Introduction to Global Educational Database

Soumita Banerjee and Christian Bach

University of Bridgeport
sobanerj@my.bridgeport.edu
cbach@bridgeport.edu

Abstract

Educational Data Mining is one of the major on-going research platforms now. Students' records need to be maintained and analyzed in a manner so that it can be utilized to predict students' behavior and learning methods. Although we know students' records need to be processed and analyzed; the primary challenge is to gather individual academic student details. This paper proposes a global database of students irrespective of geographical boundaries. Academic performance of every student from every country will be updated in this platform. Students' performance on major examinations will be available in the database. Supporting documents and performance details will be readily available and accessible to the evaluators from any geographic location. This will be helpful to standardize the evaluation process and analyze the performance of a student, irrespective of geographic boundaries. The following paper will discuss the available EDM tools and how data can be analyzed to extract information.

1. Introduction to Global Educational Database

In today's world, data is being generated constantly. To extract the information from a set of data, we need to collect and process the relevant data. It is difficult to collect and manage this huge amount of data. Therefore, technological tools are constantly being developed to process and analyze the data to obtain some meaningful information out of it.

Recently, one of the major research areas is the implementation of technical tools and systems in the academic field. Educational data is being recorded for a long time. To enhance the overall educational system is a major topic for research since decades. However, during operational stages, many ideas may not reach the expectation. In many countries, educational data might not be stored in electronic format and utilized for analysis purpose. With the advent of the internet, it is now easy to capture the educational documents in electronic format and send it to anywhere in the world within few minutes, if required. Therefore, the increased data-based academic information should be carefully analyzed.

Many students, from different parts of the world, are inclined to quality education in overseas. By using the internet, application procedure, query management, information sharing and information gathering are easy and quick. However, many students face issues to send the hard copy of the documents on time and thereby the admission process is getting delayed. Therefore, the cost of processing educational documents is raising with a tremendous growth rate. Not only in universities, for employment opportunities, medical checkup, and identification purpose educational documents are being considered. Therefore, there is a need to remove this obstacle and prepare an alternative option for smooth processing.

The education system varies from country to country, even state to state. However, if we take a closer look, any education system is governed by some educational boards. The entire educational system is divided among several major examinations. While evaluating

a student's academic performance, the academic results of the specific major examinations are being considered. Therefore, it is important to find out the similarity among the diversified educational systems before we proceed and its utilization to reduce the social inequalities. For instance, there are some major board examinations in many countries which are governed by an educational board or council. The results are being published by the board. For example, Wester African Examinations Council (WAEC) conducts West African Senior School Certificate Examination (WASSCE). Similarly, Central Board of Secondary Education of India supervises 10th and 12th Standard board examinations. There are also some major international educational organizations. For instance, the Educational Testing Services (ETL) manages Scholastic Assessment Test (SAT), Test of English as Foreign Language (TOEFL), Graduate Record Examination (GRE) – General & Subject and TOEIC (Test of English as International Communication) in more than 180 countries. The British Council organizes International English Language Testing System (IELTS) as a standardized test to measure English language ability. Pearson Test of English (PTE) is another standardized test, organized by Pearson. These examinations are standard examinations which are similar to all the participants. Consolidated records of students' performance are being managed by the respective test taking organization. Therefore, the data of the students' performance is being stored in a database but not available to others for further utilization. This creates a need for a global database where educational data of individual student can be available for further utilization in a secured manner.

2. The Idea Generation

Scientists and researchers are continuously trying to provide the best possible solutions to our problems. Once an issue is resolved, it initiates another situation which might need attention. For instance, before the internet arrives, managing paper documents was a huge task. With the arrival of the internet, electronic copies of documents replace the paper documents in many cases. However, managing the electronic data and information is a challenging task now. We are again in need of a new mechanism to store this huge data properly and utilize it for our own benefit.

The objective of the product development process is to generate an idea to provide an alternative approaches or solution. In the Product Development process, the brainstorming stage helps to identify the problem and idea generation is the following stage, where we can come up with possible solutions. In this situation, the Global Educational Database can be a new possible solution to the problem in managing diverse test scores.
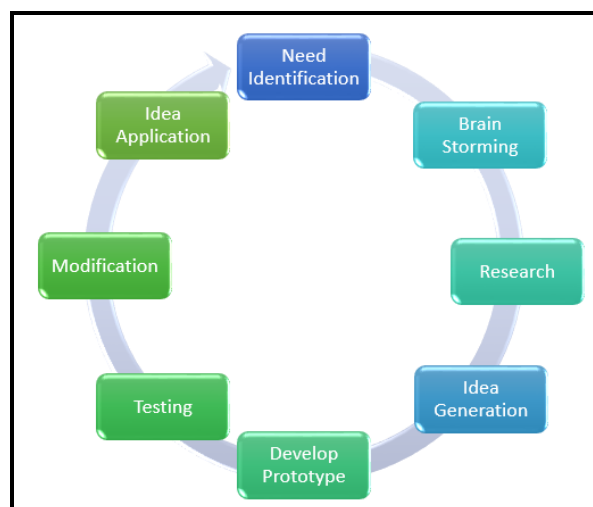

Figure 1. Product Development Process

2.1 Problem Identification

The problem identification process requires to identify all the problems and evaluate their relationships before proposing a solution. Data is being generated always. However, only structured data is mainly useful. The structured data can be stored and analyzed so that data can be converted into information. In academic domain, educational records are mostly stored based on student details. However, they are not structured in a database so that it can be accessible by others for further analysis.
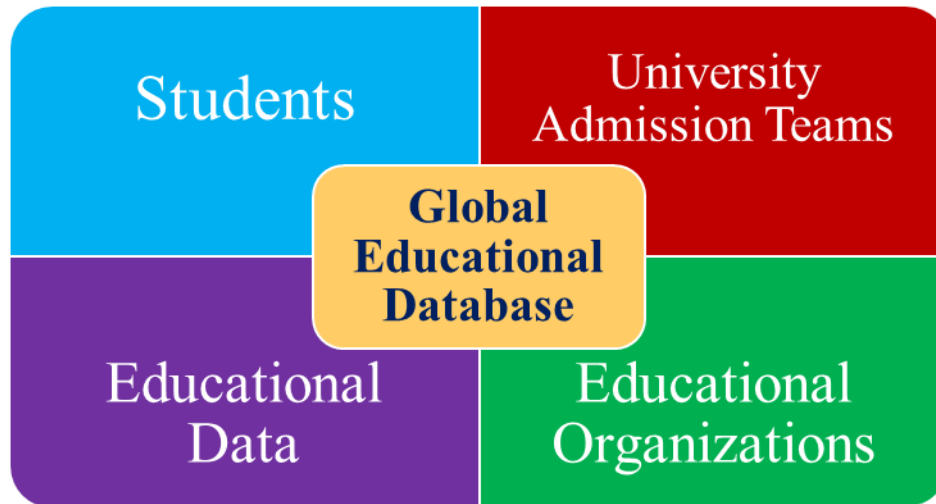

Figure 2. The Relationship Matrix between Dependent Variable and Independent Variables

When a student applies for higher education in different universities, she has to provide the hardcopy of the educational records to all the places where she applied. This is a time-consuming procedure. Also, a lot of paperwork is also an issue. Therefore, a global database, for all the students from different countries, which can store their academic results electronically is useful to save time and paper both. As per the above model, Global Educational Database is the Dependent Variable here. The Independent Variables are Students, Educational Data, Educational Organizations and University Admission Teams.

2.2 Proposed Solution

A database contains the information which user can relate with the available data. A database is a collection of data, records, and facts. In a database, data is available to store and retrieve as and when needed. An efficient database contains structured and normalized data. Only useful, relevant and correct data should be stored in a database. Data is stored in different tables in a database. These tables are interconnected and related to each other.

In this paper, we propose a global database which can be accessed from different parts of the world. The Global Educational Database (GEDB) is an academic database where academic results of students are stored directly by the test organizers. The students can be from different geographic locations. Every student will have a country specific unique student ID. The data will be stored based on the student ID so that it can be retrieved easily by using the same.
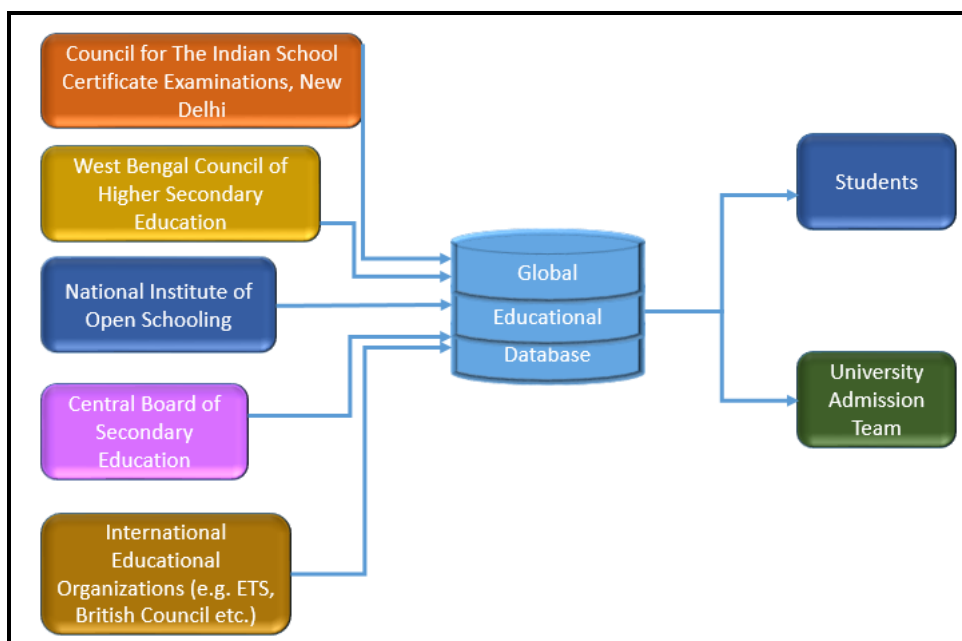
Figure 3. The Workflow of the GEDB

As overseas education becoming popular now, many students travel from one country to another for better education. Therefore, when a student from one country applies for higher education in another country, it is time-consuming to send all the hard copies of scorecards. Therefore, in Global Educational Database, University admission team can also view students' academic data by using Unique Student ID (USID). University team can view and download individual or a group of students' historical academic details. Although, they cannot edit or modify any data, the academic details can be helpful while making admission decisions.

Furthermore, there can be a question of data security and data manipulation. In order to minimize the error while inserting and editing the data, the academic data will be entered into the Global Educational Database by the test organizers only. For example, there are different state-based educational boards in India and there are several central educational boards. Additionally, there are some international educational organizations which organize different tests. The academic performance of a student will be entered by these test organizers. In a case of any mismatch, a student can contact the educational board to correct the data. When a student applies to a university, she can give her Unique Student ID to the university to access and view her academic performance and retrieve the information from this database.

The student ID can be as unique as a Passport number to an individual. Therefore, student ID can consist of country code, serial number, initials of the first name and last name, last four digits of Aadhar card number or Social Security Number (in the case of the United States of America). The USID will be system generated. No two students can have same student ID as this will be separated by country code and government identification number.

The test taker organizations upload the student scores according to USID. The data is inserted in a structured way. In the database, data is stored in different tables which are related. Therefore, data is normalized while it is stored in the database. In the database, all the tables are connected to form a Relational Database Management System (RDBMS). As mentioned by researchers, in an RDBMS, data is stored as tables, columns, and index. The relationship of the tables helps users to find the information. When a query is given by the user, the system refers to the system directory and display the information by identifying the relationship.
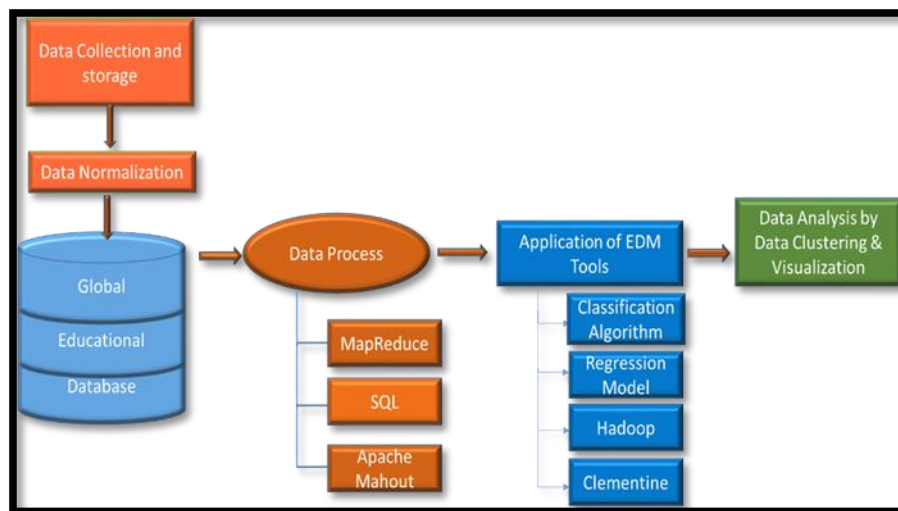
Figure 4. The Data Processing and Application of EDM tools for Data Analysis

After the academic data is being stored in the GEDB, the data is processed before analysis. The data can be processed by different programming models such as MapReduce, Hadoop, Apache Mahout etc. After data processing, the relevant data will be identified to apply different Educational Data Mining tools and algorithms for analysis. For example, classification algorithms, regression models, MS SQL, Clementine, Keel, Sodas etc. education data mining tools can be used to analyze the data by data clustering and visualization. Thereafter, the recommender systems can be applied to predict students' performance and behavior.

2.3 Benefits

The primary advantage of the Global Educational Database is student academic records is stored securely in electronic format irrespective of the geographic boundary. Every student will have a unique ID so that academic records can be retrieved using the specific ID. Moreover, the academic data is available directly from the authentic sources such as the respective educational board and international organizations of test takers. Therefore, the chances of incorrect data insertion are less. In the case of any availability of wrong data, the respective student can request to update the data by providing valid proof. Only an authorized person from the governing organization will be able to rectify the errors.
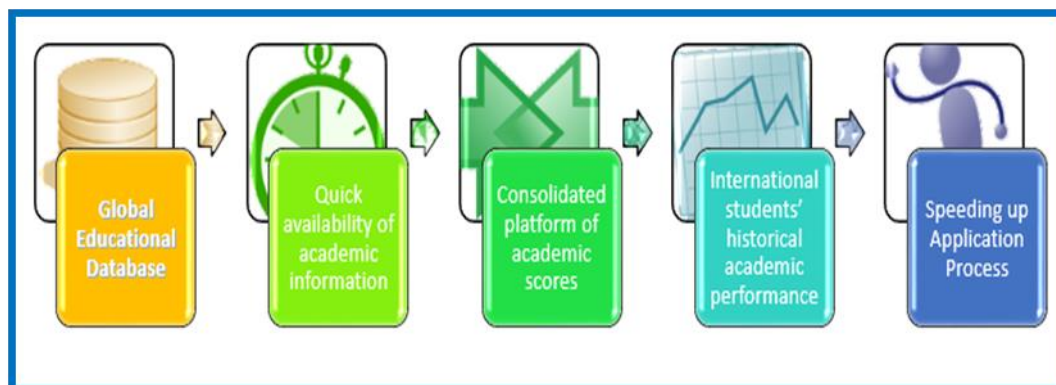

Figure 5. The Benefits of GEDB

By applying Educational Data Mining tools and algorithms, the academic data can be analyzed to predict student's performance and behavior. The easy availability of data will reduce the time consumption to collect and evaluate the data. The academic history of students can be easily available in the database so that it can be compared with other students from other parts of the world.

2.4 Challenges

At first, the Global Educational Database must be updated and maintained by skilled individuals. The academic records of millions of students must be updated correctly in the database. In a case of any mistake, the students will not be able to correct. Therefore, it is an additional responsibility of the test organizers to verify the academic record and update the correct educational details on time so that it will not have any negative impact on the student's performance prediction. On the other hand, the students will be dependent on the authorized person to rectify if there is any error in his academic records. This process can be time-consuming if there is a lack of skilled people. Therefore, the accurate data should be available from the source. It should be verified before updating it in the Global Educational Database. Additionally, the data analysis should be done by trained resources. The incorrect application of the Educational Data Mining tools and software will result in incorrect information and analysis. The data should be analyzed perfectly to predict the students' performance and behavior in best possible manner.

3. Research Methodology

The purpose of a research is to recommend and describe practical and theoretical advances. Research methods can be of different types such as experiments, surveys, content analyses, phenomenological methods, focus groups, exploratory, qualitative, evaluation research and more. Research shows, knowledge discovery in a database is the procedure to identify patterns in data and data mining is revealing patterns from the data [1]. Personalized web-based learning system offers individualistic academic performance details. However, a global database is useful to measure academic performance as a whole.

The literature review has created a theoretical foundation of this paper. Without a theoretical foundation, the interpretation of models is difficult. This new theoretical insight advances our understanding of the importance of a structured academic data [2]. Therefore, a combination of different research designs and methodologies are used for this study. One of the primary research methods is the Grounded theory, where concepts are generated by studying behavioral patterns. In this paper, practical data was considered for need identification. For instance, pending student files from different parts of the world were being studied. These files are being held due to incomplete academic documentation. The complete academic documentation might not be received by the deadline or the equivalency might not be understood. Ethnography is the another research method which is being applied here. Ethnography is about studying socio-cultural differences [3]. By observing the work culture in the admissions department, understanding the problems to meet deadlines and studying students queries about complete documentations this study is being conducted to identify the most effective and efficient way. Moreover, a collection of different journals and articles are studied which are available at the University of Bridgeport digital library, JSTOR, Google Scholar, Science Direct, and ProQuest. The literature review is essential for advancing our current knowledge. The literature review-centric approach follows three phases: 1) Brainstorming, 2) Narrowing down and 3) Evaluation. Next, the Action Research method is being followed by creating a sample database.
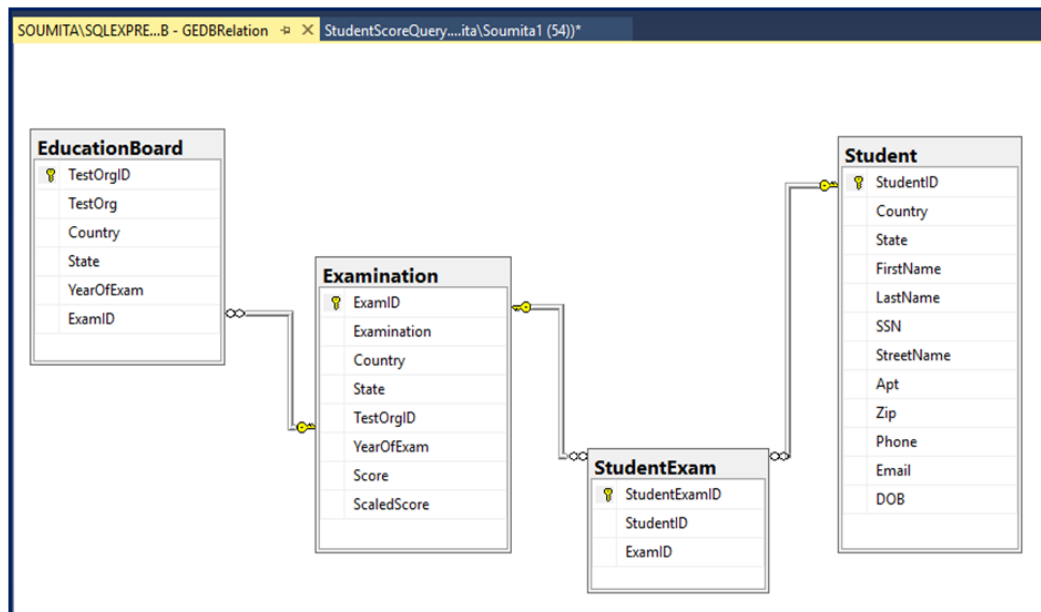
Figure 6. The Relationship between Different Tables in Database (using SQL Server)

Only student details of USA are being considered in the sample database. The database identifies the many-to-many relationship between Student Tables, Education Board table, and Examination table. StudentExam Table is the primary table for the many-to-many relationship.
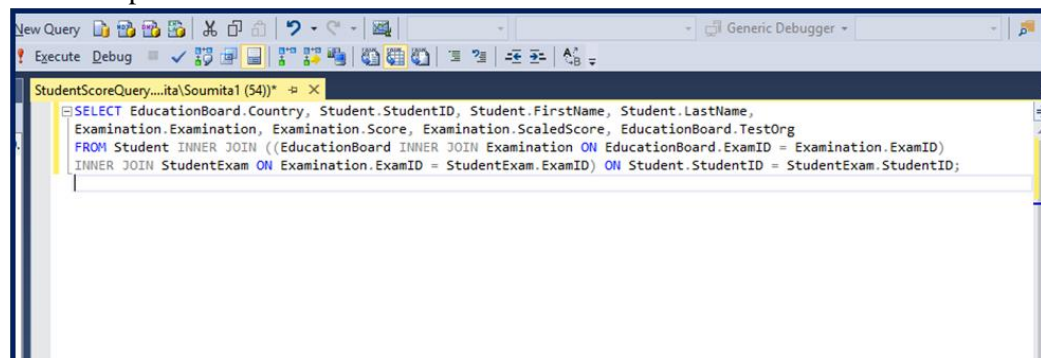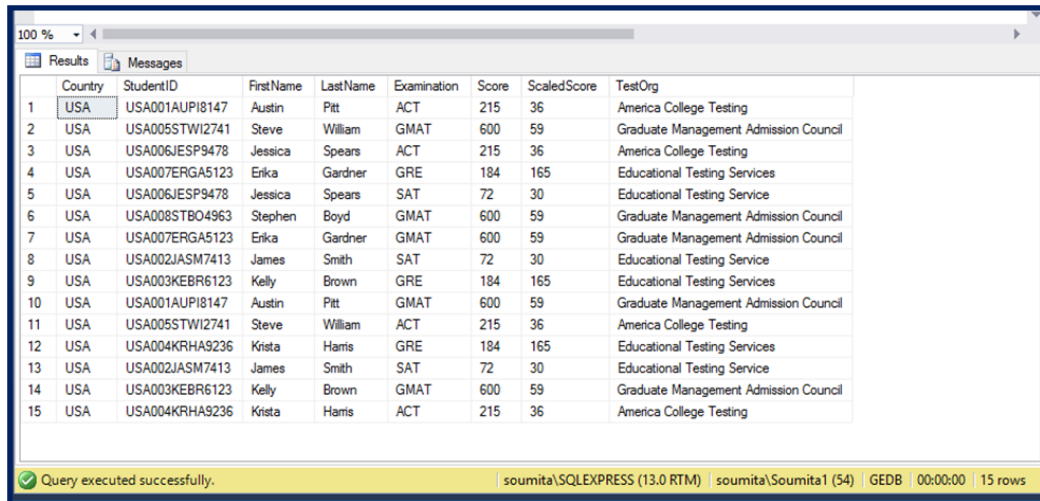

Figure 7. The SQL Query Statement to Extract the Data from the Database

After the relationship is created, the sample database is ready to extract the data from the database. SQL is the primary tool to access the database and extract the required data.
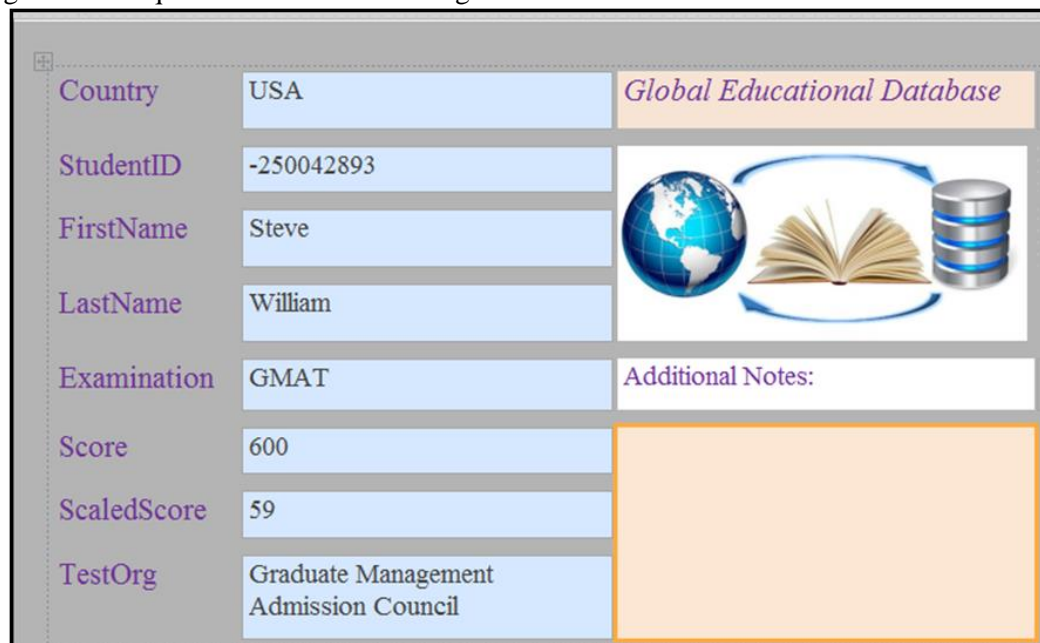
	Country	StudentID	FirstName	LastName	Examination	Score	ScaledScore	TestOrg
1	USA	USA001AUPI8147	Austin	Pitt	ACT	215	36	America College Testing
2	USA	USA005STWI2741	Steve	William	GMAT	600	59	Graduate Management Admission Council
3	USA	USA006JESP9478	Jessica	Spears	ACT	215	36	America College Testing
4	USA	USA007ERGA5123	Erika	Gardner	GRE	184	165	Educational Testing Services
5	USA	USA006JESP9478	Jessica	Spears	SAT	72	30	Educational Testing Service
6	USA	USA008STBO4963	Stephen	Boyd	GMAT	600	59	Graduate Management Admission Council
7	USA	USA007ERGA5123	Erika	Gardner	GMAT	600	59	Graduate Management Admission Council
8	USA	USA002JASM7413	James	Smith	SAT	72	30	Educational Testing Service
9	USA	USA003KEBR6123	Kelly	Brown	GRE	184	165	Educational Testing Services
10	USA	USA001AUPI8147	Austin	Pitt	GMAT	600	59	Graduate Management Admission Council
11	USA	USA005STWI2741	Steve	William	ACT	215	36	America College Testing
12	USA	USA004KRHA9236	Krista	Harris	GRE	184	165	Educational Testing Service
13	USA	USA002JASM7413	James	Smith	SAT	72	30	Educational Testing Service
14	USA	USA003KEBR6123	Kelly	Brown	GMAT	600	59	Graduate Management Admission Council
15	USA	USA004KRHA9236	Krista	Harris	ACT	215	36	America College Testing

Query executed successfully. | soumita\SQLEXPRESS (13.0 RTM) | soumita\Soumita1 (54) | GEDB | 00:00:00 | 15 rows

Figure 8. Sample Data Extraction about Student Details, Examination and Score Details

The execution of the query extracts the required data from the available data and reflects the information such as unique Student ID, Country, First Name, Last Name, Test Organization, Examination, Score, and Scaled Score. The Student ID should be auto generated. The specifications can be manually designed. For example, Student ID can be a mix of abbreviated name of the country, digits, first two initials of Last Name and First Name, last 4 digits of SSN i.e. USA001AUPI8147. The Student ID should be unique for every student irrespective of country. This will be helpful to identify only one student from any country. Therefore, the Social Security Number can be replaced by other country based identification such as Passport number, Birth Certificate number, Permanent Account Number, Aadhar Number etc. The database will be modified to generate unique Student ID considering these identification numbers.

Figure 9. Sample Student Data Form

A country-specific, sample, end-user data entry form is created in the above picture. Student ID, Country are the minimum requirement to extract the data. Data can be viewed easily in this form by the end-user. This form includes the Global Educational Database logo and an optional field to include any additional notes.

4. Data Sources

Every year many students move to different countries for better education in undergraduate level, master's level and in doctorate level. According to the report [4], international students from more than 42 countries came to study in the United States of America in 2016. In the picture below, the percentages of the international students, who came from around 32 different countries, enrolled in the Bachelor's studies in five specific countries such as the United States of America, Canada, the United Arab Emirates, China, and India are mentioned in the below table.

Dataset: Share of international students enrolled by country of origin

Year	2013	
Indicator	Share of mobile by country of origin	
Level of education	Bachelor's or equivalent level (ISCED2011 level 6)	
Unit	Percentage	

Country	Australia	Austria	Belgium	Canada	Chile	Czech Republic	Denmark	Estonia	Finland	France	Germany	Hungary	Ireland	Israel	Italy	Korea	Latvia	Luxembourg	Netherlands	New Zealand	Norway	Poland	Portugal	Slovak Republic	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	United States	Brazil
Canada	1.3	0.2	0.1	0	0.2	0	0.4	0.3	0.3	0.7	0.1	0.2	12	1.9	0.1	0.8	0.1	0.1	0.2	0.9	0.4	2	0.3	0.1	0.1	0.2	0.3	0.5	0.1	1.4	3.3	0.4
United States	0.9	0.8	0.2	6.5	1.2	0.2	1.6	0.8	1.3	1.6	0.8	2.8	9.2	14.5	0.3	1.6	0.9	0.2	0.4	7.9	1.9	3.7	1.6	0.2	0.2	2	1.6	1	0.7	2.1	0	4
United Arab Emirates	0.5	0	0	0.2	0	0	0	0	0.1	0.1	0	1.2	0	0	0.1	0	0.1	0	0.1	0.1	0	0	0	0	0.1	0	0.1	0	0.8	0.3	0	0
China	34.3	0.9	0.2	34.4	0	0.3	2.6	2.5	8.5	11.1	8	2.6	10.5	0.4	14.1	80	0.3	0.1	5.6	33.3	4.3	1.9	0.7	0.2	0.2	3.7	9	1.1	0.8	16.7	28.7	2.3
India	3.4	0.1	0.1	0.6	0.4	0	0.4	0	0.4	0.2	0.3	0.4	1.1	0.3	0.7	0.1	2.6	0.2	0.2	11.2	0.7	0.7	0	0	0.2	0	0.5	0	1	3	11.8	0.1

Figure 10. The Percentage of the International Students of Different Countries who Come to 5 Listed Countries for Bachelors' Studies

Based on the above data, in the below chart, we can see students from many different countries come to the United States of America for studying Bachelor's program. The educational systems are also different in different countries. However, when they are applying for Bachelor's program in the USA, their academic documents, score cards are also required to the universities where they have applied. It is difficult to receive academic documents from different countries within a specific time limit. Also, the educational system of a specific country needs to be understood before analyzing it. The Global Educational Database can be helpful in this situation. The educational records are easily available and accessible to every international student. The performance of the student in major academic examinations can be viewed and analyzed as well. Finally, by applying recommender systems, student performance and behavior can be predicted. Therefore, it will be easy to make admission decision by the Universities.
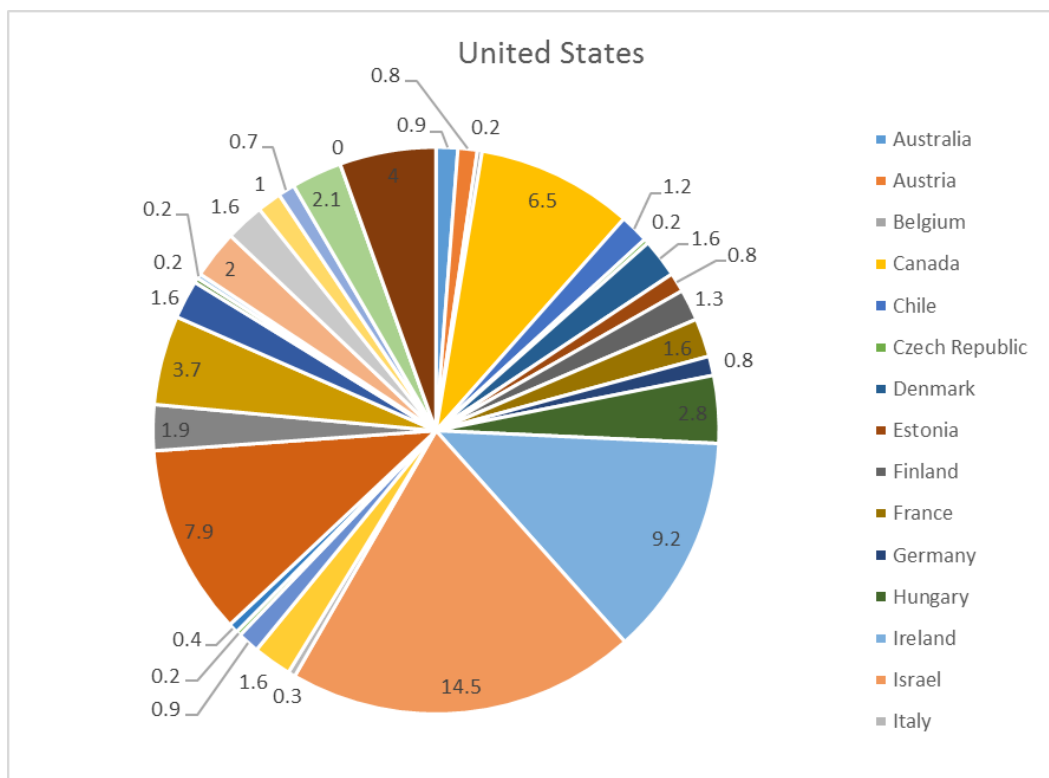
Figure 11. Percentage of the Incoming International Students for Bachelor Studies in 2013 from Listed Countries in the USA

5. Related Study

The primary objective of any educational institution is to provide best possible education to the students. The teaching pattern can be modified based on the past performance of the students. Research shows that the historical educational data can be analyzed by decision tree methods such as Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) [5]. The student database can be utilized to understand students' historical performance and predict the future performance. The enormous educational data of any academic institution can also be evaluated by applying clustering algorithms such as Course Classification Algorithm, Particle Swarm Optimization (PSO), and Expectation-Maximization (EM) [6]. After carefully analyzing the data, the recommender system can be helpful to predict students' performance. According to Bhatia and Prasad [7], Apache Mahout and Hadoop can be implemented for distributed processing of big data and to enhance the recommendation system. Not only students' assessment, evaluation of lecturer, course modules, industrial training and student registration are also important. The DM_EDU model is the data mining technology that can be implemented for thorough analysis. The Student Success System (S3) is another predictive model which specify at-risk students, areas of improvement needed, suggestions for improvement and feedback system [8].

A combination of prediction model and social networking site is another possible predictor of students' performance. According to Halawa, Shehab [9], students personality types can be identified using Myers-Briggs Type Indicator (MBTI) and students behavior can be monitored in Facebook. The results can be analyzed using the MOODLE, a learning management system. Statistical tools and mathematical models can also be applied to predict students' performance even before the semester begins [10]. The Association Mining techniques help us to identify the relationship between the student's interest for a project work, the faculty interest and the applicability of the project work

[11]. Moreover, open source software such as Moodle, e/m learning and learning through play models can be used to collect and analyze data [12]. Mathematical models can also predict student retention rate and their academic performance.

A database is a collection of data and information from where an analyst can get a conclusion. At the present time, computerization of data is necessary for everywhere. For instance, a Global Medical Knowledge database can help to examine a patient's case and get a reference immediately. A global database of global biodiversity will be helpful to analyze the biological diversity process and its effects on mankind. A global database can also be useful to students. They can learn how to manage database and extract information from the same [13].

The concept of a global database with the country-specific access is supported by the study of several researchers. According to this study, a global geographical database captures the geographical data in a Geographical Information System (GIS). This database can be accessed by different countries to study country specific detailed information. Therefore, it will be easy to monitor, collect and examine geographical information. Research shows, student information such as GPA, ACT score, the number of courses taken in high school and number courses completed can be analyzed by logistic regression model to predict their performance in college [14]. Therefore, international students' academic information can also be captured in a global database and can be analyzed using prediction models in anywhere in the world.

In 2002, Japanese scientists created a new database which contains information about human complementary DNAs. This database is useful storage of information about DNA. This open access database is available for research studies. The Electronic Database for Global Education (EDGE) was an effort to establish a global educational database. However, the success of this project is unknown. As Göker, Bülbül [15] mentioned that a database, which contains students' demographic, personal, and academic score information, is a warning system to predict students' academic success and any barriers to the same. Capturing valuable information in a structured, normalized method is important to maximize the benefits of the student database.

As we know, data should be analyzed to extract information from it. Therefore, in any large database, data can be of any format such as image file, text file, CSV etc. These different types of data should be analyzed by a tool such as Cross Industry Standard Process for Data Mining (CRISP-DM) so that information can be obtained [16]. Not only educational data, social networking data are also being considered for analyzing student behavior. Social Network Analysis (SNA) analyze the interactions of an individual with others [17]. This can be considered to predict student behavior. Research shows that the logistic regression analysis identifies the behavior patterns of the distant learning students. This can help to formulate retention strategies and modify the learning system according to the students' behavior.

According to Sperry [18], the logistic regression model can predict students' behavior and retention factors irrespective of learning community. Research shows, the predictive models are also being used in different organizations [19]. Many large organizations create their own talent pool by attracting skilled resource, nurturing and developing their skills. The repository of high potential and highly skilled employees are being analyzed to predict their behavior.

Students' performance is standardized test scored and their response to the non-cognitive questionnaire can be analyzed to measure the grade point average. Special clustering can be applied to identify the matches between data points [20]. Therefore, academic performance comparison can be easier irrespective the country-specific educational system barriers. Fuzzy systems neural networks and fuzzy systems generic algorithms can also helpful to manage the diversity of students' performances and to predict their behaviors.

6. Results & Findings

The concept of GEDB was being analyzed by the survey of the potential beneficiaries. The responses were captured from the randomly selected members of the international admission team and students of the University of Bridgeport. The survey questionnaires for admission team members and students were different.
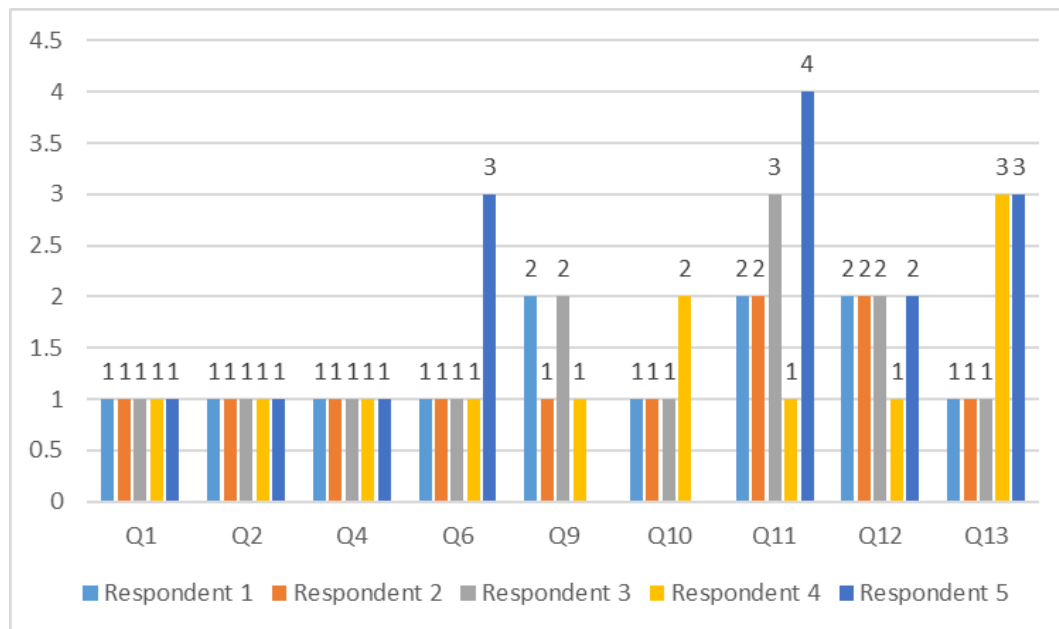

Figure 12. Analysis of Survey Responses from International Admission Team about the Feasibility and Need of GEDB

In the above picture, the analysis of the five respondents from admission team is shown graphically. Out of five members, four members accepted the need of the GEDB and confirmed that it will be helpful for them to process applications. All the admission team members accepted that GEDB will be helpful to reduce the application processing time as international students' details are quickly available and accessible. The majority of respondents think historical academic records will be helpful to make admission decision quickly. Half of the total number of respondents think GEDB will also be helpful to verify student's scores and they will recommend it to others. Most of the team members are experienced more than five years in international admission department. Out of five, two respondents are concerned about learning the process to extract student records and the time consumption to become an expert. As the responses are quantified, the mode of the responses is 1 which is positive about this concept. One respondent concerned about the cost efficiency and acceptability of the project in the institutions. Overall, the responses were good and spontaneous.

Table 1. Identified Concerns and Proposed Solutions

Cost Efficiency	The cost of implementing the project in reality
	Cost can be minimized by using web-based data management platform
Authority of data management	The authority of data management can be manipulated
	Each academic education will be provided secured id to upload data and access students' data
Unanimous Participation	Bringing all the academic boards and universities together is challenging
	By the help of ministry of education, all the academic institutions and universities will be participants
Historical Data Availability	The academic details of old students will not be available
	The academic history of students from specific year will be available as the pilot project
Legal obligation	Student might not want to show his poor performance in a specific course and legally it is not allowed to view someone's details
	Scores of major examinations will be available which are relevant and necessary to predict student behavior and performance

Figure 13. Areas of Concerns Highlighted by International Admission Team and the Proposed Solutions to Minimize the Concerns

The above figure shows the possible concerns raised by the International Admission Team members based on their practical experiences. The proposed solutions for each concern is also mentioned in the table. The unanimous acceptance of the project is needed for long term benefits. To foster the electronic data management and proper academic score maintenance, GEDB will be a major platform.
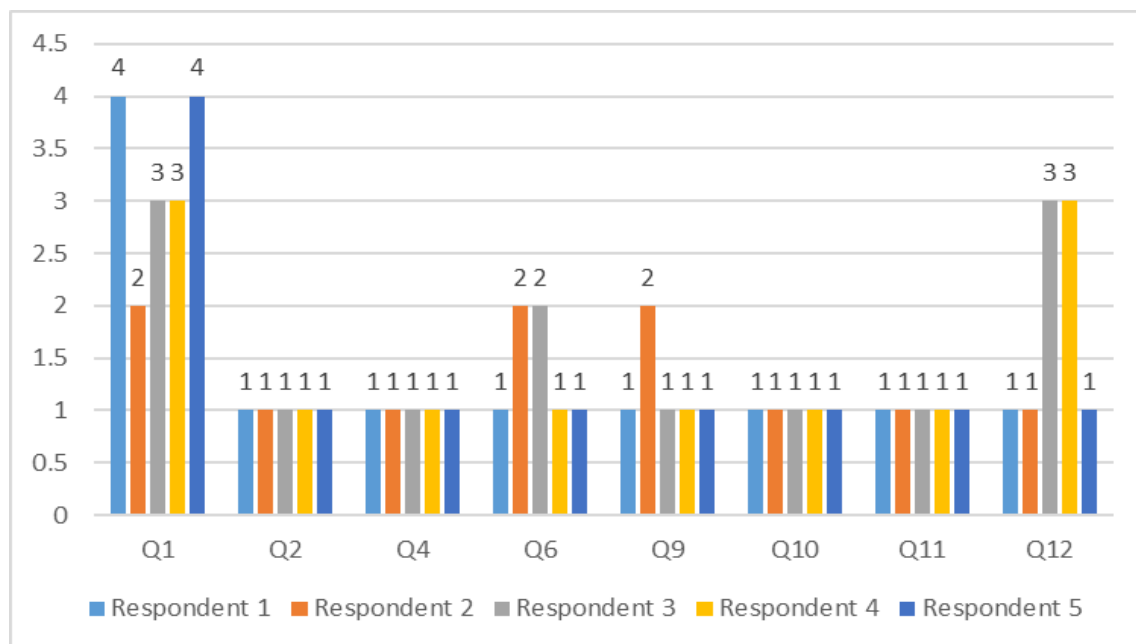
Figure 14. Analysis of Survey Responses from International Students about the Feasibility and Need of GEDB

In the above picture, the analysis of the five respondents as international students are shown graphically. All the international students have accepted the need for a global platform to maintain academic details. During this survey, all the students have accepted the lengthy procedure to receive an official transcript from their universities. The GEDB can be an alternative to the paper-based transcript as the authentic data is available to the admission teams of anywhere in the world. The convenience of sharing the academic performance in a chronological manner and the cost efficiency in processing the application documents are also widely accepted.

7. Future Direction

The Global Educational Database is a collective mechanism of structured, relevant, normalized academic and demographic student data. The data can be uploaded and retrieved by authorized people. In future, the application of different algorithms and models will result in the easy availability of student data. Recommender systems will be applicable to recommend student's performance and behavior. Accordingly, educational systems and courses can be restructured. To provide most accurate analysis, data should be collected in a structured manner from beginning.

The academic data stored in GEDB also helpful for the verification of equivalency. In future, the structured academic data is useful for comparative analysis irrespective of country-specific educational boards. The academic standards and patterns of one country are different from another country. Therefore, the equivalency mechanism is useful to compare students' data in a common platform.

Moreover, the academic history can be also used for verification by employers. To ensure authentication, employers verify the academic documents by the respective educational organizations. As academic details are directly uploaded and maintained by the test takers, the verification of data is assured. Therefore, a secured data storage tool like GEDB reduces the time to verify the documents.

GEDB can also be utilized like an Enterprise Resource Planning (ERP) which allows different educational organizations to use an integrated system to maximize the exploitation of the academic history of students irrespective of location barriers. Using

technology, innovation, skills and secured authority, GEDB can be accepted as a bane to many international applicants and admission team members.

8. Conclusion

This paper has been an effort in advancing the current educational process by data normalization and data clustering technologies using Information Technology. The research study is conducted in order to optimize the benefits of the educational system by applying Information Technology. This study reveals the high demand of a systematic educational data repository throughout the world. The model proposed by the author targets to provide superior advantages in Educational Data Mining in higher educational systems. This model is tested with hypothetical students' educational data to identify its potential benefits and future works. With the rising demand for overseas study, this database is a time saving and cost efficient method. Moreover, students' historical academic performance is helpful to plan and monitor their further educational goals. Identifying academically strong and weak students is helpful to restructure the course curriculum. Additionally, the data availability for data analysis is a great advantage. The application of different prediction models, algorithms, and recommender systems makes it easy and quick to analyze and predict student's future performance and the improvement process. Therefore, this study plays a key role in improving the overall effectiveness of the educational system globally.

As a further work, this model will be implemented to extract desired students' details and to provide adequate information to the University Admission team members and to distinguished faculty members so that selection process of students becomes more systematic.

Acknowledgments

Appendix A

This survey is a part of an academic project to find out how GEDB can be beneficial to the admission department. Thank you for your time and sharing your honest feedback. This academic project aims to provide support to international students and the admission departments of different universities by collecting authenticate, chronological academic records of the students of different countries, making data availability quick and timely, predicting students' performance and analyzing students' behavior.
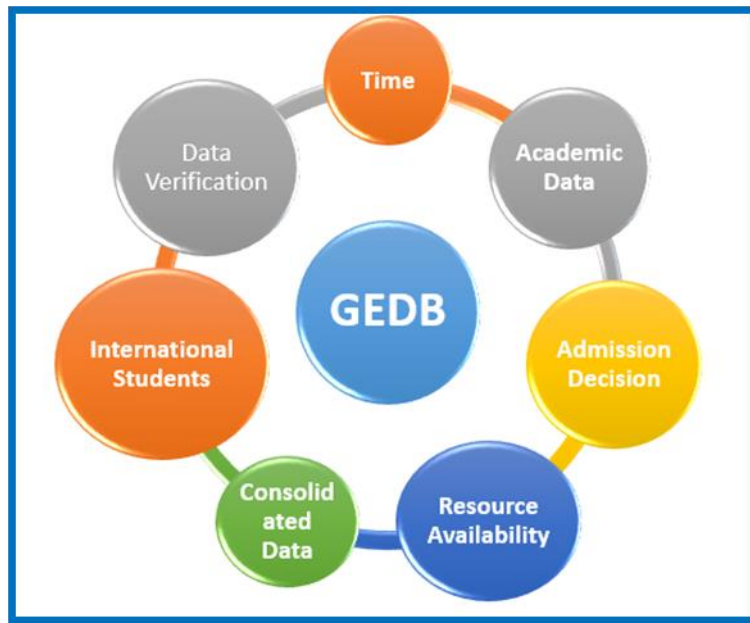
Figure 15. The factors of GEDB

1.   How long does it take to receive all academic documents?
1.   Less than 3 weeks
2.   3-5 weeks
3.   6-8 weeks
4.   More than 8 weeks

2.   Do you think a Global Educational Database will save time?
1.   Yes
2.   No
3.   May be

3.   What do you like most about GEDB?

4.   Do you think this product help international students to process their application quickly?
1.   Yes
2.   No
3.   Not sure

5.   What changes would most improve GEDB?

6.   Do you think the historical academic records of the student helpful to make admission decision?

1.   Very likely
2.   Moderately likely
3.   Slightly likely
4.   Not at all likely

7.   If you do not want to use GEDB, what are your concerns?

8.   How convenient GEDB is to verify academic score of a student?

1.   Very easy
2.   Manageable
3.   Very difficult
4.   Not at all possible

9.   If GEDB is available today, how likely would you like to recommend it to others

1.   Very likely
2.   Moderately likely
3.   Slightly likely
4.   Not at all likely

10.   How long are you working in Admission Department?

1.   Under 1 year
2.   1-2 years
3.   2-5 years
4.   Over 5 years

11. **Applications from how many countries you manage?**
1. 1-5
2. Over 5

12. **Do you think this project will be less time consuming?**
1. Yes
2. No
3. May be

14. **If you have any suggestions about this project, please provide your remarks below:**

Appendix B

This survey is a part of an academic project to find out how GEDB can be beneficial to the admission department. Thank you for your time and sharing your honest feedback. This academic project aims to provide support to international students and the admission departments of different universities by collecting authenticate, chronological academic records of the students of different countries, making data availability quick and timely, predicting students' performance and analyzing students' behavior.

1. How long does it take to collect all academic transcripts?
1. Less than 3 weeks
2. 3-5 weeks
3. 6-8 weeks
4. More than 8 weeks

2. Do you think a Global Educational Database will be useful?
1. Yes
2. No
3. May be

3. What do you like most about GEDB?

4. As an international student, do you think this product help to process your application quickly?
1. Yes
2. No
3. Not sure

5. What changes would most improve GEDB?

6.   Do you think your historical academic records should be available to admission teams?

1.   Very likely
2.   Moderately likely
3.   Slightly likely
4.   Not at all likely

7.   If you do not want to use GEDB, what are your concerns?

8.   How convenient GEDB is to verify your academic score?

1.   Very easy
2.   Manageable
3.   Very difficult
4.   Not at all possible

9.   If GEDB is available today, how likely would you like to recommend it to others?

1.   Very likely
2.   Moderately likely
3.   Slightly likely
4.   Not at all likely

10.   How many countries you send your applications?

1.   1-2
2.   2-5
3.   Over 5

11.   Do you think this project will be cost efficient?

1.   Yes
2.   No
3.   May be

12.   If you have any suggestions

References

[1]   H. Guruler, A. Istanbullu, and M. Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases", Computers & Education, vol. 55, no. 1, **(2010)**, pp. 247-254.

[2]   J. A. LePine,and A. W. King, "Developing Novel Theoretical Insight From Reviews of Existing Theory And Research", The Academy of Management Review, vol. 35, no. 4, **(2010)**, pp. 506-509.

[3]   H. Wardle and P. G. y Blasco, "Ethnography and An Ethnography in the Human Conversation", Anthropologica, vol. 53, no. 1, **(2011)**, pp. 117-127.

[4]   R. Reeves, "Integrated Postsecondary Education Data System", 2013-14. [cited 2016 September 9]; Available from: https://catalog.data.gov/dataset/201314-integrated-postsecondary-education-data-system, **(2016)**.

[5]   B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance", arXiv preprint arXiv:1201.3417, **(2012)**.

[6]   A. Dutt, "Clustering algorithms applied in educational data mining", International Journal of Information and Electronics Engineering, vol. 5, no. 2, **(2015)**, pp. 112.

[7]   L. Bhatia and S. Prasad, "Building a distributed generic recommender using scalable data mining library. in Computational Intelligence & Communication Technology (CICT)", 2015 IEEE International Conference on IEEE, **(2015)**.

[8]   A. Essa and H. Ayad, "Improving student success using predictive models and data visualisations", Research in Learning Technology, vol. 20, **(2012)**.

[9]   M. S. Halawa, M. E. Shehab and E. M. R. Hamed, "Predicting student personality based on a data-driven model from student behavior on LMS and social networks", in Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on IEEE, **(2015)**.

[10] S. Huang, "Predictive modeling and analysis of student academic performance in an engineering dynamics course", **(2011)**.

[11] R. Jindal and M. D. Borah, "A Survey on Educational Data Mining and Research Trends", International Journal of Database Management Systems, vol. 5, no. 3, **(2013)**, pp. 53.

[12] P. Juric, M. Matetic, and M. Brkic, "Data mining of computer game assisted e/m-learning systems in higher education", in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on IEEE, **(2014)**.

[13] A. Vágner, "Software application for supporting the education of database systems", Acta Didactica Napocensia, vol. 8, no. 3, **(2015)**, pp. 23.

[14] M. C. Morrison, and S. Schmit, "Predicting Success in a Gateway Mathematics Course", Online Submission, **(2010)**.

[15] H. Göker, H. I. Bülbül and E. Irmak, "The Estimation of Students' Academic Success by Data Mining Methods", in Machine Learning and Applications (ICMLA), 2013 12th International Conference on IEEE, **(2013)**.

[16] M. Riedel, A. S. Memon and M. S. Memon, "High productivity data processing analytics methods with applications", in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on IEEE, **(2014)**.

[17] G. Siemens, and R. S. d Baker, "Learning analytics and educational data mining: towards communication and collaboration", in Proceedings of the 2nd international conference on learning analytics and knowledge ACM, **(2012)**.

[18] R. A. Sperry, "Predicting first-year student success in learning communities: The power of pre-college variables", Learning Communities Research and Practice, vol. 3, no. 1, **(2015)**, pp. 2.

[19] G. Stahl, "Six principles of effective global talent management", Sloan Management Review, vol. 53, no. 2, **(2012)**, pp. 25-42.

[20] S. Trivedi, "Spectral clustering in educational data mining", in Educational Data Mining vol. 2011, **(2010)**.