MCMS

Final Report

Author(s)
Samia Oussena

With contributions from the MCMS project team

Created
29-Jun-10

Version

V1.0

Contents

Figures

Document Control

Current version

VERSION	DATE	AUTHOR	COMMENTS
V 1.0	21/04/10	Samia Oussena	Initial version

This document is the final version. The document will be submitted to for review and feedback from JISC before agreement on wider circulation.

Acknowledgements

The project team is pleased to acknowledge the support of those who have contributed to the project by providing their time, energy and knowledge in helping us deliver this report.

Email Contacts for further information

Dr Samia Oussena: samia.oussena@tvu.ac.uk

Executive Summary

Universities are under increasing pressure to improve the quality of the student experience. Issues such as competition between institutions, increasing student numbers, decreasing resource, student retention, and standards of student attainment on entry, lead to concerns over how students can be best supported throughout their programmes of study. Many institutions are seeking effective ways of monitoring student performance and of providing timely student intervention measures. This poses several practical problems including how to detect when a student shows indications of (in)activity that could lead to problems and how to engage with students in an effective way. In MCMS project, we propose to engage with students using an intervention application. A gaming metaphor was used in the intervention application. Data mining is used to process historical data and to detect problem indicator rules. The rules are then executed against live student data to generate personalized intervention events. Students engage with the events using a scoreboard system such as that used in on-line gaming, they can compare their scores against their peers and have clear indicators as to how their scores can be increased. Tutors also receive events as a consequence of changes to student behaviour, this allows them to make effective use of resources when supporting students.

The project used one of the most commonly used methodology CRISP_DM (Cross Industry standard process for data mining) methodology (CRISP_DM). For the data integration phase, we investigated the use of model driven approach, in order to automate some of the activities involved in producing an integrated model of the data sources.

The data mining process is used as two main functions; primarily to predict those students who are most likely to drop-out early, and secondly to group students into specific categories that can be targeted with personalised interventions if it is predicted that a drop-out is imminent. A prototype intervention application was built that integrate the data mining process with the intervention system. Game metaphor was applied to encourage student to actively engage. Potential problems of students are identified as early as possible, and a follow up with tailored intervention options. Tutors are also notified and supported during the intervention cycle.

Arising from this work we make the following recommendations.

- Projects need to pay careful attention to the planning and synchronisation with teaching schedules where experimental evaluation is based on live course schedules.
- Our experience suggests that any project requiring piloting an application should have a dedicated work package for managing the uptake of the application by the users.
- Business Intelligence can be part of a decision support system. The main application of Business Intelligence has been to support an institution reporting activities. In this project, we have extended the use of Business Intelligence to be embedded in the intervention application. Hence directly supporting the decision making process.
- Student engagement and motivation can be part of a retention strategy. Our experience suggests that data mining can be used to give an indication of student engagement. The benefitting factors of using a gaming environment in order to engage and motivate students in terms of retention need to be explored further.
- In MCMS project, we have developed an integrated data model for supporting student retention. It will interesting to see if the model is applicable to other institutions.

The MCMS Project Team comprised the following:

Thames Valley University	Dr Samia Oussena Prof Tony Clark Dr John Moore Hyeonsook Kim Dr Ying Zhang Maggie Stephens Dr John Waters
University of Birmingham	Dr Mark Lee

1 Background

Challenges that many universities need to address can include: students drop out rates and students taking longer to finish their degree. In the UK, according to the statistics published by HESA (Hesa 2008), the projected average percentage of full-time first degree starters expected to complete a degree is 77.4% (2005/2006) and 14.1% are expected to get no qualification. Therefore, an effective monitoring of student performance and providing good quality and timely intervention measures that will help improve student academic performance is essential.

Data mining is the field of discovering of implicit and interesting patterns for large data collections (Klesgen &Zytkow 2002). Data mining has been applied to a number of fields including bioinformatics and fraud detection. In recent years, there has been an increased interest in the use of data mining to the educational setting. Data mining has been shown to help predict student educational outcomes (Ronero & Ventura 2006). When using Data Mining, the goal is to develop a model which can infer an aspect of the student academic outcomes, such passing a module, from a combination of other data that represents student's characteristics. For example, Garbrilson (Garbrilson 2003) uses data mining prediction techniques to identify the most effective factors in determining student test scores. In (Minaei-Bidgol et. al 2004.), authors use a data mining classification technique to predict student's final grades based on their web use. In most of these applications, the results are usually presented to strategic decision makers in graphical form (for example a dashboard), which is interpreted and some intervention programme is then put in place. However the integration of the predictions provided by Data Mining, the presentation of the results produced by applying predictive patterns to live student data and the intervention processes are typically weak. In most cases intervention is provided by humans resulting in an overall process can be very resource intensive. There is therefore a lack of integration between the identification of the problem and the implementation of the intervention actions.

1.1 Aims and Objectives

The aim of MCMS project is to integrate the data mining process with the intervention process. We developed an application that allows the data mining models to be refined as a consequence of intervention actions, as well as involving students in the process. There is evidence that personalising and signposting educational 'moments' contributes to a better learning environment (Harvey, Drew et al. 2006). Although the literature on retention points to the complexity of factors influencing retention, there is evidence that linking social and academic experience, and tailoring the learning environment to individual needs increases an institution's chances of retaining its students (Anagnostopoulou & Parmar 2008).

The challenge in designing the application for the student retention has been how to maximise student engagement. Arguably the weakest link in the process arises in ensuring that students at risk are identified as early as possible. Of course, this can be achieved with unlimited resources in the form of tutors who continually monitor raw data sources and who contact students as soon as they detect a problem indicator. However, this is not feasible and our proposal is to automate the process and to present information to students in a way that will maximise their engagement and therefore reduce the resource burden.

The strong widespread appeal of computer and console gaming has motivated a number of researchers to harness the educational potential of gaming (Decastell & Jenson 2003). Here, we have looked at using the motivation power of games to encourage students to be involved with the intervention system.

In MCMS, we have designed a prototype system that has the following features:

- we have designed a uniform data model describing a student profile within a teaching and learning environment;
- historical data from a UK Higher Education Institution is processed by merging and transforming in order to fit the model;
- data mining techniques are then used to process the information and to produce rules that represent indicators of failure within the educational process;
- the rules are then run against live student data in order to raise potential indicators of failure in real-time; gaming systems have been analysed in order to produce a model whereby information can be presented to students as though their learning experience is a game; this

model has been implemented in the form of a web application and the events produced by the rules are fed into the gaming model.

2 Methodology

This section describes in detail, the methodology we have adopted to address the delivery of the aims and objectives stated above. The methodology that was used is the CRISP_DM (Cross Industry standard process for data mining) methodology (CRISP_DM). This most commonly used methodology. Luanreports that the most successful data mining projects comply with guidelines and steps in the CRross-Industry Standard Process for data Mining (Luan 2002).

Here, we elaborate each of the phases:

- **The Business Understanding Phase** This is an investigative process. The main techniques that was used in this phase interviews. The objective of the investigation is to articulate problem areas that we will be addressing in this project in form of questions relevant at the department, school, faculty and institution level. We undertook interviews with all the main stakeholders, these include the owners of the data sources such as library and VLE team and the academic processes owners such as the academic deputy vice-chancellor, faculties executives, student dean.
- **The Data Understanding Phase** This phase requires the reverse engineering of the data sources. All the data sources have been modelled in UML Diagrams. Following from the data modelling of these systems, we then proposed a unified schema in a UML class diagram.
- **The Data Preparation Phase**: Following from the previous phase, scripts and triggers were developed for the migration of data from the different sources to a unified repository. The repository has implemented in Oracle 11g.
- **The Modelling Phase** This is the phase, where data mining algorithms sift through the data to find patterns and to build predictive models. We used both supervised and unsupervised leaning algorithms. The algorithms provided with oracle miner[1] are: Apriori_association_rules, Decision_tree, generalized_linear_model, K-means, Naïve_bayes, NonNegative_matrix_factor, O_cluster, Support_vector_machines.
- **The Evaluation Phase**: In this phase, reviewed the different models constructed previously and select the ones that will best suited to the project requirements.
- **The Deployment Phase**: This is the phase that makes use of the created models. The required actions have been implemented in the intervention application.
- **Project Evaluation:** We have developed an evaluation strategy that will allow us to gather evidence to support areas such as efficiency of the approach, the extent of support of institutional strategies, tangible benefits in terms of efficiencies and student performance and engagement. The early stage of the evaluation have included gathering feedback from stakeholders in a number of forms such presentation to the university executive committee, university teaching and learning committee and executive faculty committee. We also developed a survey inviting both academic staff and student to give their feedback. These surveys can be found in:

    - Student survey (http://samsa.tvu.ac.uk/mcms/survey_student)
    - Tutor survey (http://samsa.tvu.ac.uk/mcms/survey_tutor)

[1] http://www.oracle.com/technology/products/bi/odm/pdf/data-mining-11g-datasheet.pdf

3 Implementation

In MCMS the implementation was done in two phases; one for the implementation of the data mining process and the second phase in the implementation of the intervention application. The system has been built using Oracle technology (Oracle 2010) and its architecture is illustrated in figure 1. The components of the system are described in the following sub-sections.

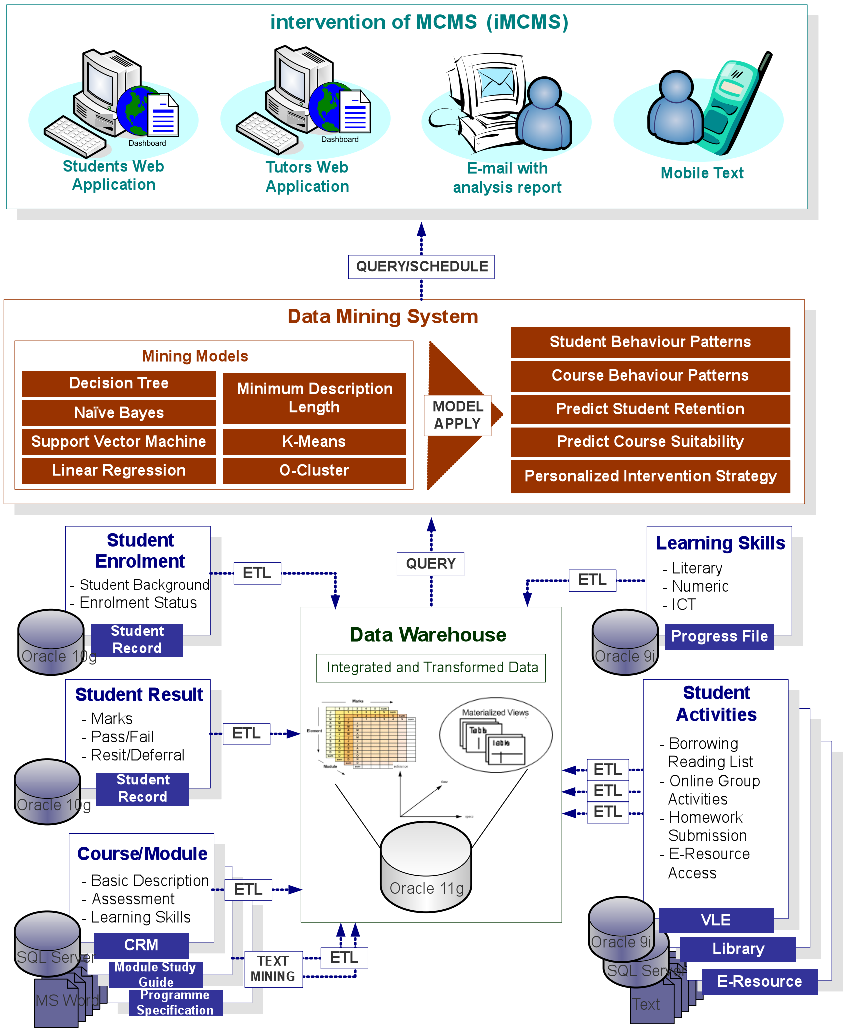

Figure 1 System Architecture

3.1 Data Integration Component

This component acts as the interface between the institution data sources and the rest of the system. A model-driven data integration (MDDI) approach was applied for the data integration (Kim et al 2009). MDDI is a data integration approach that incorporates and proactively utilises meta data across the data integration process. By coupling data and meta data, MDDI drastically reduces the complexity, and provides data integration that is aware of the context of the data. An example data model which has been used for cleansing and transforming data from the student record system data source is shown in figure 8.

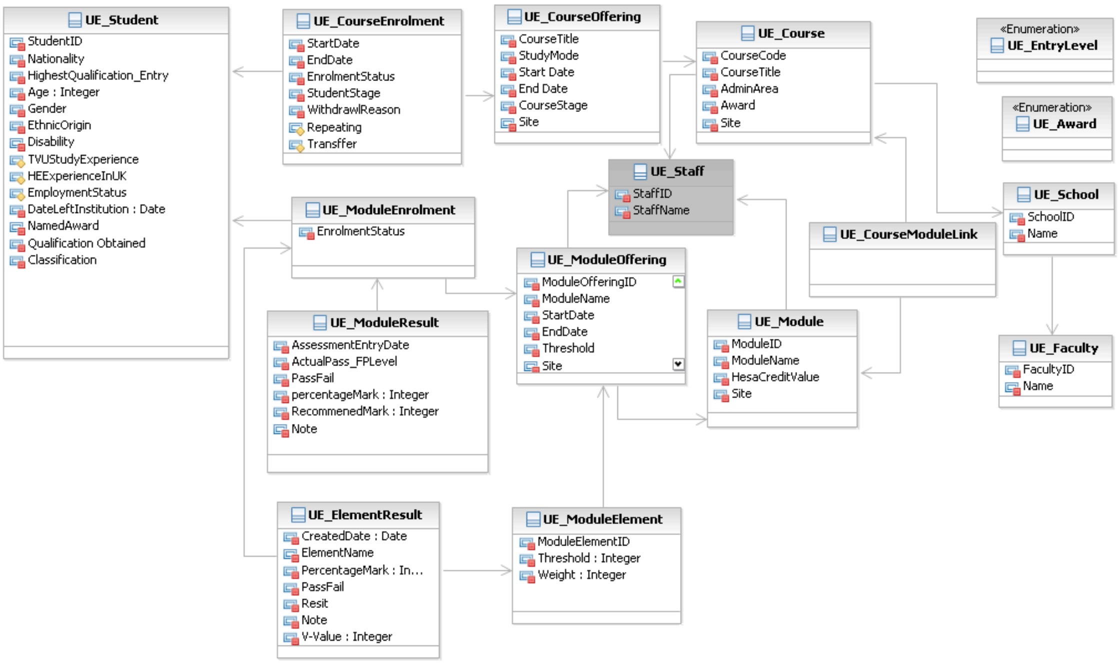

Figure 2 Integrate data model

3.2 Data Warehouse Component

This component takes the data provided by the data integration component and builds the data warehouse; i.e. the different dimensions and cubes. We have defined three cubes in this study; one for students, one for modules and the other one for courses. Using cubes, the data analysis can be effectively conducted by drilling down and sunning up along to the dimensional attributes. The statistical results can be also rendered as 3D graphs by using online analytical processing (OLAP). An example of the module cube is illustrated in figure 9.

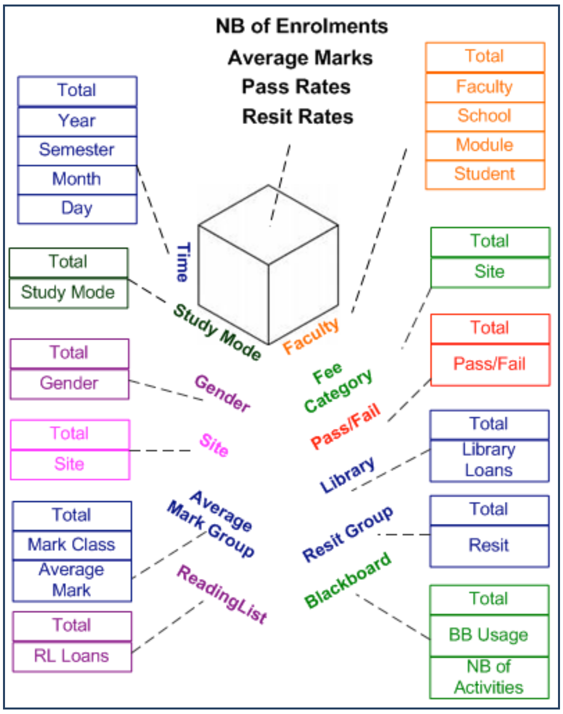

Figure 3 Module Cube

3.3 Data Mining Component

This component is based on three main types of data mining models (Ying et al 2010):

1. A model used to identify the feature's importance that helps reduce some features.
2. A model used for clustering. The output from the model allows the identification of groups on which the intervention will be targeting. Student intervention strategies are differentiated by which group a student belongs to. An example clustering model which identify a specific group is represented as a rule description in table 1.
3. A model used for the prediction of a student's academic performance; i.e. the student passing a module, student dropping out.

```
AVGMARK <= 49.8 and
AVGMARK >= 1.401298464324817E-45 and
AVGRESITNUM <= 1.67 and
AVGRESITNUM >= 0.2 and
COURSEAWARD in (Associate,BA,BSc,CPPD,Foundation,HND,PG,SPEED-,UG) and
CURRENTSTUDYLEVEL = 4.0 and
ELEMENTNUM <= 15.29 and
ELEMENTNUM >= 1.0 and
ENTRYCERTIFICATE <= 5.17 and
ENTRYCERTIFICATE >= 2.67 and
MODULETAKEN <= 7.43 and
MODULETAKEN >= 1.0 and
PG_UG equal (UG) and REPEAT in (0.0,1.0) and
RESITNUM <= 9.25 and RESITNUM >= 1.0 and
STUDYLEVEL = 6.0 and STUDYMODE in (FT, PT) and
UK in (N, Y)
```

Table 1 Example of data mining rule

3.4 Intervention Application Component

This component presents the results of the data mining models and implement the intervention system. We have used different views for different stakeholders. One view is targeted towards improving students' performance and providing them with a holistic and a detailed view of their performance. Students at risk of dropping out receive an intervention message via email or SMS. The other view is to support the academics in implementing their intervention policies. This will provide the data at different levels of abstractions depending on their role (tutor, programme leader, and head of school). Any of the intervention content is generated according to a predefined and personalised intervention rule. The different parts of the components are illustrated in figure 10:

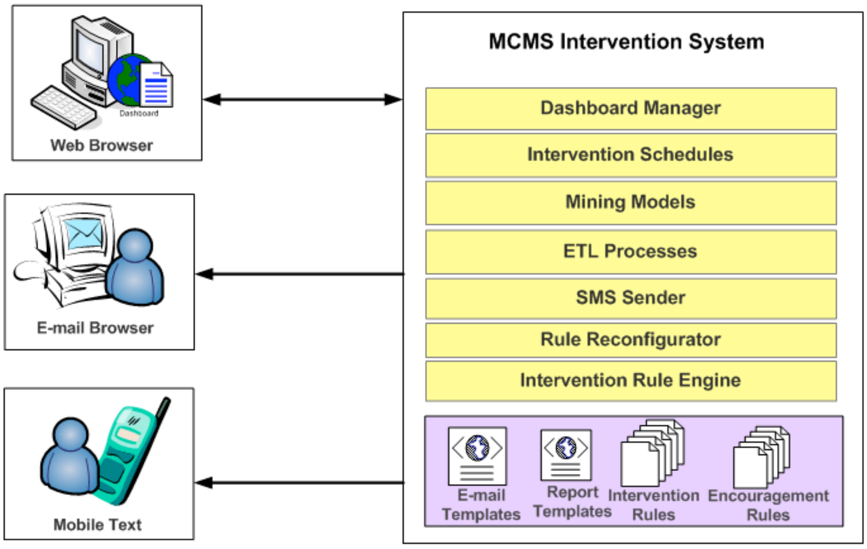

Figure 4 Intervention Component

4 Outputs and Results

The main project outputs are

- a unified data model for integration of HE data sources
- A data mining models for predicting student dropout and education performance changes
- An intervention application that combine data mining process with intervention policies

4.1 A unified data model for integration of HE data sources

Data mining involves applying algorithms that identify patterns to a database that is populated with candidate data. A Higher Education Institution, like most large organizations, does not provide a single database to which the data mining process can be conveniently applied. In such organizations, the data is distributed using many different technologies and formats. These are the *data sources* that must be merged into a single unified database before the mining can begin. Furthermore, the data sources typically overlap in terms of the information that they contain and are often inconsistent with one another.

Therefore, it is important to have a single uniform clean data model that is used as the source of the mining algorithms. This data model is created by merging information from the different sources in addition to cleaning up the data by resolving redundancy, inconsistency and incompleteness. This section describes this process.

4.1.1 The Data Model

Student who are highly integrated academically are more likely to persist and complete their degree. The models we used in our study are of two type: a data mining model that helps predict potential drop-outs, and an intervention model that is used once the potential drop-out has been identified. We have developed a system that puts in place an intervention process to retain students, once a potential drop out has been identified. In our intervention system, we believe that students play a critical role in being successful and remaining at university. Students need to be actively responsible and participate actively in their learning process.

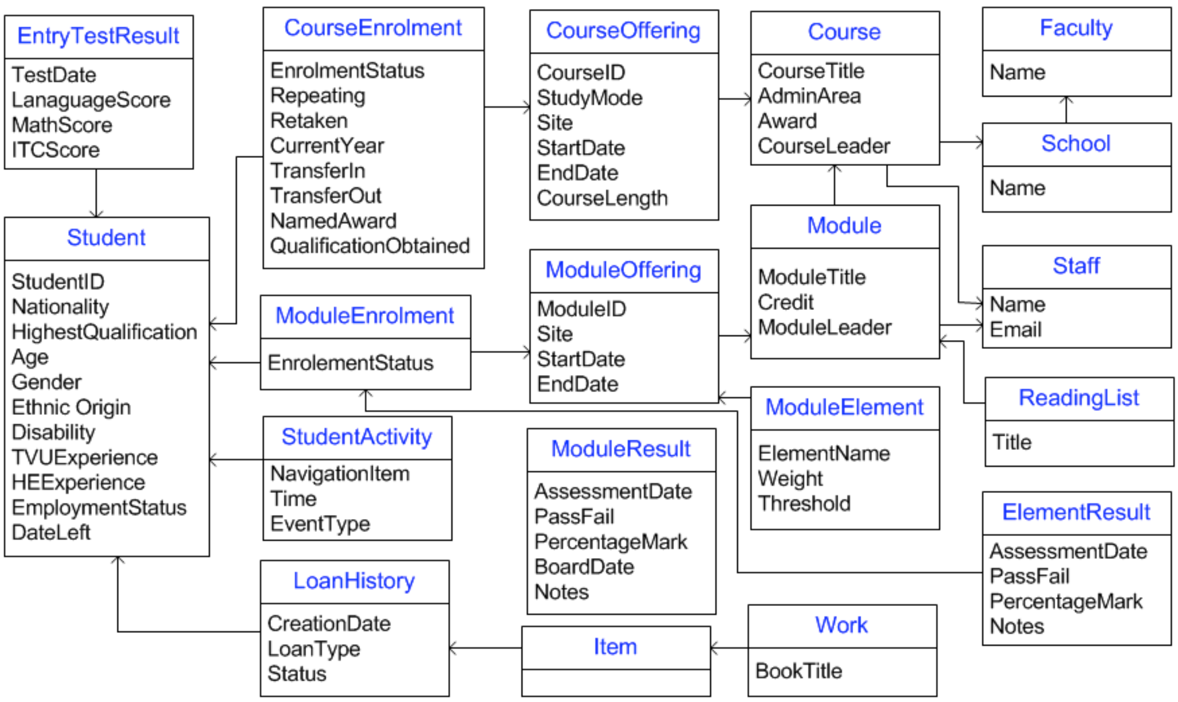

Figure 5 The unified Data Model

The data mining model is based on two main information, one related to the student and one related to the university. The information related to the student allows us to understand better the student.

The information related to university has been concentrated in this project to information related to courses and their related modules. Using information related to students, courses, and modules; we conduct data mining techniques to determine which indicators provide the best assessment of potential student dropouts. The data model used in the data mining process is illustrated in figure 1.
The data model has mainly been constrained by the institutional data sources that were available to the project. The sources that have been used to populate the data-mining model are shown in the figure below. The sources are integrated and uploaded in a data warehouse before the data-mining engine processes it.

4.1.2 Data Sources

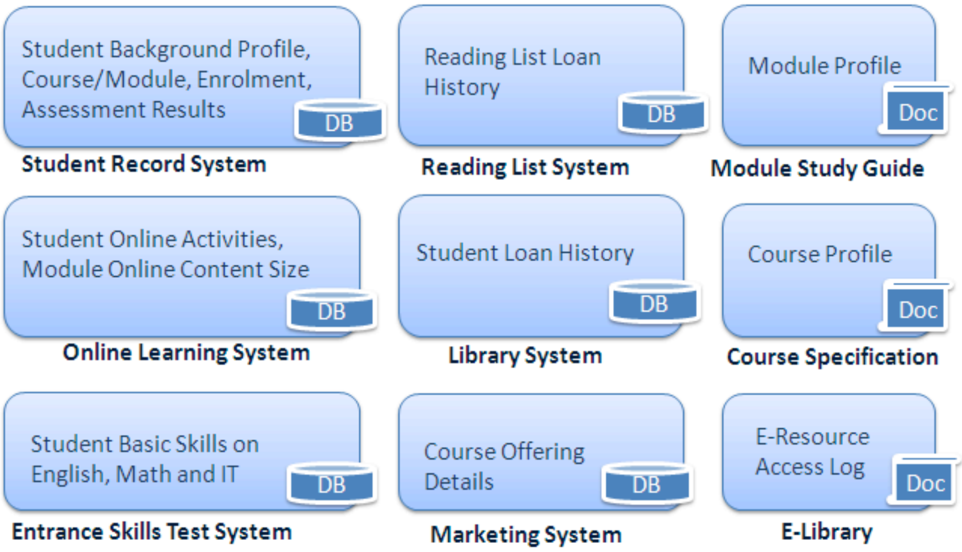

Figure 6 Data Sources

When analyzing the data sources, we have identified information that will give us a good understanding of the students' profile; i.e. data related to information before their entry to the university (student background), data related to their interaction in the university; including their goals(student interaction), and data related to their results. The key sources used in our experiment are as follows:

- The **student record system** relates to the profile of the student prior to joining the institution. Student profile will include information such as their entrance level, their ethnicity, literacy and numeracy test entrance score. In addition the student's assessment profile including marks and the number of resits taken is maintained by the record system.
- The **library system** and the **reading list system** captures information about a student's book loan activities and links this to the reading lists set for the modules that the student is registered for.
- The **online learning system** and **e-library** records how often the student logs into the system and the use of the various pages in the *virtual learning environment* (VLE). In our case the VLE can capture the number of hits on individual pages and document downloads.
- The **module study guides** provide information such as reading lists and the schedule of assessments for each module. The **course specification** provides information such as regulations about options and assessment hurdles. The study guides and course specifications also provide lists of learning outcomes that can be gained by the students when they pass individual modules or module elements. Individual assessments are broken down into different types. Institution departments that own course components are recorded in addition to the tutor responsible for delivering the module.
- The **marketing system** and **entrance test system** provides information about entry requirements and whether students have had any additional tutoring on entry. For example overseas students may be offered extra tuition in English and the test system will record whether they took up this offer and, if so, the results.

4.2 The Data Mining models

Educational data mining is a newly emerging discipline and there have been reports of a number of demonstration applications in Spanish (Romero et al. 2007) and American (Luan 2002) universities, and particularly in distance-learning institutions (Mor & Minguillón, 2004). Data Mining has already proved to be successful in e-commerce and bio-informatics, where results are achieved through the use of associators, classifiers, clusterers, pattern analysers, and statistical tools. In the educational context data mining provides analysis of the students' behaviour, navigation, frequency, and length of interaction with the e-Learning system that can identify patterns of behaviour and associations. It can classify students into groups depending on their learning behaviour rather than just ability. At the same time it also identifies students exhibiting atypical behaviour that needs early intervention and feedback. The overall process that we have used in this system is illustrated bellow. This shows that the intervention system will be used continuously. At each iteration, data is fed to the data mining process, in order to identify potential drop-out students. The intervention process will then identify specific intervention actions for these students.

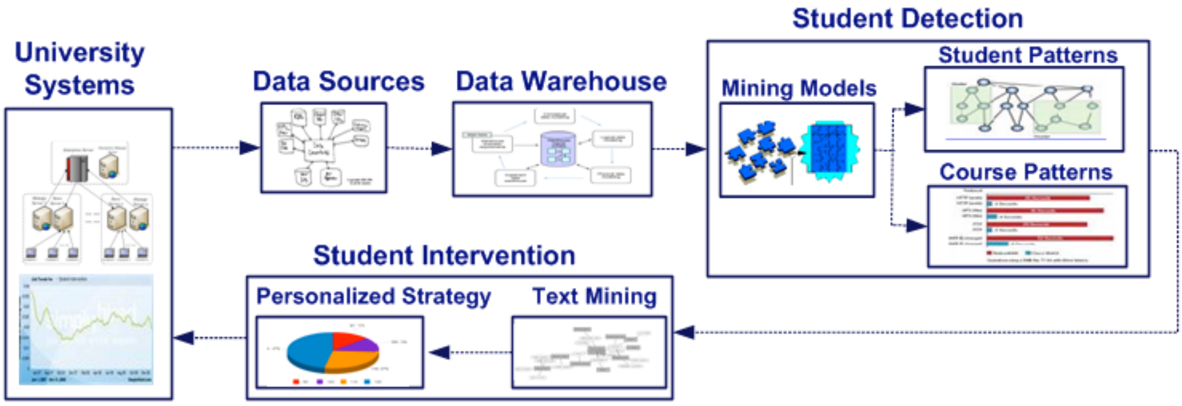

Figure 7 Overall Process

The data mining process used includes three main stages: finding the features' relations, data grouping, and making the prediction.:

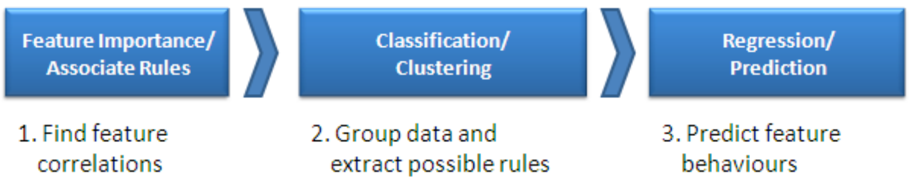

Figure 8 Data mining process

The first stage involves applying features' importance and associate rules to find the correlation among data features. This stage helps eliminate any data that is unlikely to make any impact on subsequent tasks. For example, we found that age and gender is not related to students' performance (results or drop-out), whereas the VLE interaction is related to students' performance. The next stage involves classification of data and extraction rules and patterns. Here we identify groups of students that have shown related features and will require similar intervention. Examples of such groups include first year full time students, students transferring from other institutions, non-UK students, students with low library usage or post-graduate students. The third and final stage includes applying regression to the identified groups in order to predict future behaviours; i.e. allowing us to predict the potential students that are at risk of dropping out.

The above data mining models have been built based on three years of historical data before being integrated to live feed data. The data were divided in two sets, one for building the models and one for validating the model. Once the model have been validate, the data warehouse is updated every week. The prediction stage of the process is conducted on the updated data, leading to new alerts.
Output from the data mining process tends to produce numerical reports and visualisations of clustering, association and classification procedures. In general data mining tools have been used

and interpreted by knowledge management, specialists, and therefore there has been little attention to usability, learn ability or understandability of the output produced by these systems, which is not suitable for feedback to tutors and students. In the rest of the paper, we will discuss the design of the system built in order to allow for better presentation of the data mining results. Data is presented to strategic decision-makers, course leaders, tutors, and students in a form that is easy to act on.

4.3 Intervention system based on a gaming environment

As was discussed in previous section, we designed a data mining process that allows us to identified students that potentially are at risk. The aim of this section is to discuss the intervention system that build on the data mining process.

4.3.1 Gaming Features

Gaming, and particularly on-line gaming has become very popular with young people in recent years. In such systems, players can develop a profile based on playing a number of (possibly collaborative) games. The profile can be tailored to the individual in terms of the look and feel and represents achievements in terms of goals attained, points achieved, extra features unlocked etc.

When a game is played, there are a number of features that can be attained such as completing a level or defeating a foe. In general, each game attaches points to the different challenges it presents and, although the games are different, the points are in a universal currency (or at least universal to a specific gaming platform). Points awarded to a specific gamer represents their level of achievement in terms of skills attained and challenges overcome.

Gamers can compare their aggregate performance against other gamers to produce a league table; relative positions in a league table can be a powerful motivating factor and gamers can spend a great deal of time trying to move their position up the table. In addition games can compare their performance at a more fine grain level in terms of specific skills and achievements. A gamer may have a specific interest in achieving a given skill because it is transferrable to another game.

Our proposal is that students can be viewed as gamers and learning outcomes can be viewed as being similar to points awarded when playing games. That being the case we propose that the same powerful motivating factors that lead gamers to strive to increase their performance (whether relative or absolute) can be applied to students.

4.3.2 A Model for Intervention

We have built an intervention system that put students as the main actors. Students play a critical role in being successful and subsequently remaining at the university. Studies have indicated that motivation is a perquisite for student learning (Svinicki, el. al. 1996], and the student can foster this motivation by setting clear and explicit learning goals and understanding the expectation of success. The greater the belief that a task can be accomplished the greater the motivation. In our system, at any point of time the student will be presented with what has been accomplished so far in terms of learning outcomes gained within each module, the modules that have been passed and the marks gained so far. The student is also aware of what is expected in order to acquire their qualification, for example, in terms of modules to be taken, learning outcomes to be gained within a module, and number of assignments. However, this will only make an impact, if students are engaged; i.e access and make use of the information.

Computer games being a popular and familiar medium to most students, it presents an attractive means as a motivation factor for students in engaging them. A number of attributes common to computer games are recognized in fostering active engagement; motivation and a high level of persistence in game play (Garris, et al 2002). These include the use of environment that simulates realistic experiences for the player, providing opportunities for identity exploration and play through role play (Squire et al 2003). In a number of games, a player may learn to take on attributes of their avatar (Yee & Baileson 2007). Using avatars may lead to a sense of responsibility towards the character that can lead to educationally relevant outcomes. The other main attribute is the creation of a sense of pride and accomplishment through structuring the game that will challenge the player and allow progress.

The above principals have been implemented in the intervention system in order to harness the motivation potential of gaming; including associating academic performance with scores, use of avatar, structure the levels based on learning outcomes and modules and providing a league board.

The data model on which the system is based is shown in figure 5. Based on their profile, students are associated to one of the groups that have been discovered in phase two of the data mining process (in the diagram below, they are referred to as the Zone). For each of the zone, phase three of the data mining process would have identified a threshold for when intervention actions are required.

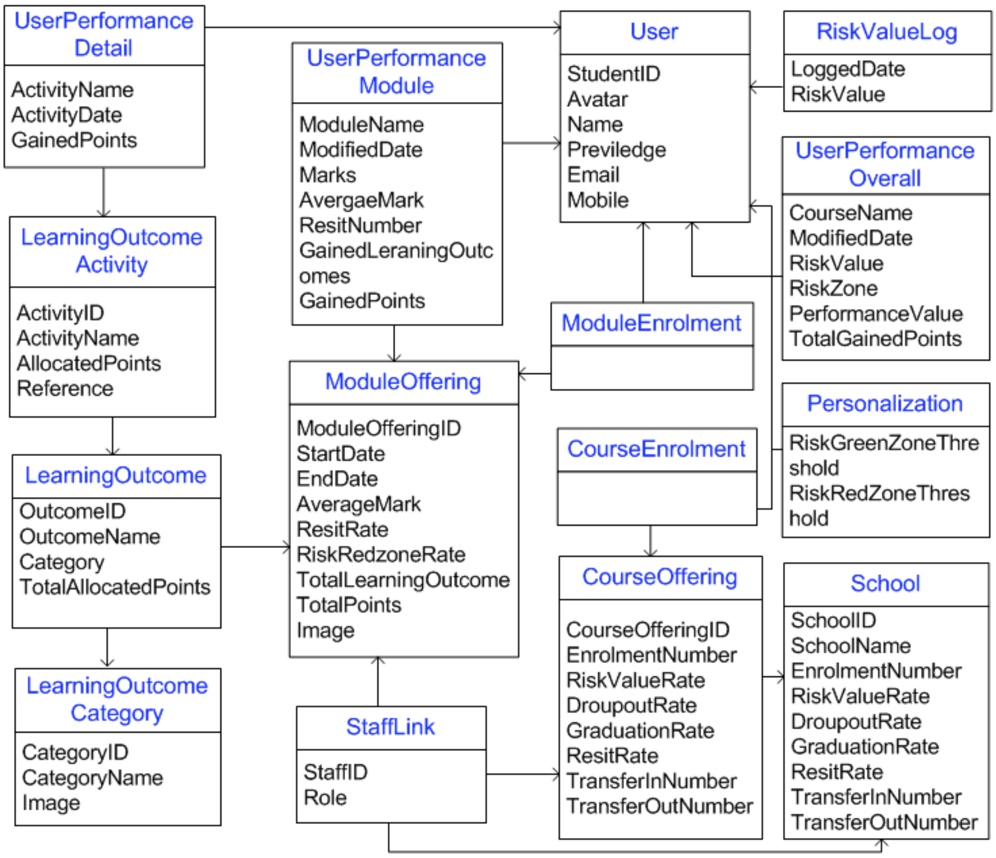

Figure 9 Intervention Data Model

4.3.3 The Intervention Interface

Students are encouraged to use avatars that have characters that they would like to relate to. Being associated with a strong avatar might motivate them to acquire high scores. Student activities are mapped to module learning outcomes activities in order to convert student performance as counted points. For example, if a student borrowing a book from the reading list, he or she would get two points for his or her module and if he or she submits an assignment, he or she gets five points. If the students reach the target points for each learning outcome of a module, they would see their mission achievement records in the web application and would encounter the next challenge for further improvement such as a usual role-playing game. We have also implemented a league table where scores can be compared within the same group. A screenshot of the implemented student web application is shown in figure 6:

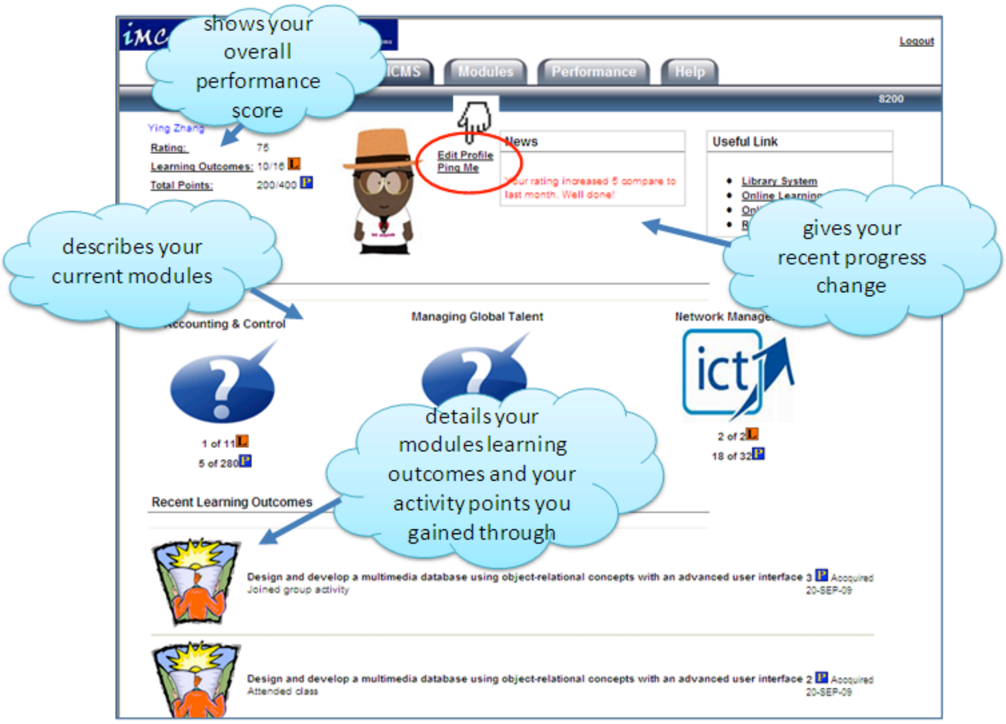

Figure 10 Student Intervention Application

Students are alerted by email or text if they are in at-risk zone. In the message, they are presented with their achievements so far in terms of the points. The message also advises them with a number of actions that they need to undertake in order to get out of the zone. A student that has taken the right action and has managed to get out of the zone is also alerted. The message that they receive is a congratulatory message that invites them to review their achievements and show them how they can unlock the next level.

The design of the messages and the actions that the student are invited to undertake are associated with each of the groups that were identified in the second stage of the data mining process. For each group; based on the pattern that describes the group, a template for the message is designed, with specific actions that will lead to correcting the low scores in the particular interaction, such as accessing the module VLE content. The templates are then personalised with live data specific to each student. The intervention messages are common to all modules, however each module leader is encouraged to refine the rules that will lead to alerts and customise the intervention messages and actions.

5 Outcomes and Implications

MCMS project at Thames Valley University aimed to apply Data Mining technology to multiple institution data sources in order to identify predictive rules that can be used to detect and improve issues related to student retention. The project has generated the following outcomes:

- A key problem in performing business intelligence within a large organization is to be able to integrate many different data sources that contain overlapping, incomplete, inconsistent and noisy data. MCMS has addressed these data integration issues, including technical, organizational and legal issues.
- The project has also engaged with institutional stakeholders in order to identify requirements for business intelligence throughout the business. Requirements have included: legal issues for data protection, institutional processes for access to data, and presentation of data to users.
- MCMS has used standard Data Mining techniques to identify rules for student retention. The data mining process has defined predictive rules that can be applied to institutional data in order to identify problems and to propose intervention strategies.
- MCMS has used a gaming metaphor for the design of the intervention system. Data mining models generate predictive rules that are then executed against live student data to generate personalized intervention events. Students engage with the events using a scoreboard system such as that used in on-line gaming, they can compare their scores against their peers and have clear indicators as to how their scores can be increased. Tutors also receive events as a consequence of changes to student behaviour; this allows them to make effective use of resources when supporting students.

5.1 Implications

From the development and implementation of the MCMS system, a number of implications arise:

5.1.1 Programme phase – project lifespan:

Our experience suggests that the project lifespan does not accommodate or align with the course structure and schedule where the deployment is planned. We were not able to make the system live at the start of the academic year. Partly this was due delays in development phase (mainly due to the delay in data source availability), and the misunderstanding of the implication of MCMS on the student enrolment process regarding the personal data act.

5.1.2 Designing and implementing an uptake strategy

Our experience suggests that any project requiring piloting an application should have a dedicated work package for managing the uptake of the uptake of the application by the users.

5.1.3 Decision support system supported by Business Intelligence

The main application of Business Intelligence has been to support an institution reporting activities. In this project, we have extended the use of Business Intelligence to be embedded in the intervention application. Hence directly supporting the decision making process.

5.1.4 Student engagement and motivation can be part of a retention strategy

Our experience suggests that data mining can be used to give an indication of student engagement. Intervention policies implemented by the

5.1.5 An integrated data model for supporting student retention

In MCMS project, we have developed an integrated data model for supporting student retention. It will interesting to see if the model is applicable to other institutions.

6 Conclusions and Recommendations

In this project, we have explored the benefitting factors of using data mining combined with a gaming environment in order to engage and motivate students in terms of retention. The prototype system built in the project has been used to integrate the data mining process with the intervention system. The data mining process is used as two main functions; primarily to predict those students who are most likely to drop-out early, and secondly to group students into specific categories that can be targeted with personalised interventions if it is predicted that a drop-out is imminent. Certain activities and accomplishments merit a specific number of points to be added on the appropriate student's game profile. The student unlocks each level in the gaming process through achieving a set score for each learning outcome. This idea of levels is a great source of motivation to students, as well as the idea of a 'leader board' in which students' are encouraged to become competitive. This also enforces the gaming environment, making 'winning' appear more appealing than in a classic university environment.

In terms of further improving the system there are a few problems that must be targeted. At each iteration the data mining models must be manually retrained, this is a hindrance, and will be rectified by implementing an algorithm that creates self-adapting models. Another area to expand upon is the range of data sources available. We hope to add new data sources, such as: financial data, timetabling, and data relating to their social involvement in universities; however this has not initially been possible due to data constraints. The next area for improvement will involve creating an additional gaming feature (i.e. settings), providing a user-friendly interface enabling students to access key information, such as: Course information, university regulations, and their personal timetable. This idea of integrating learning and games is further developed through literacy and numeracy features to support their main subject. This is once again incorporated into the gaming environment, similar to many educational games and consoles available in the current market.

7 References

Anagnostopoulou, K. & Parmar, D. (2008). ,Practical guide: bringing together e-learning & student retention, London: Cats Ltd. ISBN: 978-1-85924-301-5.

Bean, J.P. (1980). Dropouts and turnover: The synthesis of a causal model of student attrition. Research in Higher Education, 12, 155-187.

DeCastell, S., Jenson J., (2003) Serious play, Journal of curriculm Studies 35(6), 649-665 Cerrito P., (2006) Data Mining Student Performance in Mathematics Courses, CinSUG Third AnnualOne Day Conference.

Committee of Public Accounts, (2002) Fifty-eighth Report of Session Improving Student Achievement and Widening Participation in Higher Education in England, HC 588,2001-02.

CRIP-DM (2010) http://www.crisp-dm.org/ Fabro, D.D.M., Valduriez,P., (2008) Towards the efficient development of model transformations using model weaving and matching transformations, Software and Systems Modeling, 2008.

Gabrilson, S. (2003) Data Mining with CRCT Scores. Office of information technology, Geargia Department of education.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. Simulation and Gaming: An Interdisciplinary Journal, 33(4), 648 441–467.

Harvey L. Drew S, Smith M (2006), The First-year Experience: A Review of Literature for the Higher Education Academy, http://www.heacademy.ac.uk/research/Harvey_Drew_Smith.pdf.

Hesa (2008) http://www.hesa.ac.uk/index.php/content/view/1169/265/.

Kim, H., Zhang, Y., Oussena, S., and Clark, T., (2009) A Case Study on Model Driven Data Integration for Data Centric Software Development, In Proceedings of ACM First International Workshop on Data-intensive Software Management and Mining.

Kloesgen W., J Zytkow, (2002) Handbook of Knowledge Discovery and Data Mining, Oxford University Press, Oxford, 2002.

Luan, J. (2002) Data mining and knowledge management in higher education –potential applications. In Proceedings of AIR Forum, Toronto, Canada.

Mazon, J-N., Trujillo, J., Serrano, M., Piattini, (2005) Applying MDA to the development of data warehouses. DOLAP.

Mazza, R., Milani, C. 2005 Exploring usage analysis in learning systems : Gaining insight from visualisation. Workshop on Usage analysis in learning systems at 12th International Conference on Artificial Intelligence in Education, New York, USA (pp. 1-6)

Minaei-Bidgoli,B., Kortemeyer, G., Punch,W.F.,(2004) Enhancing Online Learning Performance: An Application of Data Mining Methods, From Proceeding of Computers and Advanced Technology in Education .

Mor E. , Minguillón J. ,(2004) E-learning personalization based on itineraries and long-term navigational behavior, Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, May 19-21, New York, NY, USA .

Oizilbash.H, (2008) TVU System Overview, Thames Valley University.

Oracle (2008) http://www.oracle.com/technology/products/bi/odm/index.html.

Oussena,S., (2008) Mining Courses Management Systems, Thames Valley University, 2008.

Romero, C., & Ventura, S. (2006). Data mining in e-learning, Southampton, UK: Wit Press.

Seidman, (1996) A. Spring Retention Revisited: RET = E Id + (E + I + C)Iv. College and University, 71(4), 18-20.

Squire, K., Jenkins, H., Holland, W., Miller, H., O'Driscoll, A., Tan, K. P., (2003). Design principles of next-generation digital gaming for education. Educational Technology, 700 33, 17–23.

Svinicki M., Hagen A., Meyer D. (1996) How to Reaserch on Learning Strengthens instruction, Teaching on Solid Ground: Using Scholarship to Improve Practice, Jossey-Bass publishers, San Francisco, CA.

Tinto V., (1975) Dropout from Higher Education: A Theoretical Synthesis of Recent Research, Review of Educational Research vol.45, pp.89-125,1975.

Thomas, L. (2002) Student retention in higher education: the role of institutional habitus", Journal of Education Policy, Vol. 17 No. 4, August, pp. 423-442, 2002.

Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. Human Communication Research 33, 271-290.

Yorke, M., Longden, B., (2004) Retention and student success in higher education. Society for Research in Higher Education. 2004.

Yu, C.H., DiGangi,S., Jannasch-Pennell,A., Lo,W.J., Kaprolet, (2007) A Data-Mining Approach to Differentiate Predictors of Retention, Educause Southwest Conference, Austin, TX.

Witten, I. H., Frank,E. (2005) Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco.

Zhang, Y., Oussena, S., Clark, T., Kim, H., (2010) Use Data Mining To Improve Student Retention in Higher Education – A Case Study, to be presented in 12th International Conference on Enterprise Information Systems(ICEIS).