International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Contextual Data Mining for Higher Educational Institutions

Subhashini Sailesh Bhaskaran, Mansoor Al Aali, Kevin Lu

Abstract*: Context-awareness research has been carried out by many researchers. However, little has been done in building a context-aware data mining methodology which supports decision-making in HEIs. The usefulness of Knowledge discovery data mining process (KDDM process) in HEIs were investigated to discover hidden knowledge that is contextualized, resident in student datasets and use them in decision making. It was experimented and found that not any of the KDDM processes include a contextual factor mining stage that is essential to take out hidden knowledge from datasets described by contextual factors. Therefore a new process was introduced in KDDM process that uncovered contextual data to be used to support business goal and produced a dataset at the preparation stage which generated data mining model that was contextual leading to the unearthing of course taking patterns that are contextualized. This discovery has enabled forecasting of optimum CGPA and time-to-degree.

Index Terms: HEIs, Data Mining, KDDM, Time to Degree, Student Performance, Context-Awareness.

I. INTRODUCTION

In the delivery of education in HEIs decision making processes assume significance. Those decision making processes are concerned with a number of activities in HEIs including teaching and student learning experience. For instance decisions could impact such aspects as curriculum design, categorization of students, timetabling improvements and student assessment that can be related to teaching and student learning experience (BIS, 2014). However accuracy and applicability of those decisions in HEIs are areas of major concern, because common decisions that are taken in HEIs at the institutional level do not always get implemented at the academic level (BIS, 2014). One reason for this could be that decisions usually made in HEIs use data that contain observable patterns or phenomena not entirely adequate to make accurate decisions. Observable phenomena include data and information about student gender etc. However literature points out that there is a need to consider unobservable phenomena also that could have serious implications to the decision making process. Unobservable phenomena could include data and knowledge associated to student education (for instance, student potential ,course difficulty). A major phenomenon that is found to be unobservable is the contextual factor that could contain knowledge helpful in HEIs decision making process. For instance when decisions are being made in HEIs regarding optimizing the student time to degree, knowledge related to course difficulty or student potential could help HEIs in determining the categorization of students and courses as well as providing additional support to students who need support. While literature review shows that contextual factors have the potential to enhance the quality of decision making, extracting such contextual factors from known resources and discover useful knowledge about those contextual factors has been a challenge(Vert et al., 2010). One resource that is very promising that could be used to extract contextual factors and knowledge about those contextual factors is student dataset. Recently researchers have started to focus on contextual factors that are important to HEIs that could be used in their decision making process although more needs to be done. One such area that is clearly providing evidence of this kind of interest and provides a large scope for investigation is Knowledge discovery. However serious questions are raised on the usefulness of existing KDDM processes to enable decision making involving contextual factors related to a particular business environment like HEI. For instance Vert et.al (2010) argues that current KDDM processes need to be enhanced for their usefulness in supporting accurate decision making using contextual factors, an argument that implies that current KDDM processes are not capable of addressing unobservable knowledge that is contextualized. Some argue that further research is needed in this area to support users of KDDM processes (Singh, 2003; Vajirkar, 2003). Thus on the one hand there is a need to use KDDM processes to discover knowledge and on the other those KDDM processes must be enhanced to deal with contextual factors leading to the discovery of knowledge that is contextualized and useful to make more accurate decisions. In the context of HEIs the current level of understanding of the decision makers is limited to visible and tangible factors available in the student data set (For e.g. Gender, age, nationality, grades). Most HEIs are yet to deploy any type of KDDM process to discover hidden knowledge in their decision making process. For instance, a question could be raised regarding predictability of a pattern of courses a student could register in a semester or semesters that would lead the student to achieve a lower time to degree (the time taken by a student to graduate). Such a question is unlikely to be answered straightaway using the data readily available in the dataset like student factors namely semester based number of courses enrolled or Grade Point Average (GPA) scores as the knowledge related to pattern of courses that can be used to predict the time to degree is not readily available in the dataset. Here KDDM processes could be very useful although HEIs do not appear to have deployed such processes yet in their decision making process.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Dr. Subhashini Sailesh, MIS Department, Ahlia University/ College of Business/ Ahlia University, Bahrain.
Prof. Mansoor Al Aali, President , Ahlia University, Bahrain.
Dr. Kevin Lu, Brunel University London.

The reason for this situation are that manual methods or commonly used computing methods(e.g. SQL queries) are not adequate enough to discover patterns from hidden patterns from large datasets and HEIs lack knowledge to extract those patterns from student dataset using KDDM process. Moreover the use of extracted patterns from student dataset using KDDMs in decision making can suffer in terms of accuracy or achievement of business goals as the extracted patterns lack any knowledge related to contextual factors. The main reason for this is that KDDM processes that are currently used in the business domains or discussed in the literature, that have the potential to be used in HEIs, do not have steps to integrate contextual knowledge concealed in the dataset, needed for making better decisions to attain the business goals. This is a major gap in the literature. Thus the objective of this research is to address this gap to some extent by demonstrating a modified KDDM process that extracts hidden knowledge from the student dataset characterized by contextual factors that could be used in making better decisions in HEIs to achieve specific business goals. Towards this, in this research two contextual factors namely course difficulty and semester have been taken as examples to produce course taking patterns from student dataset to forecast optimum time to degree and Cumulative Grade Point Average (CGPA). This knowledge about course taking patterns could be useful in taking many decisions in HEIs that have bearing on teaching and learning, for instance course categorization, student categorization, forecast of optimal time to degree of students and achievement of optimum CGPA.

II. RELATED LITERATURE

A. Review Stage

This section covers two aspects. One related to current knowledge on contextual factors and their utility in KDDM processes and the other related to KDDM processes. Literature review shows that contextual factors have been of interest researchers in various fields. Table1 provides a glimpse of current efforts of researchers taken in different fields that utilize contextual factors in KDDM processes to extract hidden knowledge from datasets that are contextualized.

Table 1, Current Efforts

Meaning of Context	Author and Paper	Remarks
Context, the cumulative history that is derived from data observations about entities (people, places, and things), is a critical component of analytic decision process. Without context, business conclusions might be flawed.	Context-Based Analytics in a Big Data World: Better Decisions, An IBM® Redbooks® Point-of-View publication Sokol, 2013	By using context analytics with big data, organizations can derive trends, patterns, and relationships from unstructured data and related structured data. These insights can help an organization to make fact based decisions to anticipate and shape business outcomes. Entities are defined as people, places, things, locations, organizations, and events. Entities are an important focus of big data analytics. Context is defined as a better understanding of how entities relate. Cumulative context is the memory of how entities relate over time.
A general approach for context-aware adaptive mining of data streams that aims to dynamically and autonomously adjust data stream mining parameters according to changes in context and situations	Context-Aware Adaptive Data Stream Mining, Haghighia et. al, 2009	The researchers proposed an overall method for context-aware adaptive data mining that includes context-awareness into universal data stream mining and allows real time examination of data on board mobile devices in a clever and cost-effective manner. They achieved Context-awareness through Fuzzy Situation Inference (FSI) that assimilates fuzzy logic in the CS model, an official context modeling and cognitive approach for assisting pervasive computing environments.
Context– background information Contextual item set mining extracts frequent associations among items considering background	Contextual Item set Mining in DBpedia, Rabatel et al, 2014	The authors exhibit the capacity of contextual item set mining. Contextual item set mining excerpts frequent associations between items bearing in mind the background

Context as a concept has been identified as important in the field of KDDM recently, due to the potential it has in adding value to the discovered knowledge. Lack of context in the discovered knowledge has been argued to be a limitation in various KDDM processes by researchers (Schilit et al., 1994; Dey, 2001; Bolchini et al., 2007) due to the important role context could play in decision making. (Vert et al., 2010)claim that a dataset characterized by contextual attributes if fed to the KDDM processes at a definite stage, then it is possible that the knowledge discovered through the mining processes will not only contain the normal attributes but also the contextual attributes. These arguments could be extended to the HEIs also.Thus based on these arguments and as explained earlier there is a necessity to investigate the usefulness of KDDM processes in HEIs to discover hidden knowledge that is contextualised, resident in student datasets and use them in decision making. Notably if a method to extract contextual factors is identified and employed in KDDM processes then it is probable to dig out hidden knowledge that is contextualised for better decision making.

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication (BEIESP)

Furthermore, amongst the various factors that have been discussed by researchers in supporting decision making process in HEIs, prediction of optimum time to degree and CGPA by means of students' course taking patterns in association with contextual factors namely course difficulty and semester do not find place. While there are arguments supporting a linkage between course taking patterns, CGPA and time to degree, there is hardly any research outcome that has substantiated this claim using data mining techniques particularly considering semester and course difficulty as contextual factors. This research aims to address this aspect.

A. KDDM Process

A number of KDDM processes have been used in many domains namely medicine, manufacturing, education and other businesses. Widely used KDDM processes are tabulated in table 2 which indicates the steps involved in them. While literature shows the usefulness of KDDM processes in extracting hidden knowledge from datasets for decision making, those processes are also found to have limitations. Table 3 provides a good idea about the limitations affecting different KDDM processes.

Table 2, Evaluation of KDDM (adapted from Kurgan and Musilek, 2006)

Model	Fayyad *et al.*	Cabena *et al.*	Anand & Buchner	CRISP-DM	Cios *et al.*	Generic model
Area	Academic	Industrial	Academic	Industrial	Academic	N/A
No of steps	9	5	8	6	6	6
Refs	(Fayyad *et al.*, 1996d)	(Cabena *et al.*, 1998)	(Anand & Buchner, 1998)	(Shearer, 2000)	(Cios *et al.*, 2000)	N/A
Steps	1 Developing and Understanding of the Application Domain	1 Business Objectives Determination	1 Human Resource Identification 2 Problem Specification	1 Business Understanding	1 Understanding the Problem Domain	1 Application Domain Understanding
	2 Creating a Target Data Set	2 Data Preparation	3 Data Prospecting 4 Domain Knowledge Elicitation	2 Data Understanding	2 Understanding the Data	2 Data Understanding
	3 Data Cleaning and Preprocessing		5 Methodology Identification	3 Data Preparation	3 Preparation of the Data	3 Data Preparation and Identification of DM Technology
	4 Data Reduction and Projection		6 Data Preprocessing			
	5 Choosing the DM Task					
	6 Choosing the DM Algorithm					
	7 DM	3 DM	7 Pattern Discovery	4 Modeling	4 DM	4 DM
	8 Interpreting Mined Patterns	4 Domain Knowledge Elicitation	8 Knowledge Post-processing	5 Evaluation	5 Evaluation of the Discovered Knowledge	5 Evaluation
	9 Consolidating Discovered Knowledge	5 Assimilation of Knowledge		6 Deployment	6 Using the Discovered Knowledge	6 Knowledge Consolidation and Deployment

Table 3, Limitations of KDDM

Process Model	Limitation	Limitation Identified By	Modified step or stage	New Concept
KDD(Fayyad et al. (1996)) 5 step)	Lack of data collection step which is vital for the KDD techniques in some real applications such as information security and medical treatment	(Ruan, 2007)	Data Collection step using previous mining results. It was added before data selectionas shown in Fig 2	Inclusion of Data Collection step in data mining process to filter irrelevant data leading to better decision making. (Ruan, 2007).
	Lack of domain knowledge leading to decision making that maybe useful if such a knowledge is not part of the mined data.	(Redpath & Srinivasan, 2004)	Proposed an architecture based on domain knowledge as shown in Fig 3	Introduction of Domain knowledgebefore the data selection (Redpath & Srinivasan, 2004)
	Unlike in other process models loop back to the second, third or fourth step are necessary due to prepared data which is not suitable for the mining process.	(Kurgan & Musilek, 2006)	Not addressed in the literature	Not addressed in the literature
	Lack of contextual information in the data	(Vert et al., 2010)	Not addressed in the literature	Not addressed in the literature
CRISP - DM	Insufficiency to handle multidimensional temporal data resulting in knowledge that cannot support decision making which is dependent on temporal issues	(Catley et al., 2009)	phases 1 (business understanding), 2 (data understanding), 4 (data modelling), and 6 (deployment) were enhanced to suit temporal data as shown in Fig 4	Introduction of Intelligent Data Analysis architecture in the mining process for ensuring the mined knowledge to support decision making that need temporal aspects(Catley et al., 2009).
	The current CRISP -DM model is limited to address data that is fr ee of human intervention	(Li et al., 2009)	Proposed model had two backbones of the model, namely data mining and applied intelligent system, three participation elements namely On-Line Analytical Processing (OLAP), six sigma, and domain knowledge as shown in Fig 5	On-Line Analytical Processing (OLAP), six sigma, and domain knowledge (Li et al., 2009).
	Lack of integrated process model.	(Sharma et al., 2012)	Identification of task-task dependencies (between tasks of the same phase and different phases) is the first step	Building an integrated process model(Sharma et al., 2012)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The previous information reveals that not any of the KDDM processes include a contextual factor mining stage which is essential to unearth unseen knowledge from data identified by contextual factors. Even though there are few papers that have talked about context knowledge and context driven data mining, such knowledge do not propose how to link data mining to contextual factor to help in decision making in businesses including HEIs that lead to accomplishment of business goals. When such addition could be attained, the resultant KDDM model can have a higher foretelling power which is essential for making better decisions. This has been addressed in this research in a limited manner by adding the CRISP-DM process with a contextual factor mining stage. In this context, this research relies upon KDDM process developed by Chapman. CRISP-DM model is given in Figure 1.

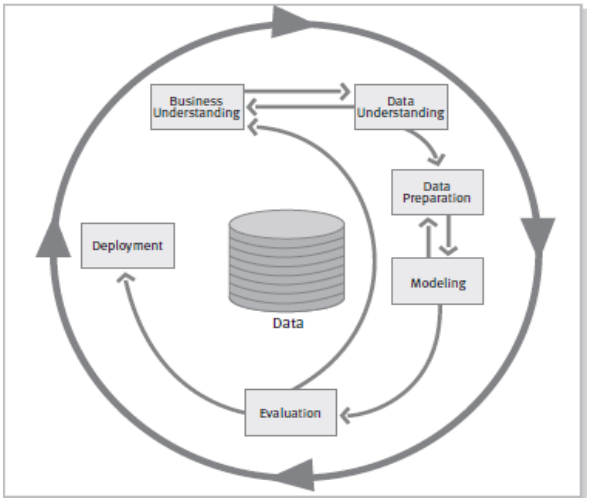


Figure 1, Current CRISP-DM model described in the literature (Adapted from Chapman et al., 2000)

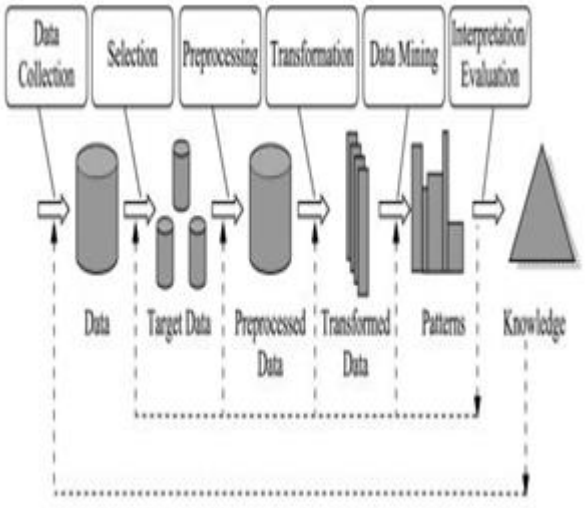


Figure 2, KDDM process model adapted from (Ruan, 2007)

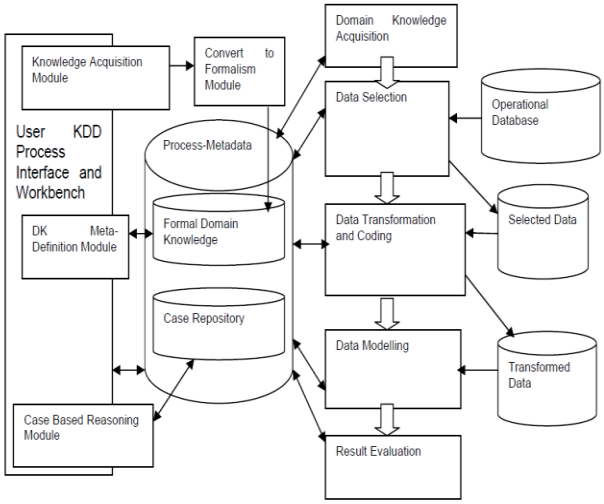


Figure 3, KDDM process model adapted from (Redpath and Srinivasan, 2004)

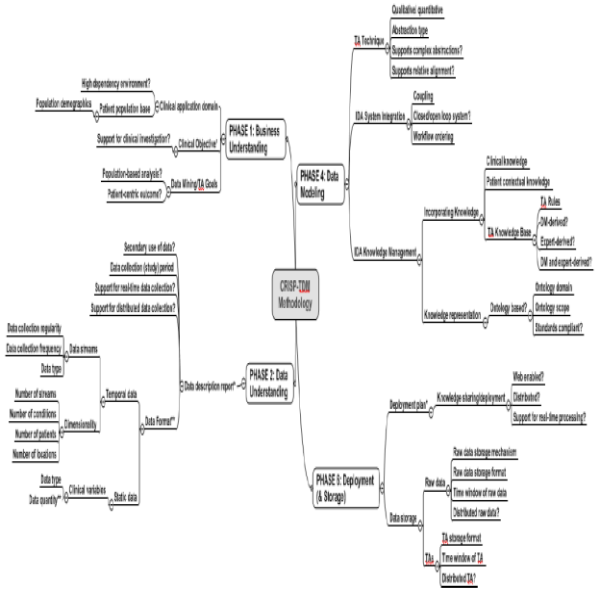


Figure 4, KDDM process (Catley et al., 2009)

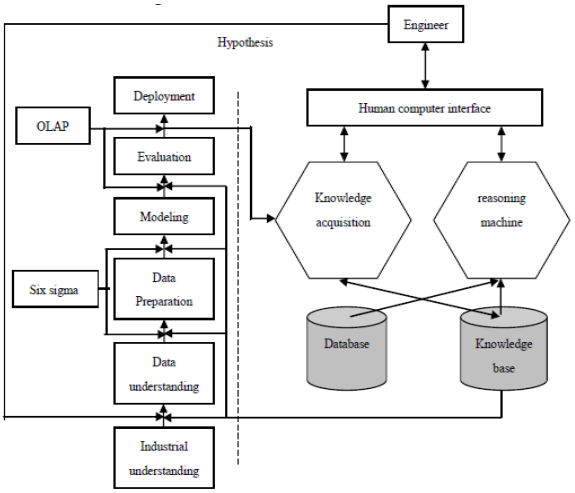


Figure 5, KDDM process model adapted from (Li et al., 2009)

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication (BEIESP)

I. THREE STAGE PROCESS TO MODIFY KDDM

This section covers three stages. The first one covers a discussion on CRISP-DM process including its limitation to handle contextual data, modification needed to enable into handle contextual data and how a dataset with contextual factor can be used to discover knowledge through the process. The dataset thus identified is used for the next stage of discovering hidden knowledge. Second stage informs how the CRISP-DM method is actually modified to handle contextual dataset. The third stage provides information on the experiments conducted on a contextualised dataset by means of the tailored CRISP-DM process which include a method to detect the presence or absence of a particular contextual factor within a dataset using an algorithm developed for the purpose. In addition the third stage also depict how the modelling stage is linked to the contextual dataset that enable the CRISP-DM process to extract hidden knowledge from that student dataset described by contextual factor that could be useful to make decisions. The results of the experiment have been provided. In order to discuss the above mentioned aspects it is important to note that the researcher used the dataset of a HEI in Bahrain pertaining to BS programme in Accounting and finance (BSAF). The student dataset was extracted from the student registration system and has limited data pertaining to time to degree and CGPA. The method used in this research to determine the presence of context factor (course difficulty and semester) was the one suggested by Vert et al. (2010) which uses pseudo code. Each one of the stages mentioned above is explained below.

II. TESTING AND EVALUATION

Use Refer figure 1. A broad summary of the different stages of CRISP-DM process has been provided in table 5 because the detailed discussion on the CRISP-DM process is beyond the scope of this paper. Further due to paucity of space each stage and its components have not been discussed in detail. Instead important components that are necessary to support the experiments conducted in this research are given below (for a detailed description of various components of CRISP-DM refer to Chapman et.al (2000). In addition this section does not include discussion on any step related to contextual factor as the aim of this section is to fetch to see whether it can produce hidden pattern from the mined data including contextualised pattern.

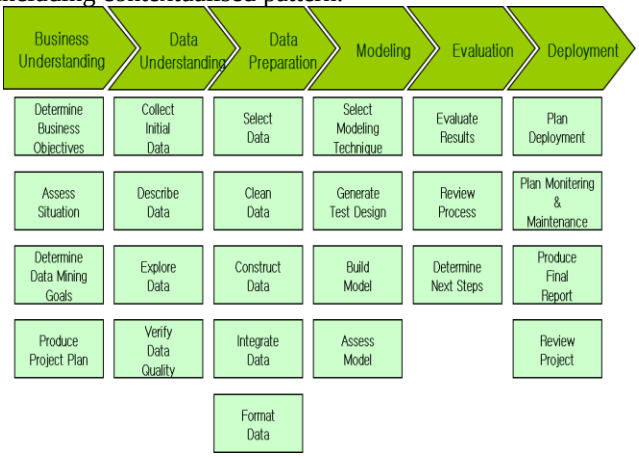


Figure 6, Current CRISP-DM model steps described in the literature (Adapted from Chapman et al., 2000)

The essential steps of CRISP-DM are given below.

The detailed specification of the stages and components can be had from the authors.(i) Business Understanding: To forecast the optimal time to degree using course taking patterns and CGPA and to categorize courses and students with unearthed knowledge. For this purpose data from student registration system of a university in Bahrain was used. Weka was used as the data mining tool.

(ii) Data Understanding: SQL queries was used to retrieve the data (see table 3). The outcome was a dataset that can be mined to attend the business problem. An investigation of the dataset discovered probable association between semester wise student course registration data and CGPA. The dataset was evaluated for quality problems like extreme values, incomplete values.

Table 3, Data set without contextual factors extracted

Student_ID	gpa	len	sem_gradepoints	sem_pass_credit	Course_C
2E+08	3.78	3.5	3.668	15	ACCT 101
2E+08	3.78	3.5	3.668	15	ARAB 101
2E+08	3.84	3.5	4	15	ARAB 101
2E+08	3.84	3.5	3.6	15	ACCT 101
2E+08	3.84	3.5	3.6	15	ARAB 101
2E+08	3.84	3.5	3.776667	18	ACCT 101
2E+08	2.18	4	1.553333	9	ARAB 101
2E+08	2.18	4	1.75	12	ACCT 101
2E+08	2.62	4	2.934	15	ARAB 101
2E+08	2.62	4	2.61	18	ACCT 101
2E+08	2.62	4	2.666667	9	ACCT 101
2E+08	2.62	4	2.666667	9	ARAB 101

(iii) Data Preparation: This stage concerned the choice of records of students from the dataset, clean-up, building, combining and arranges data. The resulting dataset enclosed partial information associated to enrolled student courses, semester information, CGPA and time to degree. The dataset contained 50 entries.

(iv) Modelling: This stage involved the use of modelling algorithm that enabled selection of modelling technique, generation of test design, building a model and assessment of the model.

The modeling technique chosen was classification and the chosen algorithm is Genetic Algorithm which is a heuristic search algorithm that uses cross over, mutation and fitness function to produce the course taking pattern. As part of the test design training, test and validation data sets were generated. The model was created using the following parameter setting: CGPA ≤ 4, time to degree ≥ 3.5 and ≤ 6, programme under consideration – bachelors in accounting, number of courses registered in ≥ 4 and <= 6 and semester = 3. The final model is a table with student records and fields (CGPA, semester and time to degree) (see Table 5) having columns student ID (student_id), Cumulative CGPA (GPA), time to degree(len), and course taking patterns(course).

The time to degree were assessed using the course taking patterns and CGPA see table 5 which showed that time to degree can be associated courses in terms of registered number of courses and CGPA of students only.

(v) Evaluation: The results in Table 5 implied that the time to degree could be interpreted only in terms of the registered number of courses of students (without an observable pattern) for a particular CGPA, considered the highest amongst students belonging to the same programme (BSAF).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

However in the absence of an observable pattern the relationship between time to degree and the registered number of courses can only be interpreted using the number of courses registered in to achieve lower time to degree as in a semester as the attribute .However it can be seen that higher the number of courses, lower will be the time to degree which is perhaps obvious regardless of the CGPA which does not require data mining. Such a result will not be very useful to the HEIs to accurately classify the courses or students to allow them to attain optimal time to degree and CGPA because of lack of additional information that could be used to link course pattern to time to degree significantly. Besides it lacks predictive power because of the obviousness associated with the outcome. Instead if a contextual factor was included in the dataset, for instance course difficulty, then it could act as missing information that could determine a enrolled course pattern of students. In this case the level of course difficulty associated with every course in the set of registered courses of students could be used to determine the optimum time to degree for a given CGPA. For instance student (stud21) has achieved the maximum CGPA of 3.88 possibly because all the six courses could be having a lower difficulty rating. However whether this actually is the case or not, cannot be confirmed due to lack of knowledge about course difficulty in the outcome produced by CRISP-DM process. This point towards the usefulness of contextual factors in discovering knowledge pertaining to course taking pattern. How this can be achieved is discussed in Section V.

(vi) Deployment: The model could not be deployed as evaluation of the outcome of CRISP-DM (see Table 5) showed that the discovered knowledge is not useful to accurately predict the course taking pattern and optimal CGPA and time to degree and classify either the students or courses which is the business goal.

At this point it can be summarized that if the business goal has to be achieved there is a need to discover hidden patterns in the dataset characterized by context which is only possible when the CRISP-DM process has steps to include contextualized dataset and mine it. Thus this research has developed an altered CRISP-DM process that could achieve the above as described next.

Table 5, Course taking patterns generated from Dataset devoid of Contextual Factors Extricated from Student Data

III. DEVELOPMENT OF MODIFIED CRISP=DM PROCESS (STAGE 2)

The modification of CRISP-DM process in line with the summarization provided in the previous paragraph is explained in a step-by-step fashion below taking into consideration the steps described in Section IV. The modified process is depicted Figure 7 and steps A to F in this figure must be related to the discussions below. However in this section only the design aspects of the modification are discussed whereas the corresponding experimental details, numerical examples and interpretations are provided in Section VI. Hence every sub-section in this section could be directly cross referred to sub-sections (A1 to F1) in Section VI to understand how the design aspects have been translated to.

a) Business Understanding (Step A): This step does not significantly change and remain as explained in sub section (i) of section IV except that contextual factors namely course difficulty is also taken into account in prediction of time to degree.

b) Data Understanding (Step B): With respect to subsection (ii) of section IV this step has been divided into namely Contextual (Step B1) and General Data Understanding (Step B2). While step B1 enable the identification of course difficulty (contextual factor) in the dataset and extraction of hidden contextual data (course difficulty) from the dataset (here a data mining method needs to be introduced to discover contextual knowledge), the step B2 is the same as the one described in subsection (ii) of section IV. The outcome of step A can be seen to be linked to both the steps B1 and B2 in a reversible fashion which indicate a basic modification of the original CRISP-DM process.

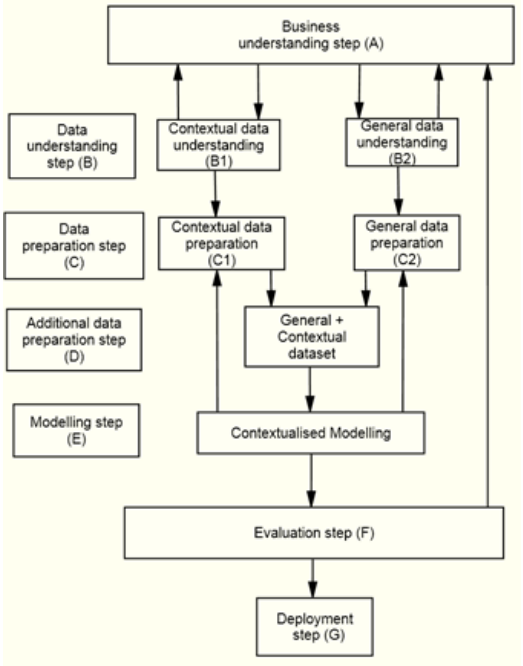


Figure 7, Modified CRISP-DM model

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication (BEIESP)

c) Data Preparation (Step C): Like in Step B this step also have been divided into 2 subsections which step C1 and C2 with respect to subsection C of Section IV. While step C1 deals with contextual data namely course difficulty, in this step data preparation is conducted in similar line as described in subsection (iii) of section IV. Again step C2 is implemented in the same way as described in sub section (iii) of section IV. It can be seen in Figure 7 that the step B1 is linked to step C1 and step B2 is linked to step C2 in a reversible fashion which is another modification of the original CRISP-DM process.

d) Added Data Preparation (Step D): The outcomes of this step are a dataset that has extra information about course difficulty. This step is not originally part of the CRISP-DM process developed Chapman et. al, 2000.In this step the outcomes of steps C1 and C2 are merged into a single dataset that will be mined further. This is a major modification of the original CRISP-DM process.

e) Contextualized Modelling (Step E):This step although very similar to subsection (iv) of section IV differs from that step significantly as it has been modified to handle contextual data using the additional link provided to connect with step C1 and C2 in the reverse direction as well as using a modified Genetic Algorithm that uses course difficulty data. In addition the model was created The model was created using the following parameter setting: CGPA $\leq$ 4, time to degree $\geq$ 3 and $\leq$ 6, programme under consideration – bachelors in accounting and finance (BSAF), number of courses registered in $\geq$ 4 but <=6, semester = 3 and course difficulty which is measured using a 5 point scale (very difficult, difficult, average, easy and very easy). The final model expected to be produced is a table with student records and fields CGPA and time to degree) (see Table 6) which has columns student identification (student_id), time to degree(len), Cumulative CGPA (GPA), course difficulty(difficulty) and course taking patterns. The time to degree were assessed using the course taking patterns and CGPA see table 5 which showed that time to degree can be associated courses in terms of registered number of courses and CGPA of students only.

f) Evaluation (Step F): Time to degree could be interpreted in terms of the registered number of courses of students, with an observable pattern that could be linked to course difficulty for a particular CGPA, considered the highest amongst students belonging to the same programme (BSAF). The use of course difficulty lies in enabling the interpretation of the most optimum course taking pattern of students that could lead to achieve the optimum time to degree and not by mere number of courses. For instance a student could register in six courses all of which could fall under the category of difficult and yet achieve a time to degree of 3 and the highest CGPA. On the other hand a student could register in six courses that could fall under the category of course difficulty "easy" yet achieve a time to degree higher than 3 and lower CGPA. In this case it is possible to interpret that despite registering in the maximum number of courses in a semester, with the difficulty levels measured as "difficult", a student can achieve a lower time to degree and highest CGPA. Such an interpretation could help in identifying and categorizing students who are capable of achieving lower time to degree regardless of the fact that they have registered in courses whose difficulty level could be higher and score high CGPA. The same argument could be used to categorize other students depending on the course taking pattern and the difficulty levels of those courses they have registered in but have taken longer time to degree and scored lower CGPA. This process could be continued to categorize students and the courses they have registered in to determine the level academic support that needs to be provided to enable them achieve optimum time to degree and high CGPA. Thus it can be seen that the business goals of categorizing courses and students to predict optimum time to degree and highest CGPA could be achieved accurately using the contextual factors and a modified process which is demonstrated practically in Section VI.

g) Deployment (Step G): The deployment of the modified CRISP-DM process involves an analysis and understanding of the results produced by the process and at this stage a decision to deploy cannot be taken. Hence this section has been addressed in Section VI where results of the analysis can be used to determine the deployment.

In summary it can be seen that the newly designed modified CRISP-DM process has the potential to achieve the business goal of predicting reasonably accurately the optimum time to degree and CGPA in terms of course taking pattern of students and course difficulty leading to decision making about categorizing courses and students and provide better support to students. To verify this aspect the next section provides the details of the experiments conducted and analysis of the results.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Table 6, Course-taking patterns extracted from Dataset with Contextual Attributes mined from Student Data

Student ID	Timetodegree	CGPA	SGPA	difficulty	course
stud1	4.5	3.33	2.982	Difficult,Difficult,Difficult,Average,Difficult	ACCT 301,FINC 310,FINC 321,FREN 101,STAT 202
stud2	4.5	3.06	2.4	Difficult,Difficult,Difficult,Average,Difficult	ACCT 301,FINC 310,FINC 321,FREN 101,STAT 202
stud3	5	2.44	2.666	Average,Difficult,Average,Difficult,Difficult	ACCT 312,BANK 220,CULT 101,ENGL 202,ITMA 201
stud4	6	2.28	2.4	Average,Difficult,Average,Difficult,Difficult	ACCT 312,BANK 220,CULT 101,ENGL 202,ITMA 201
stud5	4.5	3.1	3.334	Difficult,Average,Difficult,Difficult,Difficult	ACCT 311,ACCT 321,ENGL 202,FINC 321,STAT 202
stud6	4.5	3.49	3.468	Difficult,Average,Difficult,Difficult,Difficult	ACCT 311,ACCT 321,ENGL 202,FINC 321,STAT 202
stud7	4.5	3.75	3.6	Difficult,Difficult,Easy,Difficult,Difficult	ACCT 301,ACCT 311,ARAB 102,FINC 421,ITMA 201
stud8	4.5	3.41	2.8	Difficult,Difficult,Easy,Difficult,Difficult	ACCT 301,ACCT 311,ARAB 102,FINC 421,ITMA 201
stud9	4	3.4	3.668333	Difficult,Difficult,Average,Easy	ACCT 321,ARAB 201,BANK 220,FINC 431,ITCS 121,PHOT 101
stud10	4	3.46	3.723333	Difficult,Difficult,Average,Easy	ACCT 321,ARAB 201,BANK 220,FINC 431,ITCS 121,PHOT 101
stud11	4.5	2.55	2.168333	Difficult,Average,Difficult,Easy,Difficult,Easy	ACCT 321,ACCT 402,BANK 302,ENGL 201,FINC 421,PHOT 10
stud12	4.5	2.33	2.333333	Difficult,Average,Difficult,Easy,Difficult,Easy	ACCT 321,ACCT 402,BANK 302,ENGL 201,FINC 421,PHOT 10
stud13	5	2.5	1	Average,Difficult,Easy,Difficult,Difficult	ACCT 403,BANK 302,ENGL 201,FINC 320,FINC 421
stud14	5.5	2.36	0.8	Average,Difficult,Easy,Difficult,Difficult	ACCT 403,BANK 302,ENGL 201,FINC 320,FINC 421
stud15	4	3.57	3.732	Difficult,Average,Difficult,Difficult,Average	ACCT 301,CULT 102,ENGL 202,FINC 320,ITCS 121
stud16	4	3.84	3.666	Difficult,Average,Difficult,Difficult,Average	ACCT 301,CULT 102,ENGL 202,FINC 320,ITCS 121
stud17	4	3.6	3.934	Average,Average,Difficult,Easy,Easy	ACCT 402,ACCT 403,BANK 302,ENGL 201,VDEO 101
stud18	4	3.22	3.202	Average,Average,Difficult,Easy,Easy	ACCT 402,ACCT 403,BANK 302,ENGL 201,VDEO 101
stud19	4	3.42	3.398	Average,Average,Easy,Difficult,Difficult	ACCT 312,ACCT 320,ACCT 341,BANK 302,FINC 310
stud20	4	3.23	3.2	Average,Average,Easy,Difficult,Difficult	ACCT 312,ACCT 320,ACCT 341,BANK 302,FINC 310
stud21	3	3.88	3.778333	Difficult,Difficult,Difficult,Difficult,Difficult,Difficult	ACCT 404,BANK 302,ECON 301,FINC 320,FINC 421,STAT 202
stud22	3	3.53	3.556667	Difficult,Difficult,Difficult,Difficult,Difficult,Difficult	ACCT 404,BANK 302,ECON 301,FINC 320,FINC 421,STAT 202

The input to the KDDM process is the dataset characterised by contextual factor course difficulty. The business goals to be achieved are determining the optimum time to degree and CGPA, categorisation of courses that could predict optimum time to degree and CGPA using course registration patterns and course difficulty. Categorisation of students based on optimum time to degree, CGPA, course patterns and course difficulty. Data preparation and data understanding stage covers the data quality factors of data. For detailed steps, readers have to refer to chapman et.al(2000). Genetic algorithm was used with knowledge of contextual factors course difficulty. Model evaluation stage comprises of pattern extracted through the process with and without contextual detection stage.The overall CRISP-DM process has not been discussed in detail and the discussion here is limited to demonstrating the difference between the knowledge discovered using patterns extracted through CRISP-DM process with and without the detection of course difficulty and semester as contextual factor. As far as the introduction of contextualised dataset was concerned initially the characteristics of course difficulty pertaining to every course under consideration in a particular semester 3 were utilised which included 'course GPA', 'weighted average of the course GPA', 'current course'; 'course equivalent to the current one', 'Set of equivalence courses for course', 'Total number of students in the course'. Using these characteristics pertaining to course difficulty a scale was developed to signify the level of difficulty of each course. The course difficulty was measured over 5 point scale very easy, easy, average, difficult and very difficult.

I. EXPERIMENTS ON MODIFIED CRISP-DM PROCESS AND ANALYSIS OF RESULTS(STAGE 3)

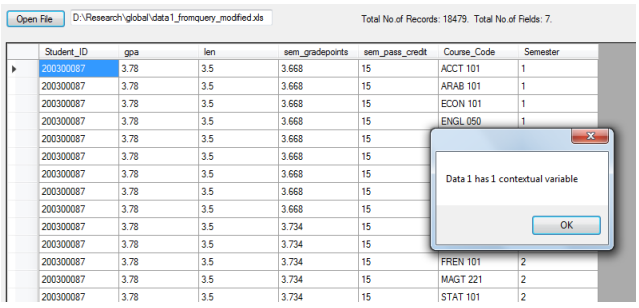
This section describes experiments conducted on modified CRISP-DM process given in figure 7. The subsection (a1) to (g1) in this section has one to one correspondence to the subsection A to G.in section V. Thus each subsection provides the experimental and analytical details conducted according to the design specification provided in the subsection A to G in section V.

a1) Business understanding(Step A): predicting reasonably accurately the optimum time to degree and CGPA in terms of course taking pattern of students and course difficulty leading to decision making about categorising courses and students and provide better support to students.

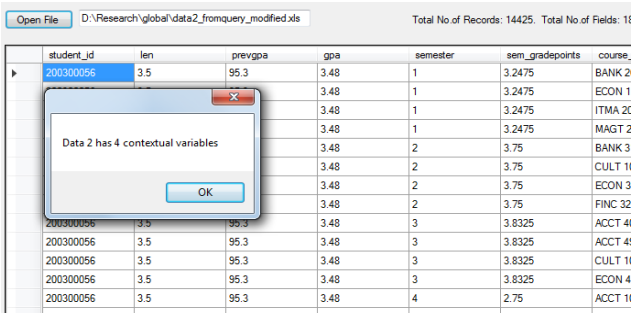
b1) Data Understanding (Step B): This step is divided into two: one addressing contextual data and the other addressing general data.

b1.1 Finding the existence or nonexistence course difficulty using a specially designed algorithm that was based on the formula devised by Zainuddin,2012 to compute the course difficulty level(refer Zainuddin,2012) and a pseudo code written based on the guideline given by Vert et.al,2010.

The outputs derived from the execution of the pseudocode is shown in the screen shot below.



Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication (BEIESP)



b1.2 General Data

This section is same as previous sections as it deals with general data of students. The data extracted using the SQL queries is provided in Table 3.

c1) Data Preparation (Step C): Both the outputs of b1.1 and b1.2 were subjected to the same steps mentioned subsection iii of section IV.

d1) Added Data Preparation (Step D): The data prepared in step C1 and C2 see figure 7 are merged and the resulting dataset is shown in Table 4. This dataset is further fed into the modeling stage.

Table 4, Data set with contextual factors extracted

student_id	len	prevgpa	gpa	semester	sem_gradepoints	course_code	difficulty
2E+08	3.5	95.3	3.48	4	2.75	ACCT 101	Very Difficult
2E+08	3.5	95.3	3.48	4	2.75	ARAB 101	Difficult
2E+08	4	97.6	3.87	1	3.868	ACCT 101	Very Difficult
2E+08	4	97.6	3.87	2	3.466	ACCT 101	Very Difficult
2E+08	4	97.6	3.87	2	3.466	ARAB 101	Difficult
2E+08	4	97.6	3.87	5	3.868	ARAB 101	Difficult
2E+08	3.5	95.4	3.78	1	3.668	ACCT 101	Very Difficult
2E+08	3.5	95.4	3.78	1	3.668	ARAB 101	Difficult
2E+08	3.5	96.2	3.84	1	4	ARAB 101	Difficult
2E+08	3.5	96.2	3.84	2	3.6	ACCT 101	Very Difficult
2E+08	3.5	96.2	3.84	2	3.6	ARAB 101	Difficult
2E+08	3.5	96.2	3.84	5	3.776667	ACCT 101	Very Difficult
2E+08	4	77.3	2.18	1	1.553333	ARAB 101	Difficult

e1) Modelling (Step E): This step provides the model generated at step E see of the modified CRISP-DM process using the experimental details provided in subsection e in section V. The model is a report is shown in Table 7 (KJ Lu, 2016).

Table 7, Course taking patterns extracted from Dataset with Contextual Factors Extracted from Student Data (KJ Lu, 2016).

Student ID	Timetodegree	CGPA	SGPA	difficulty	course
stud1	4.5	3.33	2.992	Difficult,Difficult,Difficult,Average,Difficult	ACCT 301,FINC 310,FINC 321,FREN 101,STAT 202
stud2	4.5	3.06	2.4	Difficult,Difficult,Difficult,Average,Difficult	ACCT 301,FINC 310,FINC 321,FREN 101,STAT 202
stud3	5	2.44	2.666	Average,Difficult,Average,Difficult,Difficult	ACCT 312,BANK 220,CULT 101,ENGL 202,ITMA 201
stud4	6	2.28	2.4	Average,Difficult,Average,Difficult,Difficult	ACCT 312,BANK 220,CULT 101,ENGL 202,ITMA 201
stud5	4.5	3.1	3.334	Difficult,Average,Difficult,Difficult,Difficult	ACCT 311,ACCT 321,ENGL 202,FINC 321,STAT 202
stud6	4.5	3.49	3.468	Difficult,Average,Difficult,Difficult,Difficult	ACCT 311,ACCT 321,ENGL 202,FINC 321,STAT 202
stud7	4.5	3.75	3.6	Difficult,Difficult,Easy,Difficult,Difficult	ACCT 301,ACCT 311,ARAB 102,FINC 421,ITMA 201
stud8	4.5	3.41	2.8	Difficult,Difficult,Easy,Difficult,Difficult	ACCT 301,ACCT 311,ARAB 102,FINC 421,ITMA 201
stud9	4	3.4	3.668333	Difficult,Difficult,Average,Easy	ACCT 321,ARAB 201,BANK 220,FINC 431,ITCS 121,PHOT 101
stud10	4	3.46	3.723333	Difficult,Difficult,Average,Easy	ACCT 321,ARAB 201,BANK 220,FINC 431,ITCS 121,PHOT 101
stud11	4.5	2.55	2.168333	Difficult,Average,Difficult,Easy,Difficult,Easy	ACCT 321,ACCT 402,BANK 302,ENGL 201,FINC 421,PHOT 101
stud12	4.5	2.33	2.333333	Difficult,Average,Difficult,Easy,Difficult,Easy	ACCT 321,ACCT 402,BANK 302,ENGL 201,FINC 421,PHOT 101
stud13	5	2.5	1	Average,Difficult,Easy,Difficult,Difficult	ACCT 403,BANK 302,ENGL 201,FINC 320,FINC 421
stud14	5.5	2.36	0.8	Average,Difficult,Easy,Difficult,Difficult	ACCT 403,BANK 302,ENGL 201,FINC 320,FINC 421
stud15	4	3.57	3.732	Difficult,Average,Difficult,Difficult,Average	ACCT 301,CULT 102,ENGL 202,FINC 320,ITCS 121
stud16	4	3.84	3.666	Difficult,Average,Difficult,Difficult,Average	ACCT 301,CULT 102,ENGL 202,FINC 320,ITCS 121
stud17	4	3.6	3.934	Average,Average,Difficult,Easy,Easy	ACCT 402,ACCT 403,BANK 302,ENGL 201,VDEO 101
stud18	4	3.22	3.202	Average,Average,Difficult,Easy,Easy	ACCT 402,ACCT 403,BANK 302,ENGL 201,VDEO 101
stud19	4	3.42	3.398	Average,Average,Easy,Difficult,Difficult	ACCT 312,ACCT 320,ACCT 341,BANK 302,FINC 310
stud20	4	3.23	3.2	Average,Average,Easy,Difficult,Difficult	ACCT 312,ACCT 320,ACCT 341,BANK 302,FINC 310
stud21	3	3.88	3.778333	Difficult,Difficult,Difficult,Difficult,Difficult,Difficult	ACCT 404,BANK 302,ECON 301,FINC 320,FINC 421,STAT 202
stud22	3	3.53	3.556667	Difficult,Difficult,Difficult,Difficult,Difficult,Difficult	ACCT 404,BANK 302,ECON 301,FINC 320,FINC 421,STAT 202

f1) Evaluation (Step F): In line with the details of the design given in sub-section (f) of Section V the results obtained were evaluated. Evaluation included exploratoring the course taking pattern by allocating the course difficulty level for each enrolled course and links the course pattern to time to degree and assess the scored CGPA (Table 7). The course difficulty level was calculated using a five point scale (see Section V). The conjecture here is that more the number of enrolled courses of a student having difficulty level is greater than or equal to 'difficult' then more would be the time to degree and lesser will be the CGPA. For instance, take a student e.g. stud 6 in Table 7 who is shown to have registered in five courses of these the course difficulty is 'Difficult' with the left over 1 course as 'Average'.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The student CGPA is 3.49. It is sensible to anticipate that this student will take more time to graduate as he was able to finish the entire number of 44 courses in 4.5 years on an average of five courses in a semester and a CGPA around the same figure of 3.49. The course difficulty stage has possibly not allowed the student to enrol in more courses in a semester, say 6. Had 'stud6' enrolled in 6 courses/semester, then the student might have graduated in 4 years. Thus course taking pattern and the corresponding course difficulty measure of a course can be believed to have an effect on the time to degree. On the other hand if 'stud6' had registered in 6 courses whose course difficulty scores were a combination of 'Easy',' Average' and 'Difficult' (for instance, a course taking pattern of 6 courses may be ARAB 102, ACCT 320,ENGL 201, BANK 302,STAT 202 with course difficulty measured as'Easy', 'Average','Easy','Difficult','Difficult' then there is a probability that 'stud6' might have scored the same CGPA of 3.49 or more and had taken lesser time to degree (4 years). Therefore it can be witnessed that course taking pattern if linked with the course difficulty then it is probable to forecast the optimal time to degree at an optimal CGPA. That is to say, it is possible to anticipate that when students enrol in courses where difficulty level is "difficult" then lesser would be the number of such courses a student could enrol owing to harder effort to complete with a higher CGPA. From this situation, it is probable to deduce that when students enrol in a set of courses in which difficulty levels are "difficult", then the no. of courses in that set is liable to be lesser. This might be because, more the course difficulty level, difficult is the attempt needed to finish the course with higher CGPA. Therefore lower number of courses in the set provides greater opportunity to work hard leading to better chance of scoring higher CGPA at the cost of time to degree. That is:

f11) Course difficulty negatively influences the number of courses registered in a semester.

f12) Course taking pattern characterized by high number of difficult courses negatively influences CGPA and positively influences time to degree.

f13) Course taking pattern characterized by low number of difficult courses positively influences CGPA and negatively influences time to degree.

However this inference was not so obvious to predict and generalize when one considers the hidden knowledge in the dataset and results provided Table 7. Entirely opposite results have been obtained in the case two students namely 'stud21' and 'stud22'. From Table 7 it can be seen that the statements in (f11), (f12) and (f13) are completely falsified. The evaluation of the results of the students 'stud21' and 'stud22' clearly shows:

f14) Course difficulty does not influence the number of courses registered in a semester. The students 'stud21' and 'stud22' have registered in a maximum of six courses all of them rated as 'difficult'.

f15) Course taking pattern characterized by high number of difficult courses does not influence CGPA and negatively influences time to degree. It can be seen that the pattern of courses registered in by the student 'stud21' although comprises six courses whose level of difficulty is rated as 'difficult' has the scored the maximum CGPA and achieved the lowest time to degree of 3 years. Similarly for the student 'stud22' the course taking pattern although comprising of six courses whose level of difficulty is rated as 'difficult' has not scored the maximum CGPA but achieved the lowest time to degree of 3 years. This shows that the course taking pattern characterized by high number of difficult courses does not influence CGPA but negatively influences time to degree.

f16) Course taking pattern characterized by low number of difficult courses positively influences CGPA and negatively influences time to degree is falsified because high number of course rated as difficult can also positively influence CGPA and negatively influence time to degree as mentioned in (f15).The previous findings and discussions evidently point out that the optimal time to degree and the maximum CGPA can be fixed by a set of six courses (highest authorized by a HEI in a semester, refer section III), with every course fixed as 'Difficult'. When this finding is treated as optimal, in that case the following decision could be facilitated in the HEIs.

1. Categorizing students akin to 'stud21' and 'stud22' to allow them to graduate in 3 years.
2. Categorizing students akin to 'stud6' who have enrolled in just 5 courses and investigate their performance in terms of their ability to score the same CGPA or higher to enable them to achieve a performance similar to 'stud21' and 'stud22'.
3. Categorizing students like 'stud17' and 'stud18'who have taken 4 courses and analyse their performance in terms of their ability to score the same CGPA or higher to enable them to achieve a performance similar to 'stud21' and 'stud22'.
4. Assert the performance of students who have enrolled in 4 or 5 courses in a semester at an early stage in their academic career in the university using modified CRISP-DM to forecast their optimal time to degree with higher CGPA based on past student results.

g1) Deployment (Step G): The deployment of the modified CRISP-DM process involves an analysis and understanding of the results produced by the process. From the discussions provided in above section f1 it is clear that the modified CRISP-DM process can enable HEIs to achieve business goals using the discovered knowledge in terms of course taking patterns and the context of course difficulty to determine optimum time to degree and CGPA. The results obtained are consistent and hence the tailored CRISP-DM process can be deployed through teaching and learning and achieve the business goal of supporting students to achieve to optimize the time to degree and CGPA using course taking patterns and course difficulty as parameters. The tests deployed in modifying CRISP-DM process are very similar to original CRISP-DM process.

I. CONCLUSION, LIMITATION AND FUTURE RESEARCH

The results obtained in this research as explained in Sections V and IV clearly point out to that the main objective of demonstrating a modified KDDM process that extracts hidden knowledge from the student dataset characterized by contextual factors that could be useful in making better decisions in HEIs to achieve specific business goals has been achieved.

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication (BEIESP)

The modified process uses CRISP-DM as an example which has not been deployed in the context HEIs. Thus the outcome of this research also provides a demonstration of the use of CRISP-DM process in HEIs. Finally the modified CRISP-DM process provides a DM process that could uncover contextual data to be used to support business goal and produce a dataset at the preparation stage which is contextualized. This provides a new way of dealing with the model development at the DM stage in the CRISP-DM process. At this stage a modified genetic algorithm has been specifically developed to enable the CRISP-DM process to generate a model that is contextual leading to the discovery of course taking patterns that are contextualized. This discovery has enables prediction of optimum time to degree and CGPA, an aspect not covered in the extant literature. As far as limitations of the research are concerned, it must be pointed out that this research has used only one contextual factor namely course difficulty of the registered courses of just 25 students in only one semester. There is a need to cross-check this with more number of semesters and additional contextual factors for greater number of students. Future research could investigate into more number of semesters and use additional contextual factors including student potential, course weight age and course complexity which promise to enable decision makers to make more accurate decisions.

REFERENCES

1. Adelman, C., 1999. Answers in the toolbox: Academic Intensity,attendance patterns, and Bachelor's Degree attainment. Washington DC, Department of Education, Office of Educational Research and Improvement.
2. Astin, A., 1971. Predicting academic performance in college: Selectivity data for 2300 American colleges.. New York: The Free Press.
3. Athiyaman, A., 1997. Linking student satisfaction and service quality perceptions: The case of university education. European Journal of Marketing, 31(7), pp. 528-540.
4. Bahr, P. R., 2010. The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification.. Research in Higher Education, 51(8), pp. 724-749.
5. Bolchini, C. et al., 2007. A dataoriented survey of context models. SIGMOD Rec.(ACM) doi:10.1145/1361348.1361353. ISSN, 36(4), pp. 19-26.
6. Brachman, R. & Anand, T., 1996. The process of knowledge discovery in databases:A human-centered approach.In U. Fayyad, G. Paitestsky-Shapiro, P. Smith, & R. Uthuruswamy (Eds.), Advances in knowledge discovery and data mining. AAAI Press, pp. 36-57.
7. Catley, C., Smith, K., McGregor, C. & Tracy, M., 2009. Extending CRISP-DM to Incorporate Temporal Data Mining of Multi dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study. s.l., Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium.
8. Chapman, P., Clinton, J., Kerber, R. & Khabaza, T., 2000. CRISPDM 1.0 step-by-step data mining guide. Technical report, CRISP-DM , s.l.: CRISP-DM.
9. Daempfle, P. A., 2003. An Analysis of the High Attrition Rates Among First Year College Science,Math,and Engineering Majors. Journal of College Student Retention, 5(1), pp. 37-52.
10. Davenport, T. . H., 2010. The New World of "Business Analytics", s.l.: International Institute for Analytics.
11. DeShields, O. W., Kara, A. & Kaynak, E., 2005. Determinants of business student satisfaction and retention in higher education: Applying Herzberg's two-factor theory.. International Journal of Educational Management, 19(2), pp. 128-139.
12. Dey, A. K., 2001. Understanding and Using Context. Personal Ubiquitous Comput., 5(1), pp. 4-7.
13. Elliott, K. M. & Healy, M. A., 2001. Key factors influencing student satisfaction related to recruitment retention. Journal of Marketing for Higher Education, 10(4), pp. 1-11.
14. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R., 1996. Advances in knowledge discovery and data mining.. MIT Press..
15. Fettke, P., Vella, A. L. & Loos, P., 2012. From Measuring the Quality of Labels in Process Models to a Discourse on Process Model Quality: A Case Study. Maui, HI , IEEE.
16. Helgesen, O. & Nesset, E., 2007. What accounts for students' loyalty? Some field study evidence.. International Journal of Educational Management, 21(2), pp. 126-143.
17. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.
18. John, G. H., 1997. Enhancements to the data mining process. s.l.:PhD thesis, Stanford University.
19. Knight, W. E., 1994. Why the five-year (or longer) bachelors degree ?:An exploratory study of time to degree attainment. New Orleans, LA, Association for Institutional Research forum.
20. Kovacic, Z. J., 2010. Early prediction of student success: Mining student enrollment data. s.l., Proceedings of Informing Science & IT Education Conference .
21. Kurgan, L. & Musilek, P., 2006. A survey of knowledge discovery and data mining process models.. Knowledge Engineering Review, 21(1), pp. 1-24.
22. Li, J., Yang, B. & Song, W., 2009. A New Data Mining Process Model for Aluminum Electrolysis. Qingdao, P. R. China, Proceedings of the International Symposium on Intelligent Information Systems and Applications (IISA'09).
23. Lotkowski, V. A., Robbins, S. B. & Noeth, R. J., 2004. The Role of Academic and Non-Academic Factors in Improving College Retention, Iowa City, IA: ACT Policy Report.
24. Marbán, Ó., Mariscal, G. & Segovia, J., 2009. A Data Mining & Knowledge Discovery Process Model. I-Tech, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, Julio Ponce and Adem Karahoca (Ed.),ISBN: 978-3-902613-53-0.
25. Minaei-Bigdoli, B., Kashy, D. A., Kortemeyer, G. & Punch, W. F., 2003. 33rd ASEE/IEEE Frontiers in Education Conference. Boulder,CO, IEEE.
26. Pheng, L. S. & Arain, F. M., 2006. A KNOWLEDGE-BASED SYSTEM AS A DECISION MAKING TOOL FOR EFFECTIVE MANAGEMENT OF VARIATIONS AND DESIGN IMPROVEMENT: LEVERAGING ON INFORMATION TECHNOLOGY APPLICATIONS. ITcon, Volume 11.
27. Rao, K., Govardhan, A. & Rao, K., 2012. SPATIOTEMPORAL DATA MINING: ISSUES, TASKS AND APPLICATIONS. International Journal of Computer Science & Engineering Survey (IJCSES), 3(1).
28. Redpath, R. & Srinivasan, B., 2004. A Model for Domain Centered Knowledge Discovery in Databases. Budapest, Hungary, Proceedings of the IEEE 4th International Conference On Intelligent Systems Design and Application August,(ISDA 2004), ISBN: 9637154302.
29. Rennolls, K. & Al-Shawabkeh, A., 2008. Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. Intelligent Data Analysis, 12(2), p. 147–163.
30. Ronco, S. L., 1996. How Enrollment Ends: Analyzing the Correlates of Student Graduation, Transfer and Dropout with a Competing Risks Model, Tallahassee, Fla.: AIR Professional File, No. 61 Association for Institutional Research.
31. Ruan, T. . L. D., 2007. An extended process model of knowledge discovery in database. Journal of Enterprise Information Management, 20(2), pp. 169 - 177.
32. SAS, I., 2008. SEMMA data mining methodology. [Online] Available at: http://www.sas.com
33. Schilit, B., Adams, N. & Want, R., 1994. Context-Aware Computing Applications. s.l., First International Workshop on Mobile Computing Systems and Applications.
34. Sharma, S., Osei-Bryson, K.-M. & Kasper, G. M., 2012. Evaluation of an integrated Knowledge Discovery and Data Mining process model. Expert Systems with Applications, Volume 39, p. 11335–11348.
35. Tinto, V., 1975. Dropouts from higher education: A theoretical synthesis of recent literature. A Review of Educational Research, Volume 45, pp. 89-125.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

36. Vert, G., Chennamaneni, A. & Iyengar, S. S., 2010. Potential Application of Contextual Information Processing To Data Mining.. Las Vegas Nevada, USA., Proceedings of the 2010 International Conference on Information & Knowledge Engineering, IKE 2010, July 12-15, 2010.
37. Vialardi, . C., Chue, J., Peche, J. P. & Alvarado, G., 2011. A data mining approach to guide students through the enrollment process based on academic performance. User Modeling and User - Adaptation Interaction, Volume 21, pp. 217-248.
38. Zhang, C., Yu, P. . S. & Bell, D., 2010. Introduction to the Domain-Driven Data Mining Special Section. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 22(6).
39. BIS (2014) Improving the Student Learning Experience – a national assessment (BIS Research Paper no 169). London: Department for Business, Innovation and Skills.
40. G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
41. Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, "Data Mining: A prediction
42. for Student's Performance Using Classification Method", World Journal of Computer Application and Technology 2(2): 43-47, 2014
43. Lu, K.J. and Sailesh, S.B., 2016. Context driven data mining to classify students of higher educational institutions.
44. R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. IEEE Trans. Plasma Sci. [Online]. 21(3). pp. 876—880. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar

AUTHORS PROFILE

Dr. Subhashini Bhaskaran Sailesh, Assistant Professor, Ahlia University, Bahrain Research interests include Data Mining, Data Science Business Analytics, Data Analytics

Prof. Mansoor Al Aali, President, Ahlia university, Bahrain Research interests include Data Mining, Data Science, Artificial Intelligence

Dr. Kevin Lu Brunel University London