forecast, which were described in the business case as stages in the development of the BI support.

**Reactive Solutions**

1. Supply based on demand. The reactive solution explores historical records. Tracing the dynamics of offering the course by semesters, including the dynamics of the structure of student population. Structure of the student population is presented by clustering students’ profiles. Profiles, described by demonstrated interest in a major (from admission database for new students, and by registered major for junior and senior) and
by history of already taken courses. This allows defining every student's individual road-
map by mapping major's road-map to student's history. This enables defining a distance
and applying cluster analysis. In principle, the history of the offering is the current base
for offering courses. The BI component allows assessing the probability that a potential
candidate, belonging to the particular category, will register the course. This approach
was thoroughly discussed in class and the brainstorming bared two proactive and more
reliable solutions.

2. Resolving Time slots Conflicts. The solution explores data frequent pattern analysis of
historical data. This is a typical "basket-analysis" problem. After applying apriori
algorithm (the one described in Han et al., 2011 was actually implemented) to old
registration data, the measure "support" for association rules given in the form {student
who registers course X also registers course Y in the same semester; support N} shows
the potential number of affected students if the courses X and Y are offered in the same
time slot. The intuitive solution applied recently makes similar analysis locally - by
offering within departments. There is no instrument to identify cross-departmental time
conflicts and this simple and well-known BI solution allows exposing the impact of time
conflicts and facilitates improvement in the quality of offering.

3. Information regarding courses. Individual assessment regarding a particular course is
usually subjective. Information regarding common students' assessment of the course is
recorded in registrar's office in students' evaluation of the course. This information is
confidential and used to evaluate performance of the faculty. Possible use of this
information is via extracting a summary of evaluations, but only from those students,
who belong to the same cluster of students' profiles. In this way, only the opinion of
students with the same status, including similar grades in similar courses and standing,
will be provided, which is more reliable than the opinion of a randomly selected peer.
Additionally, for repeating courses, recent syllabi can be provided to inform students
regarding course difficulties, out of class time needed, etc. All this information is
recorded, but in different forms, and needs careful processing to resolve the
confidentiality problem, to implement technologies for processing languages to extract
semantic information from different texts (see, for example, D. Karagiozov, 2015) , and
to maintain students' profiles and cluster analysis.

4. Road-map Projection. Any decision of choosing a course in a given semester results in
adjusting individual's road-map. The liberty students have in selecting courses allows
them to choose not only "must take" courses, but also courses for fun or courses to extend
understanding in areas far from the selected professional area. In all those cases, the
student must keep balance in courses to graduate on time. In addition, not all courses are
offered every semester, thus, missing to register a needed course may influence the
desired schedule toward graduation on time. Departments have developed "major's road-
maps" with the recommended sequence of courses leading to successful graduation on
time; also, there is information in the academic catalog about the frequency of offering a
course and the usual semester when the course is offered. Combining this information
with what the student has already accomplished, and with the selection of courses for
next semester, it is possible to generate scenarios leading to completion of graduation
requirements. In the earlier years (freshman and sophomore standing), such scenarios
could be too many, but it is important to inform the student that with a given selection,
graduation on time is possible and explain the strategy she or he must follow in the next semesters. The student may choose one of the road-maps to follow, and his or her choice may provide additional information to departments regarding demand.

**Proactive Solutions**

1. **Demand.** Two alternative approaches were discussed in class:
   a. **Early preregistration.** Organizing an early (not later than during the third week of the semester) preregistration with offering all courses from the catalog may accumulate information regarding actual demand. The major problem with this approach is how to motivate students to participate, knowing that this will not guarantee them that the pointed courses will be offered in reality.
   b. **Having an individual student’s scheduler** (see Shehi R., 2014). Regi Shehi, motivated by in-class discussions, added particular functionality in the application developed as his Diploma Thesis. Smartphone usage is common nowadays among students, and he developed an application, which reminds students about all scheduled events in one’s study, as deadlines for homework, quizzes, exams, submissions etc. Regi added the registration component that allows the student to specify not only his or her time for registration (missing this time may result in missing the opportunity to register desired courses), but also to specify the targets in registration for next semester. He added functionality, which allows extracting this information from individual devices and makes it available to registrar’s office. Setting student’s goals in registration early enough makes it possible to collect objective information for expected demand.

2. **Information regarding courses.** Combining the "re-active" solution with shared students’ unofficial information may augment reasons for selecting or postponing the course. Information regarding how many out-of-class hours per week a successful student dedicates to the course, and good practices in studying and preparing, is usually shared among students by using social networks as Facebook. Extracting, structuring and summarizing such information may allow improving the judgment, extending the assessment of complexity to the entire composition of courses for a given semester, and balancing the weekly workload. Applying student's profile to the grades within the cluster may enable even to predict what could be the grade of the student under standard circumstances for this course and to improve the ground for what the student may expect as the outcome.

3. **Composition of classes.** This was the most controversial issue discussed in class. On one hand, there was a mutual understanding that psychological compatibility among students may create an atmosphere, which stimulates learning and may result in better achievements. On the other hand, how to achieve this? The approach discussed was to add more attributes to the student's profile, as interests, habits, friends lists and/or "black lists". The idea is to allow students to see whether somebody from these lists is already registered for that course and to make the decision based also on this information. Interests can be associated with membership in student clubs or participation in student activities. Assuming two students who participate in the same set of activities have
similar interests in general. Technology for face recognition may apply to identify students in the events holding the only visual documentation.

The range of shared ideas was unlimited, but the major objective was to select only ideas, which are beneficial, mostly to students, with the assumption that all those ideas may also serve to improve utilization of available resources, mostly facilities and faculties, and to improve in general the distribution of students in the courses, making it closer to uniform distribution. In general, this will lead to minimizing costs and reducing exceptions, like last minute course cancellations, which create problems to both students and faculties.

Scenarios

Figures 1 and 2 illustrate possible scenarios in visualizing some of the ideas shared in the previous section. The first one illustrates the scenarios when a student logs in to the registration system.

- The student is shown the list of available courses colored according to his or her roadmap. For example, red is for courses the student must register this semester; green courses are just recommended;
- When pointing one of the courses, an information window opens up, showing well-structured information regarding the course, including predictions for the workload and grade. Information regarding already registered students may be provided as well. To avoid confidentiality issues or even discrimination, this information is presented in a form of a warning of potential personal conflicts.
- When selecting a course, an accumulated expected weekly workload is calculated.

Figure 1. Support for students at the time of choosing courses for registration
The second figure illustrates what is generated after selecting all of the courses for the semester:

- All possible road-map scenarios are generated, filtering and showing only those of them that allow graduation on time;
- In case there are no road-map scenarios that allow graduation on time, a warning message is issued;
- After final approval of the selection, the courses are recorded in the student’s profile.

Figure 2. Data processing after confirmed choice

Discussion and Conclusion

There were two major outcomes of the way this course was designed and implemented. The first one is pedagogical. The theory of constructivism (see Fosnot, 2005) was leading in designing the course. Students previous knowledge and experience in course registration practice was assumed as essential in training application of Business Intelligence. The eclectic implementation of Active Learning (see Bonwel and Eison, 1991) and Goal-directed training approaches were applied. The course combined:

- Learning by Discovery - students were asked to investigate institutional readiness by adjusting particular methodology (Gartner’s EIM MM) to discover how to assess it, as results students were motivated to discover how to improve the registration process; and
- Learning by Doing (see Christozov, Galletly, Karagiozov, Bonev, 2007). The process combined brainstorming, for generating ideas how to apply BI techniques in improving the process, with actual implementation of some ideas on available historical data.
Engaging students in research with no preliminary known results, and later reflecting their findings within the training process proved to be a highly effective pedagogy. Their understanding that the work done in class may really improve the system and may lead to better operation ensures high motivation. This proves conclusion of several researchers (see for example Rusbult, C., 2007) that the Constructivism combined with Active Learning techniques is an efficient educational practice.

The discovered problems and proposed solutions developed within the class discussions were real. Implementation of applications by following the proposed solutions is possible and even feasible, in most of the cases. More careful cost-benefit analysis is needed to justify the actual implementation of the proposed BI solutions for AUBG, but the stage for identifying and pointing out how the current systems can be improved by real data-oriented decision making is a significant step forward in this direction and initiate the long-way process of building understanding related to application of BI in the university. The benefits for all constituencies within the university include higher students' satisfaction, less dropped courses, and probably, better utilization of available resources - facilities and faculties.

The work done as class project was fully supported by the three groups of stakeholders. The President, the Provost and many officers found time to speak with students, nevertheless that their opinion is that the university does not need such tools to improve operations. The documents developed within the class project - description of the problem, analysis of information resources needed, business cases, and suggested, in some cases even implemented, software solutions may serve as an excellent instrument in "informal training". Training that will lead to improvement in understanding among different constituencies that applying BI approach may help in improving decision-making with a firm foundation on rationality and actual data.

The initial goal was to engage students in learning, but in the process of acquiring theories and developing understanding and skills about how to apply them, it turns to go beyond pure pedagogy. The course was supplemented by a series of workshops on different technologies. Applying them to explore available historical data, makes it possible even to illustrate that implementation of some of the proposed solutions is not difficult, and to place them in production is just a problem of awareness. This was the most significant outcome of teaching this course.

Acknowledgement

The course was designed and implemented within the project BG051PO001-3.1.07-0001- ,,Updating Computer Science Curriculum According to the Job Market.

References


Authors’ Biographies

Dimitar Christozov is a Professor of Computer Science at the American University in Bulgaria, since 1993. He has more than 35 years of research and education experience in areas as computer science, applied statistics, information systems. His recent interests are in the field of business analytics and data science. He graduated Mathematics from Sofia University “St. Kliment Ohridski” in 1979. He completed his Ph.D. thesis “Computer Aided Evaluation of Machine Reliability” in 1986 and D.Sc. thesis “Quantitative measures of the quality of informing” in 2009. At CIME (1979-1986) Dimitar Christozov was engaged in numerous software development projects. His major contribution was design and development of "Relia-Soft" for different platform widely used by Bulgarian industry. At ICTT “Informa” (1986-1993) Dr. Christozov was involved in establishing the national information network for technology transfer and research in the areas of technologies assessment, integral quality measures and information systems for quality management. In these areas he was recognized as one of the leading experts in Bulgaria. At the American University in Bulgaria, he was the leading person in curriculum development, launching and development of the majors of Computer Science (1993) and Information Systems (2008). At the University of Library Studies and Information Technologies (since 2002) he proposed and implemented the major of Information Brokerage. Professor Christozov has more than 100 publications as separate volumes, journal papers and papers in refereed proceedings. He is a founding member and fellow of Informing Science Institute and chair of Bulgarian Informing Science Society; founding member and fellow of Applied Knowledge Management Institute and founding member of the Bulgarian Statistical Society and the Bulgarian Telework Association. He had chaired and serve as member of Program and Organizing Committees of many international conferences and other events.

Alexandr Bronasco was born and raised in Balti, Republic of Moldova. He received his high school diploma in 2010 and, being awarded the Open Society Institute Full Scholarship, headed towards university education at the American University in Bulgaria. In May 2015, Alexandr is going to graduate with Bachelor of Arts degree in Information Systems and Bachelor of Arts
degree in Business Administration. While studying for his degree in Information Systems, Alexandr came across the term Business Intelligence and found it extremely interesting. He is fascinated how Business Intelligence is helping the business world make the right decisions and explore the before unnoticed connections among business data. He is using all the available resources and time to create detailed knowledge and expertise in the Business Intelligence software solutions and projects. Alexandr’s academic journey has just started. He is currently applying for a Master degree focusing on Business Intelligence. Alexandr Bronasco is an energetic young man, passionate about sports and geopolitics. He loves new challenges and experiences, and is always ready to learn something new.