Intelligent Context Driven Data Mining To Analyse Student Performance in Higher Educational Institutions (HEIs)

Subhashini Sailesh Bhaskaran

Abstract: Context driven research has been conducted by many. However, very less work has been conducted in constructing a context driven data mining that helps in HEI decision-making. A Student Information System that interacts with Students, Faculties, Student Parents and Management might not have enough information of the background. Context driven data mining is an application intelligent enough to detect and examine the context from different sources and take suitable actions to improve performance and efficiency of decision-making by discovering the hidden factors. This paper recommends a context driven data mining method for understanding student performance from Student Information System in HEIs.

Keywords: HEIs; Context-Awareness; Data Mining; Student Performance;

I. INTRODUCTION

Environments and semantics of meaning explaining an event inspires context driven processing. An illustration of an event can be elucidated by means of spatiotemporal dataset that is likely to clarify about a spatial or temporal incident that may happen at a specific time t and location x (Rao et al., 2012). For example catastrophes like earthquakes or floods could be measured as spatiotemporal events. The datasets possessing data on these events could be considered to depict the location and time when the event took place. Considering similar arguments to HEI student performance, this research aspires to construct an intelligent context-aware data mining methodology for mining student data and unearth knowledge. This unearthed knowledge can enable HEIs to take decisions that are high quality and enhanced those made using existing processes. Data mining helps in extracting hidden knowledge from large volumes of data named as Big Data for helping users to make better decisions in their business, usually, giving recommendations and predictions. The importance of data mining has lately led to the boom of a number of context-aware data mining. Generally widely considered context dimensions comprise of physical location, time, user preferences and capabilities. HEI context information is seldom taken into account. For instance information including demographic characteristics of students enrolled in various programmes, time to degree of students and GPA of students in HEIs is purported to be available as a part of student data and warehoused using computer systems..

Since this knowledge is not easily visible as tangible factors, institutions are not been able to use them to make decisions in many areas by HEIs like student assessment, student performance, strategies related to student support, student administration and teaching issues. Further the knowledge extracted using data mining seem to be deficient in deepness and might be lack of essential information that is concealed in the data cannot be extracted

The outcome is the utilization of partial knowledge in HEI decision making. For example HEI students' course taking and contextual knowledge related to such patterns cannot be generated from HEI student dataset, by means of simple query or algorithms or manual ways. Decisions made in HEIs regarding student performance which does not engage contextual information could be incomplete. A context-aware process will sense and examine the context from numerous sources and takes activities suitable to contexts. Context driven data mining help in adding various types of intelligence to our environment. Our contribution is presented in a context-aware data mining methodology for analysing student performance in HEIs and a proposed pseudocode for validating a dataset for presence of context. The remainder of the paper presents taxonomy of context from some of the main researches and proposes a data mining system architecture for HEIs, and then concentrates on the context driven data mining method.

II. RELATED LITERATURE

A. Taxonomy of Context

In the literature various concepts or theories or terminologies related to context are found which are summarised as follows. Context analysis is a method to analyse the environment in which a business operates. Contextual investigation is an area in which data collection technique that is used to get comprehensive information about the way the consumer of a product interacts with the product in their normal surroundings. In mobile applications, main characteristic of the technique is a comprehensive cataloguing of their physical surroundings factors, basically attained by cataloguing mobile sensor data. The contextual design process contains the following: contextual inquiry, data consolidation and analytics, visioning, storyboarding, user environment design, and prototyping. Contextual application design in digital media is the adaptation of the contextual design process A contextual data analytics collects data logs, events and context awareness parameters and streams them into an easy to process database which may be queried to provide insights. The contextual analysis helps to assess the text. The focal point of the Context Data Model is to offer a basis for unifying, integrating, and sharing identity-related data.

Retrieval Number: B1842078219/19©BEIESP
DOI: 10.35940/ijrte.B1842.078219
Journal Website: www.ijrte.org

856

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Specifically, we are paying attention on information regarding a person, a group or an entire organization. This will comprise of authentication data, contact information, email addresses, preferences, interests, employer-related information. However many researchers like the above have emphasized on the significance of context to HEIs and data mining. Baker(2008) has highlighted that problems of context, time and sequence demonstrate substantial role in the investigation of educational data. It is also stated that presence of enormous students' data from alike education experiences however in very diverse environments urges for reviewing the effect of contextual factors on learning and learners. There is an argument by many researchers that without contextual information, decisions taken in business might be erroneous. When context, is used establishments can develop tendencies, patterns, and associations that can help a business to create fact-based decisions. Creation of data inside the relevant context brings higher quality models that can help in better decisions and results. In spite of a lot of emphasis on context, very little has been done in datamining and HEIs. The above-mentioned arguments evidently indicates the need to investigate the importance and effect of context on data mining in HEIs.The taxonomy of context present in some of the major researches is shown in Table 1.

III. STUDENT INFORMATION SYSTEM MODEL FOR HEIs

Figure 1 demonstrates the abstract level of the model for
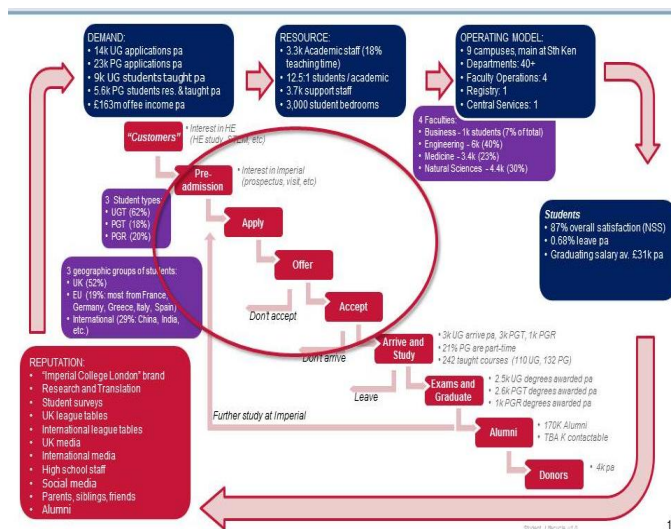analysing student performance in HEIs.

Fig 1: Logical Process of Student Life Cycle in HEI (Adapted from Imperial College, London)

The student life cycle in HEI starts with application process as a prospective applicant after which the HEI processes the application to offer or reject the application. The applicant then accepts or rejects the offer from the HEI. If the applicant accepts the offer the status improved to a student after which the student studies in the HEI as full time or part time. The student is assessed through examinations and finally graduates after completing all the graduation requirements and completion of the study plan. There are a

lot of environment factors that can affect the student performance and graduation in HEI like student background factors, HEI factors, faculty factors and other factors that are unknown. The student faces a lot of challenges in liberal study plan or in American education system where the student has the liberty to enrol to courses as per choice (except for courses with pre requisites) and can choose the learning path. The learning path even though is within the study plan provided for programme but has a lot of variations and combinations of courses that have to be studied in a semester. Having in mind that there are semesters varying between 8 and 16 in a student life cycle within which around 44 courses have to be registered and completed with probabilities of success associated with the courses are voluminous. There is a lot of information or knowledge hidden in the student learning environment or in the study plans which will have to be unearthed in order to advise the students and for planning resources in HEIs.

Table 1: Taxonomy of Context

Meaning of Context	Author and Paper	Field	Remarks	Experimentation or Analysis	Terms
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 By Lisa Sokol, Ph.D, IBM Software Group, US Federal CTO, and Steve Chan, Ph.D, 2013	Big Data	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.	Nil	Context Analytics, Big data Analytics, Real-time analytics, Deep reflection analytics, predictive analytics
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, Pari Delir Haghighia , Arkady Zaslavskya , Shonali Krishnaswamya , Mohamed Medhat Gaberb , and Seng Lokec, 2009	Mobile devices/PDA Datamining	The researchers proposed a 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.	A prototype was tested on the Nokia N95 mobile phone to represent a real-world scenario in the area of mobile healthcare for monitoring patients suffering from blood pressure fluctuations	Context, fuzzy logic, data mining
Context – background information Contextual itemset mining extracts frequent associations among items considering background information.	Contextual Itemset Mining in DBpedia, Julien Rabatel1 , Madalina Croitoru1 , Dino Ienco2 , Pascal Poncelet1, 2014	Data mining	The authors exhibit the capacity of contextual itemset mining. Contextual itemset mining excerpts frequent associations between items bearing in mind the background information. Each resultant itemset is specific to a particular context and contexts can be related to each other's following the ontological structure. They used contextual mining on DBpedia data and proved the use of contextual information can re_ne the itemsets obtained by the knowledge discovery process.	Algorithm has been proposed and experimented	contextual frequent pattern (CFP),Linked Open Data,
Context awareness is a general concept that refers to the capability of a system to be aware of its physical and logical environment (i.e., its context) and to intelligently react according to this awareness.	A Contextual Data Mining Approach Toward Assisting the Treatment of Anxiety Disorders Theodor Chris Panagiotakopoulos, Dimitrios Panagiotis Lyras, Miltos Liyaditis, Kyriakos N. Sgarbas, Member, IEEE, George C. Anastassopoulos, and Dimitrios K. Lymberopoulos, Member, IEEE,2010	Health care, data mining	In their research the general context information is provided through the patient context which has four subcontexts, that is environmental, personal, stress and symptoms context. Four treatment-supportive services like lifestyle and habits pattern detection service, context and stress level pattern detection service, symptoms and stress level pattern detection service, and stress level prediction service were proposed by the researchers. The first three facilities target at determining possible associations inside the patient's multiparameter dynamic context data hence Apriori association rule mining algorithm was used. The last one aims at creating a patient-dependent stress level prediction model that would be able to accurately predict the stress a patient will present under a specific environmental context hence the Bayesian network predictive classifier was employed. In order to examine the efficiency and usefulness of the proposed treatment-supportive services they performed quantitative analysis that showed that all the services are computationally efficient and that they can be embedded into real-time healthcare applications.	Apriori Association rule, Bayesian network	Context awareness, machine learning, mental health, user modeling.
Context refers to the background in which the consumer review was given in opinion mining	Mining Context Information from Consumer's Reviews Silvana Aciar,2010	Data mining	The authors have tried to address the most critical issue on opinion mining which is how to extract information that can be understood and utilized by computers from written text by users/consumers in natural language. They employed classification text mining techniques to identify review's sentences containing contextual information to be then processing and incorporated in a recommender system.	use of text mining tools to obtain classification rules to identify contextual sentences containing contextual information into a review	Contextual Information, Opinion Mining, Text Mining, sentiment analysis
Processes are performed in a particular context, but this context is often neglected during analysis in process mining which is used to discover and analyze business processes based on raw event data	Process Mining Put Into Context Wil M.P. van der Aalst and Schahram Dustdar,2012	Data mining	The authors argue that the contexts in which the events take place should be considered when the processes are analysed. They distinguish four types of contexts like instance context, process context, social context and external context.	Nil	Process mining,
Contextual information like vehicle dynamics, drivers' physiological condition, driver's profile and environmental conditions could help to explain why a crash is imminent and improve the accuracy of crash prediction	Towards situation-awareness and ubiquitous data mining for road safety: rationale and architecture for a compelling application, Shonali Krishnaswamy, Seng Wai Loke, Andry Rakotonirainy, Osnat Horovitz and Mohamed Medhat Gaber, 2005	Mobile devices , Data mining	The authors presented an approach based on Ubiquitous Data Mining (classification algorithms) to decrease human errors by monitoring driving risks in real time. They proposed an architecture of proficient of estimating risks. The type of risks that could be possibly monitored by their system include fatigue, roll over, speed and inexperience which are among the factors addressed by the Australian road safety priorities.	Classification algorithms	Ubiquitous Data Mining, Situation Awareness, and Road Safety

Use of data mining applied to sensor data to better understand unobservable system operating environments to support the requirements elicitation process.	A Case Study of Applying Data Mining to Sensor Data for Contextual Requirements Analysis Angela Rook, Alessia Knauss, Daniela Damian, Alex Thomo Department of Computer Science IEEE , 2014	Data mining, mobile data	Initially the J48 (C4.5) data mining machine learning algorithm is applied to historical sensor data to automatically identify which context situations are relevant to five specific requirements. Next, the results of applying the data mining algorithm to historical sensor data is compared against the actual context situations in which specific requirements are valid. The results of this approach inform a discussion of the application of data mining algorithms in this domain, as well as the application of the approach to other domains.	J48 algorithm	Context-aware systems
Context could consist of any circumstantial factors of the user and domain that may affect the data mining process.	Context-Aware Data Mining Framework for Wireless Medical Application Pravin Vajirkar, Sachin Singh, and Yugyung Lee, 2003	Datamining	The authors proposed a context aware framework which considers context factors during datamining. In their framework, Context-aware Process Component deals with context factors for Data mining and Query processor components. In this component, Context factors such as Application Context, User Context, Domain Context, Data Context are defined. In order to identify context factors from the implicit information of the user's query, processes including checking the attributes of the user query; learning about the entities of the query context are performed. They showed 2 scenarios as a proof of their concept.	Nil	Context-aware data mining
Notion of a context with that of the idea that information could be used to characterize a situation and thus could be responded to.	Potential Application of Contextual Information Processing To Data Mining, Vert et. al IKE 2010	Data mining	The authors have proposed contextual information processing to be as a part of data mining process. No experiments or statistical proofs are given for the same.	Nil	Contextual information processing

The following are some of the challenges in developing a context-aware data mining methodology for analysing a real time student performance in HEIs (1) predicting the next course for a student so he/she can graduate on time with high GPA (2) finding student and HEI information relevant to the ontext and content of the event; (3) searching and bringing the potentially needed prediction information closer to the student that can be practically applied to get desirable results. and (4) finding if a dataset that is fed into the data mining framework has contextual information in it or not which will help the dataminer to be assured of the modelling results which will decide the future of the student.

IV. STUDENT PERFORMANCE CONTEXT-AWARE DATA MINING FRAMEWORK FOR HEIs

The proposed student performance context-aware data mining framework for HEIs is given in fig 2.
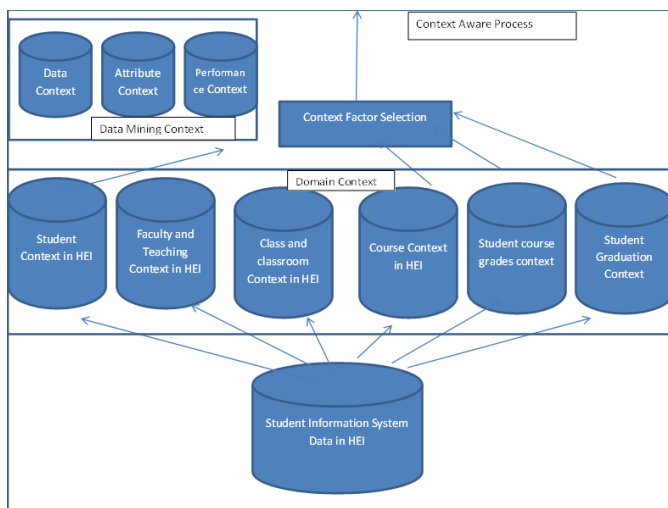

Fig 2: Student Performance Context-Aware Data Mining Framework for HEIs

The types of context factors for HEIs are defined as follows:

Domain Context illustrates domain precise context. In the case of HEIs below are some of the contexts:

Student Context: This contains all information about the student covered by the Business understanding or problem, their background, prior education, family data, income level and the like. Course Context: This context contains all information about the course similar to description of course, credits, weight-age, type of course, importance of course, timings, size of class, location of class. Faculty and Teaching Context: This contains all details about the teacher, environment, experience, education and all types of techniques or methods of teaching. Student Transcript Context: Has all details about the student transcript like grades of students of the courses enrolled in semesters, GPA scored in semester, passed credits scored in semester. Student Graduation Context: Has all details about the graduation of student, concluding GPA, destination ,time to degree. Data Mining Context defines the characteristics related to the data-mining task. This context has data context, attribute context and performance context. In Data Context the dataset to be used for mining process is defined and Attribute Context defines the attributes on which the prediction has to be made and the attributes to be used for the prediction. Performance Context actually states the time consumed for data mining process. There is always a dilemma in data mining whether speed is important or accuracy. Certain data mining applications demand speed while certain where the prediction values decides the future (HEI) or life (Medical or health) demands accuracy. Below is a pseudo code of an algorithm Context Check in Data Sets (CCDS) which checks whether a dataset has contextual attributes in it.

V.   EVALUATION

The data was collected from Ahlia university, Bahrain student information system data. Graduated students data of a particular programme for instance BSAF was collected. Initial data set included 337 student records. The data collected were pertaining to graduated students, their grades, courses registered, course details. Some fields were readily available some were computed like GPA, time to degree(length of study), course taking pattern codes. 2 datasets were fed into the algorithm. Dataset1 contained student id, GPA, time to degree, course code and semester.

Dataset2 contained student id,GPA,Time to degree, course code, semester, course difficulty, course complexity, student potential, course weight, course type, course level, class size, student prior learning details. The algorithm showed DataSet2 to have rank 1 having more contextual information then DataSet1. Genetic Algorithm was executed on both the datasets to find the course taking patterns of students in HEIs and to predict the course taking patterns for a given GPA or time to degree and vice versa.

Fig 3: Course taking patterns from Dataset 1

Fig 4: Course taking patterns predicting GPA and Length from Dataset 1

From the results got in Fig 3 and Fig 4 it is clear that the recommendation and prediction cannot be fully accurate as the data does not contain contextual information about the students and courses.

Fig 5: Course taking patterns from Dataset 2

Fig 6: Using context variables predicting Course taking patterns, GPA and Length from Dataset 2

A. Discussion

The results of the above tests represents that significant difference existed between the results produced by Dataset1 without context and Dataset 2 with context. The algorithm was able to detect the dataset with context as rank1 so as to make the dataminer use it rather than the dataset without context. The results confirm that the predictions are more effective and efficient than those produced by the dataset with no context. This step of context aware process has brought out more hidden meaningful knowledge which can be utilised for decision making in terms of resource allocation for HEIs and for students optimum course taking patterns which can lead to optimum time to degree and high GPA. Such patterns will guide us as we examine prerequisites and placement policies. Selection of courses to register, and in what combinations, can intensely impact student's likelihood for academic success, advancement towards programme degree and retention within the field and the HEI. Students are helped by academic advisers to select courses that are likely to lead them towards success in their selected programme. Earlier, academic advising has been mainly based on adviser's intuition, previous experience and institutional knowledge.

With the recent developments in data mining, advisors can start to conclude how a specific student may accomplish in the subsequent semester, by associating the data from earlier students who were in similar condition.

VI. 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. 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 weightage and course complexity which promise to enable decision makers to make more accurate decisions.

REFERENCES

1. 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).

2. 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.

3. 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.

4. 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.

5. 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.

6. 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).

7. 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.