Realistic Analysis of Data Warehousing and Data MiningApplication in Education Domain

Dr. Akhilesh Saini[1], Ms. Priyanka R. Gondaliya[2]
[1]*Associate Professor Om Sterling Global University, Hisar*
[2]*Assistant Professor, Sardar Patel College of Engineering, Bakrol, India*

Abstract—Data-driven decision support systems, such as datawarehouses can serve the requirement of extraction of information from more than one subject area. Data warehousesstandardize the data across the organization so as to have a single view of information. Data warehouses can provide the information required by the decision makers. In the education sector, data warehousing and data mining have increasingly become essential tools for enhancing the quality of education, understanding student behaviour, improving learning outcomes, and optimizing institutional performance. This analysis outlines their practical applications, challenges, and benefits within the educational domain.

INTRODUCTION

Now a day, the educational institutes have to generate funds for their research and other operational activities as thegovernment funding has been limited to aided institutes. Utilizing a decision support system is a proactive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by distilling data from any combination of education records maintenance system. The data mining fromdata warehouse can be a ready and effective system for the decision makers. A data warehouse is a subject oriented support of management decisions .Data warehouse obtains the data from a number of operational data base systems which can be based on RDBMS/ERP package, etc. The data from these sources are converted into a form suitable for data warehouse. This process is called Extraction, Transformation and Loading (ETL). In addition to the target database, there will be another database to store the metadata, called the metadata repository. This data base contains data about data-description of source data, target data and how the source data has been modified into target data. The client software will be used to generate reports.

MOTIVATION

1. Data Warehousing in Education

A data warehouse is a centralized repository for storing large volumes of data from different sources. In education, data warehousing helps institutions consolidate student records, administrative data, financial information, and academic performance reports, creating a unified system for analysis.

Applications:

- Student Information System (SIS): Educational institutions can integrate data from multiple systems—attendance, performance, disciplinary
- integrated, non-volatile, and time variant collection of data in records, and health data—into a data warehouse for a holistic view of each student's journey.
- Decision Support: Administrators can make data-driven decisions on resource allocation, curriculum development, and policy-making by analysing historical data on student success rates, drop-out trends, and faculty performance.
- Accreditation and Reporting: Universities and schools are required to report data to accreditation bodies. A centralized data warehouse simplifies the process of aggregating and reporting necessary metrics (e.g., graduation rates, test scores).
- Predictive Maintenance and Budgeting: Facilities and resources data (like infrastructure wear and tear or technological resources) can be stored in a warehouse, enabling predictive analytics for maintenance and financial planning.

Challenges:

- Data Integration Issues: Education systems often use disparate data sources (legacy systems, third-party applications, etc.), making it difficult to unify the data into a single repository.
- Privacy Concerns: Educational data contains sensitive information. Safeguarding student privacy under regulations like FERPA (Family Educational Rights and Privacy Act) requires robust security measures.

- Cost of Implementation: Establishing and maintaining a data warehouse can be expensive, especially for small institutions with limited budgets.

2. Data Mining in Education

Data mining refers to the process of discovering patterns, correlations, and insights from large datasets. In education, data mining is used to improve learning outcomes by identifying trends in student performance, personalizing learning experiences, and detecting at-risk students.

Applications:

- Personalized Learning: Data mining can help track a student's learning habits and tailor personalized recommendations. For example, algorithms can analyse a student's performance across different subjects to recommend additional resources or coursework based on areas of improvement.
- Early Warning Systems: Data mining can be used to identify students at risk of failing or dropping out by analysing their attendance records, assignment submissions, and academic performance. Educators can then intervene with personalized support.
- Adaptive Learning Technologies: Ed-tech platforms use data mining to adapt educational content in real-time, offering students content at the right difficulty level based on their learning pace and success.
- Curriculum Effectiveness Analysis: By analysing how students perform across different curricula and teaching methods, schools can modify or improve their instructional approaches to increase student engagement and success rates.
- Cheating Detection: Data mining can identify unusual patterns in student behaviour, such as similarities in test answers, helping institutions detect academic dishonesty.

Challenges:

- Data Quality and Completeness: Data mining algorithms require large, high-quality datasets to produce reliable results. However, data inconsistencies, missing information, or inaccuracies in educational data can hinder the effectiveness of mining techniques.
- Ethical Concerns: Mining student data raises concerns around ethics and consent. Students, parents, and faculty must be aware of how data is collected, processed, and used.
- Interpretability: Educators may find it difficult to interpret complex data mining results, and there's a risk of relying on statistical patterns without understanding their educational context.
- Bias in Algorithms: If historical data contains biases (e.g., unequal resources allocated to different demographics), data mining models could reinforce these biases rather than alleviate them.

3. Synergy Between Data Warehousing and Data Mining

When data warehousing and data mining work together, educational institutions gain powerful tools for both storing data efficiently and extracting actionable insights.

- Data warehousing provides a structured and scalable environment for long-term storage of massive educational data, from course enrolments to test scores and attendance.
- Data mining transforms this raw data into actionable insights, enabling personalized education, enhancing curriculum designs, and improving institutional performance.

For example, a data warehouse might store years' worth of test scores and attendance records. Data mining techniques can be applied to this dataset to identify factors that predict student success or failure, allowing educators to intervene earlier and adjust teaching methods.

4. Benefits in the Educational Domain

- Improved Learning Outcomes: By personalizing learning experiences and identifying at-risk students early, educational institutions can significantly improve overall student performance.
- Informed Decision-Making: Data mining provides insights that help school administrators make better decisions about resource allocation, faculty management, and curriculum planning.
- Increased Efficiency: Automating data collection, storage, and analysis reduces the administrative burden on educators and frees up resources for improving teaching and learning.
- Enhanced Student Experience: Students receive more personalized support, adaptive learning materials, and timely interventions, resulting in a more engaging and effective learning experience.

5. Future Trends and Considerations

- Learning Analytics: Data mining in the form of learning analytics will continue to play a pivotal role in shaping future learning environments, focusing more on personalized and adaptive learning.

- AI and Machine Learning Integration: Advances in artificial intelligence and machine learning will drive new levels of sophistication in educational data mining, allowing real-time feedback and recommendation systems for students and teachers.
- Increased Focus on Data Privacy: As the use of data mining and warehousing grows, institutions must balance innovation with strict adherence to data privacy laws and ethical practices.

Motivation for building data warehouse for the educational institute is from two sources, internal sources like inability ofcurrent operational systems to provide required information for parameter driven analysis and external sources like competitive factors. A survey is carried out by visiting several educational institutes to gather information regardingthe current practices the institutes have implemented as decision support systems.

The findings are summarized below.

1) The data is stored in different sources in distributed locations.
2) Users find difficulty in locating the reports needed by them.
3) The user interface for the current operational system is notsatisfactory and is confusing and hard to use for decision makers.
4) When the consolidated report from two or more different subject area is required, it is almost impossible.
5) There is no easy way to get assistance.

Utilizing a decision support system is a proactive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by distilling data from any combination of education records maintenance system. The purpose of this paper is to investigate current system of information delivery and propose a better system for timely, accurate, consistent information delivery to the decision makers of the educational institutes.

LITERATURE REVIEW

Following section briefly describesthe different application areas for which data warehouses are built.

A. Retail Sales:- Data is collected at several interesting places in a grocery store. Some of the most useful data is collected at the cash registers as customers purchase products. Modern grocery store scans the bar codes directly into the point_of_sale system. The POS system is at the front door of the grocery store where consumer takeaway is measured. The back door, where vendors make deliveries, is another interesting data collection point. At the grocery store, management isconcerned with logistics of ordering, stocking, and selling products while maximizing profit. Some of the most significant management decisions are on pricing and promotions. Both store management and marketing spend a great deal of time tinkering with pricing and promotions. In such scenarios, data warehouses come to rescue.

B. Telecommunications

A telecommunications company generates hundreds ofmillions of call-detail transactions in a year. For promoting proper products and services, the company needs to analyze these detailed transactions. The data warehouse for the company has to store data at the lowest level of detail.

C. Transportation

In this case, the airline's marketing department wants to analyze the flight activity of each member of its frequent flyer program. The department is interested in seeing which flights the company's frequent flyers take, which planes theyfly, what fare basis they pay, how often they upgrade, how they earn. These requirements can be fulfilled by data warehouse.

D. Education

There are some efforts in the area of data warehouse for building data warehouse for education domain. The paper by Carlo DELL'AQUILA summarizes the experience in designing and modeling an academic data warehouse. Existing facilities and databases affect the chosen data warehouse that brings them together to support decisional activities leading the whole university environment, including administrators, faculties and students. The choice to develop a dedicated system is mainly forced by the peculiar information type that defines the basic information in data warehouse widely different from institution to institution .In the article titled 'What academia can gain frombuilding a data warehouse' by David Wierschem, et. The authors have identified the opportunities associated with developing a data warehouse in an academic environment. They begin by explaining what a data warehouse is and what its informational contents may include, relative to the academic environment. Next they addressed the current environment drivers that provide the opportunities for taking advantage of a data warehouse and some of the obstacles inhibiting the development of an academic data warehouse. Finally, the article

provides strategies to justify developing a data warehouse for an academic institution.

DATAWAREHOUSE ENVIRONMENT Utilizing a decision support system is a proactive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by distilling data from any combination of education records maintenance system. The data mining from data warehouse can be a ready and effectivesystem for the decision makers. ETL activities are performed to extract the data from heterogeneous sources and load into staging and then load thedata into dimension and fact tables as per the schedules. We proceed to extract the BI report from data warehouse on demand based on requirement from the management. In an educational institute, main information required will be regarding key components of the education institute, namely students, employees and infrastructure. The purpose of this paper was to investigate current system of information delivery and proposing a better system for timely, accurate, consistent information delivery to the decision makers of the educational institute. The paper has been prepared in order to extend the usage of current available technology in decision making processes of educational institute.

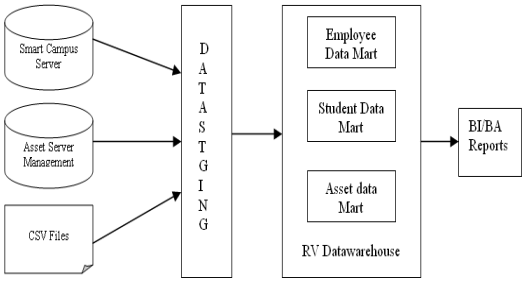

Fig. 1. EnggData warehouse architecture

Data warehouse enables the decision makers with benefitslisted below.

1) Phenomenal improvements in turnaround time for dataaccess and reporting
2) Standardizing data across the organization so that therewill be one view of information.
3) Merging data from various source systems to create amore comprehensive information source.
4) Reduction in costs to create and distribute informationand reports.
5) Encouraging and improving fact-based decision making.

BI-REPORTING

This refers to the variety of capabilities that can be provided to the users to leverage the presentation area for analytic decision making. All data access tools query the datain the data warehouse presentation area. A data access tool can be as simple as an ad hoc query tool or as complex as sophisticated data mining application. The majority of the users use pre-built parameter driven analytic applications to access the data. This enables them to retrieve the required information and analyze hidden pattern in the retrieved data .Using suitable data mining techniques, the useful information can be extracted from the data warehouse. Data mining form three main components of the institute, namely Employees, Students and Infrastructure. Employee data mart can provide the users with the information such as career growth and attrition rate. Student mart can provide the information related to the student like best outgoing student considering his academic and non academic activities. Information regarding assets such as the investment in a particular financial year can also be accessed.

RESULTS

Once the data warehouse is deployed, it invariably becomes a mission-critical application. Users depend on the data warehouse to provide them with the information they need to function properly. To make certain that the ETL process runs and completes, it must be actively monitored and supported. Some of the results observed after querying the data marts are documented below. The results are cross checked with the requirements specified by the differenttypes of users. The requirements with regard to asset information were to extract the information regarding the number of assets of each type in the Institute. The different data marts are queried using SQL query. The results returnedby the queries are found accurate and meeting users demands. The sample screen shots of queries and the results are shown.

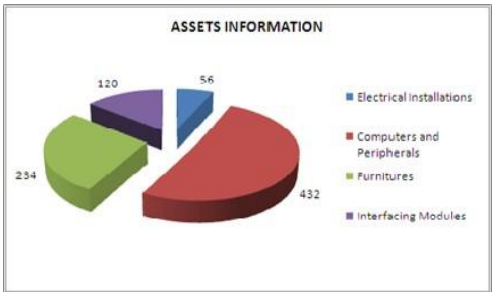

Fig. 2. Assets Information

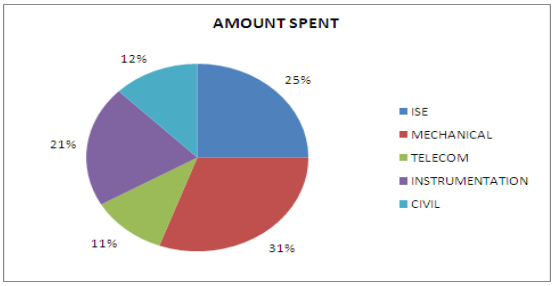

Fig. 3. Amount spent

Fig. 2 gives the asset information of the institute like electrical installations, computers and peripherals, furniture's and interfacing modules. Fig. 3 gives amount spend on each department viz EE, MECHANICAL, TELECOM, INSTRUMENTATION and CIVIL. Fig. 2 and Fig. 3 are output from asset mart.

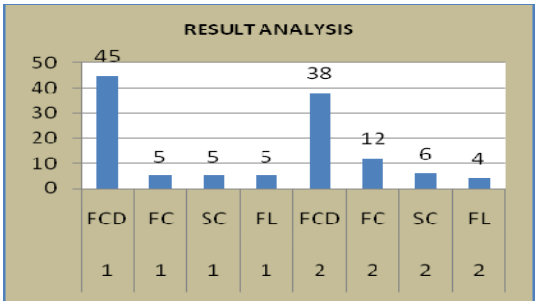

Fig. 4. Result Analysis

Fig. 4 shows the detailed result analysis which shows number of students who have obtained different classes; thisis the output from student mart.

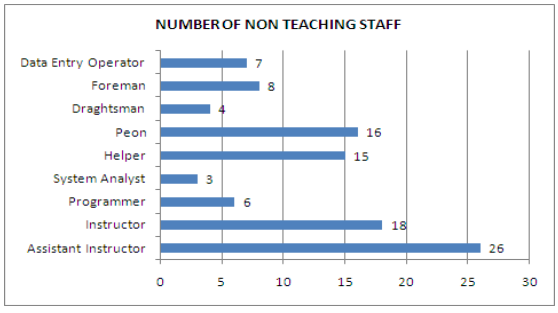

Fig. 5. Staff details

Fig. 5 gives the information regarding was the number of employees in each cadre of two particular departments. This is the output from employee mart.

CONCLUSION

Justifying a data warehouse project can be very difficult. Usually, analysis of the success of the data warehouse project is done considering the financial benefits against the investment. Since most of the educational institutes are nonprofit organizations and service oriented, the evaluation of the usefulness of the data warehouse can be done on the basis of its ability to meet user's requirements. The academicdata which was spread all across different sources has been loaded into single platform. The decision makers can extract information regarding three main components of the institute, namely Employees, Students and Infrastructure. Employee data mart can provide the users with the information such as career growth and attrition rate. Student mart can provide theinformation related to the student like best outgoing student considering his academic and non academic activities. Information regarding assets such as the investment in a particular financial year can also be accessed. In educational institute, decision makers ask "What are the expected resultsand benefits?" when making a data warehouse project rather than "What is the anticipated return on investment?". The data warehouse developed has met their expectations. Benefits of the present project can be more if the Institute haspositive approach towards new technologies. They can take micro-level decisions in a timely manner without the need to depend on their IT staff. They can perform extensive analysisof stored data to provide answers to the exhaustive queries to the administration cadre. This helps them to formulate strategies and policies for employees and students. This helps students and Employees in making decisions. They are the ultimate beneficiaries of the new policies formulated by the decision makers and policy planner's extensive analysis on student and employee related data.

Over all 80 to 85% of decisions are made based on the reportsgenerated by the proposed system. The realistic productivityis about 85%.

FUTURE SCOPE

The future scope of data warehousing and data mining in education is vast, driven by advancements in technology and the increasing importance of data-driven decision-making. These trends will shape the next phase of educational innovations, helping institutions better serve students, faculty, and administration. Below are key areas where data warehousing and data mining are likely to evolve and have a significant impact:

1. Advanced Learning Analytics

- Personalized Learning at Scale: As data mining techniques become more sophisticated, they will enable personalized learning experiences at an unprecedented scale. Machine learning models will analyze vast amounts of student data to customize

teaching methods, materials, and pacing for each learner.

- Real-Time Feedback Loops: With real-time data mining, educators will be able to offer immediate feedback to students, helping them improve learning outcomes more quickly. For example, platforms can automatically suggest additional resources or alternate teaching methods when a student struggles with a topic.
- Predictive Analytics for Learning Outcomes: Data mining models will become more accurate in predicting learning trajectories, allowing educators to proactively intervene when a student is at risk of falling behind. These predictive models will also help institutions forecast student retention, graduation rates, and job placement.

2. Integration of AI and Machine Learning

- AI-Powered Adaptive Learning Systems: Artificial intelligence (AI) will increasingly be integrated with data warehousing and mining systems to create adaptive learning platforms. These systems will tailor educational content in real-time, adjusting lesson difficulty and content delivery based on a student's performance and preferences.
- Automated Administrative Processes: AI-driven data mining can automate many administrative tasks, such as student admissions, resource management, and faculty scheduling. Predictive models can help in forecasting resource needs, staffing, and course enrollments.
- Natural Language Processing (NLP) for Educational Insights: NLP can be applied to analyze student interactions, feedback, and discussion forums. AI systems will interpret textual data to provide insights into student sentiment, understanding, and engagement levels.

3. Data-Driven Decision-Making for Institutional Management

- Strategic Planning and Policy Development: Data warehouses will enable institutions to analyze multi-year trends in enrollment, academic performance, and resource utilization. Data mining will help university leaders and administrators make informed strategic decisions about curriculum design, faculty development, and financial planning.
- Improved Resource Allocation: Predictive models can analyze student data to determine how resources (e.g., faculty, classrooms, technology) can be allocated more efficiently. For example, data can be used to optimize class sizes, course schedules, and resource distribution based on demand.
- Benchmarking and Institutional Research: Institutions will be able to use data warehouses for internal and external benchmarking, comparing their performance with similar schools or universities. This will foster continuous improvement by identifying best practices and areas for growth.

4. Student-Centric Predictive Models

- Student Success and Retention Models: The next generation of data mining algorithms will predict not only academic outcomes but also the likelihood of student success in broader contexts, such as career readiness or graduate program admission. Institutions will use these insights to tailor support services and interventions.
- At-Risk Student Identification: Advanced data mining will make early detection of at-risk students more precise by incorporating a wider variety of data points, including social behavior, financial factors, extracurricular activities, and well-being data. This can lead to targeted and timely interventions.
- Learning Path Optimization: Mining data on learning behaviors and preferences will lead to more optimized learning paths for students, helping them complete courses or degrees more efficiently while still achieving desired educational outcomes.

5. Enhanced Data Privacy and Security Practices

- Advanced Encryption and Privacy-Preserving Techniques: With the increasing volume of sensitive student data being stored and mined, privacy-preserving data mining techniques, such as differential privacy, will become more prevalent. These methods allow institutions to extract useful insights from student data without compromising individual privacy.
- Compliance with Global Data Protection Laws: As more countries adopt strict data protection laws (e.g., GDPR, CCPA), educational institutions will need to implement robust compliance measures. Future data warehousing systems will likely integrate features that ensure full compliance with global data protection regulations.
- Ethical AI Practices: The rise of ethical concerns around AI and data mining will push institutions to develop transparent and fair algorithms, ensuring that they are free from biases and provide equal opportunities to all students.

6. IoT and Wearable Technology Integration

- Learning Environment Data Collection: As educational institutions adopt Internet of Things (IoT) devices and wearable technology, the amount of data

available for warehousing and mining will increase exponentially. These devices will collect data on student movement, interaction with physical spaces (like classrooms or labs), and engagement with educational content.

- Physical and Mental Health Monitoring: Wearables can collect data on student physical activity, sleep patterns, and even stress levels. Mining this data will help institutions provide better support for student well-being, leading to improved learning environments and outcomes.

7. Blockchain for Secure Data Warehousing

- Decentralized and Immutable Records: Blockchain technology has the potential to transform data warehousing in education by providing a decentralized, immutable record of student performance, qualifications, and academic history. This will enhance the transparency and security of data, giving students ownership over their own educational data.
- Credential Verification: Educational institutions will use blockchain for credential verification, reducing the risk of diploma fraud and ensuring that student qualifications are easily verifiable by employers and other institutions.

8. Gamification and Behavioral Analytics

- Data Mining in Educational Games: Educational games generate vast amounts of behavioral data about how students interact with learning materials. Data mining can analyze this data to offer insights into student engagement, problem-solving strategies, and cognitive development.
- Motivation and Engagement Models: Data mining can identify what keeps students motivated by analyzing their behavior in online courses, discussion boards, and social interactions. This will allow educators to introduce game-like elements, such as badges or progress trackers, that enhance motivation.

9. Collaboration and Global Educational Networks

- Data-Driven Global Education Platforms: With the rise of massive open online courses (MOOCs) and international education platforms, data warehouses will store data from learners across the globe. Data mining algorithms can analyze global student behavior and performance, offering cross-institutional insights that can improve curriculum design and pedagogy on a worldwide scale.
- Inter-Institutional Data Sharing: Universities and schools will collaborate more extensively, sharing anonymized student and performance data to help improve educational standards. Federated data warehouses will enable such collaborations while maintaining individual data ownership.

10. Cross-Domain Applications

- Interdisciplinary Data Mining: Educational data will be combined with data from other domains like employment, social services, and health to gain deeper insights into the broader factors influencing student success. For example, mining data on student employment outcomes can help institutions fine-tune curricula to meet the demands of the job market.
- Lifelong Learning and Skill Mapping: With the future focus on lifelong learning, educational data mining will track an individual's learning progress over time, providing insights into skill acquisition, retraining needs, and career development paths.

CONCLUSION

The future of data warehousing and data mining in education is transformative. By harnessing advanced technologies like AI, blockchain, and IoT, educational institutions will create more personalized, efficient, and secure learning environments. Ethical considerations and privacy concerns will need to be carefully addressed, but the potential benefits are immense: from improved learning outcomes to better institutional performance and student well-being, data-driven approaches will play a pivotal role in shaping the future of education.

REFERENCES

[1]. Ralph Kimball, the Data Warehouse ETL Toolkit, Wiley India Pvt Ltd.,2006.
[2]. KV. K. K. Prasad, Data warehouse development Tools, Dreamtech Press, 2006.
[3]. W. H. Inmon, Building the Data Warehouse. Wiley; 3rd edition March 15, 2002.
[4]. Alex Berson, Data Warehousing Data Mining & OLAP, Computing Mcgraw-Hill, November 5, 1997.
[5]. Arshad Khan, SAP and BW Data Warehousing, Khan Consulting and Publishing, LLC (January 1, 2005)
[6]. Carlo DELL'AQUILA,'An Academic Data Warehouse' World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA ©2007.
[7]. McMillen and Randy McBroom.,'what

academia can gain from building a data warehouse' no.1,pp.41-46.2008

[8]. Channah F. Naiman, Aris M. Ouksel "A Classification of Semantic Conflicts in Heterogeneous Database Systems", Journal of Organizational Computing, vol. 5, 1995.

[9]. John Hess, "Dealing With Missing Values In The Data Warehouse" A Report of Stonebridge Technologies, Inc-1998.

[10]. Manjunath T.N, Ravindra S Hegadi, Ravikumar G K."Analysis of Data Quality Aspects in DataWarehouse Systems",(IJCSIT) International Journal of Computer Science and Information Technologies, vol. 2(1) , 2010, 477-485.

[11]. Jaideep Srivastava, Ping-Yao Chen,"Warehouse Creation-A Potential Roadblock to Data Warehousing", IEEE Transactions on Knowledge and Data Engineering January/February 1999 (vol. 11, no. 1) pp.118-126.

[12]. Amit Rudra,Emilie Yeo (1999) "Key Issues in Achieving Data Quality and Consistency in Data Warehousing among Large Organizations in Australia", Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999.

[13]. Jesús Bisbal et all,"Legacy Information Systems: Issues and Directions", IEEE Software September/ October 1999.

[14]. Scott W. Ambler "Challenges with legacy data: Knowing your data enemy is the first step in overcoming it", Practice Leader-2001.

[15]. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.).

[16]. Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

[17]. Baker, R. S. J. d., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1(1), 3-17.

[18]. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A Reference Model for Learning Analytics. International Journal of Technology Enhanced Learning, 4(5-6), 318-331.

[19]. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review, 46(5), 30-40.

[20]. Sharma, S., & Singh, N. (2016). Big Data and Predictive Analytics in Education: A Review and the Future Scope. International Journal of Engineering Research & Technology, 5(6), 455-460.

[21]. Zorrilla, M. E., & García-Saiz, D. (2013). A Service-Oriented Architecture to Provide Data Mining Services for Non-Expert Data Miners. Decision Support Systems, 55(1), 399-411.

[22]. Rodríguez, P., Nussbaum, M., López, X., & Sepúlveda, M. (2010). A Monitoring and Assessment Framework for Learning Communities. Educational Technology & Society, 13(2), 274-285.