

Analysis of Data Mining Model for Successful Implementation of Data Warehouse in Tertiary Institutions

Amit Mishra

ABSTRACT

This research work with title “analysis of data mining model for successful implementation of data warehouse in tertiary institutions” is a proposal that provides a framework that is used to structure, plan, and control the process involved in information discovery for tertiary institutions. It equally deals with series of steps or procedures which governs the analysis and design of this particular Data Mining Model for Tertiary Institutions. The methods, techniques and procedures which are used to collect and analyze information were also provided and implemented. Object oriented Analysis and Design methodology (OOADM) was deployed to develop this model. The current system was analysed by using Use Case approach to identify the existing objects in the current system. Use Case diagrams were equally developed and the relationship between the objects was drawn. A diagram was drawn to represent the identified problems. Finally, the High-Level Model (HLM) that would be used to design the proposed system was drawn to enable the System Designer to develop a computer-based model that forecasts educational behaviours with little or no user intervention.

Keywords: Framework, Information Discovery, Tertiary Institutions

Article Info

Publication Issue :

Volume 8, Issue 5
September-October-2022

Page Number : 305-316

Article History

Accepted: 10 Oct 2022
Published: 28 Oct 2022

I. INTRODUCTION

The last decade has experienced great change in information availability via the internet. Therefore, more and more businesses and organizations began to collect data that are relevant and related to their own operations. While analysts have been seeking efficient and effective way of storing, retrieving and manipulating data; the machine learning community has focused on building and developing various techniques for learning and acquiring knowledge from the data. In this regard, it is often referred to as

data mining (DM) and knowledge discovery in databases (KDD).

Since human life began people have been seeking pattern in data mining. Politicians seeks pattern in voters' opinion, hunters seek pattern in animal migration behaviour and farmers seeks pattern in crop growth. A scientist's work is to make intelligence out of data, to determine the core or important model that governs the performance of the physical world and encapsulate the same in theories that can be used to predict the future. As the background of all the scientific discoveries especially theories has been same, what

then is new about data mining (DM)? The simple answer is that, in data mining the amount of the stored data is in the digital or electronic format and the search is automated or improved by a computer. In data mining, it is important to understand the difference between a model and a pattern. Model is an intangible representation of reality or universal summary of the dataset and makes statement about any point in the full measurement space. While pattern describe an arrangement, a relationship to a relatively small part of the data or the space in which data would occur. In 1960's, computer was increasingly applied to data analysis problems and it was renowned that if one searches long enough, one can always find some model to fit in the dataset but complexity and size of the model were important considerations.

Analysis of Data Mining Model for successful implementation of Data Warehouse in tertiary institutions was geared towards data mining tertiary institutions' huge data warehouse which involves digging through heaps of data in search of information which the management uses for proper running of the institution as well as to make prospect predictions based on the current trend of events. Object Oriented Analysis and Design (OOADM) methodology was deployed to design this model. Source data was collected from the department of Admissions/academic planning and Exams and records sections in the University.

STATEMENT OF THE PROBLEM

The topic focuses on understanding how the present system works, the problems that are inherent in the present system as well as analyze and design a proposed system using Object oriented Analysis and Design (OOADM) methodology.

OBJECTIVES OF THE STUDY

The objectives of this research work are to develop a system that should be able to:

- (i) Find relationship between

- a) Results and admission procedure.
 - b) Results of courses and the lecturers that taught them in different sessions.
- (ii) Investigate the existing system and the problems affecting it.
 - (iii) And analyzing a new system so as to enable the System Designer, build up a computer-based model that predict educational behaviours with little or no user intervention.

II. METHODOLOGY

Methodology is defined as a system of wide ideology or rules from which specific methods or procedures may be derived to deduce or work out different problems within the range of a particular discipline. Methodology is not a formula but a set of practices. It is used to refer to a specific series of steps, methods, techniques and measures which governs the collection, analysis and design of a particular project. Methodology is defined as a outline that is used to structure, plan, and control the process of developing an information system. It is used. The researcher used Object oriented Analysis and Design (OOADM) methodology to analyze the present system as well as to design the proposed system with the primary aim of:

- (i) Identifying the problems innate in the present system.
- (ii) Investigating the causes of these problems
- (iii) And proffering a likely solution to these system

RESEARCH DESIGN

The researcher adopted a survey for this study. Fritz Scheuren defines survey as a way often used to describe a technique of assembling information from a sample of individuals. This sample is usually just a fraction of the population being studied. Olagbewo (2006) also defines survey as design which documents

the nature, scope, dimension, direction of events, behaviours, attitude etc about things or persons.

AREA OF THE STUDY

The study was conducted in Ibrahim Badamasi Babangida University Lapai, Niger State.

POPULATION

The research/study population consists of the Examination Office, Faculty of Natural Science and Academic department IBB University Lapai.

INSTRUMENT OF DATA COLLECTION

The research instrument for data collection was interview method. The researcher conducted an interview in order to find out the stages of admission processing, how exams records are collated, Association between results and admission and the process of data storage and mining.

ANALYSIS OF THE PRESENT SYSTEM

A study to these two departments/units revealed the different methods they engaged in carrying out their daily activities.

ADMISSIONS PROCEDURE

The admissions department is charged with the duty of processing: -

- (i) Students' admission application
- (ii) Students' change of Department application.
- (iii) Students' direct entry application.

The processing of these applications takes the following stages for undergraduate studies

- (i) Unified Tertiary Matriculation Examination (UTME) sends list of successful candidates to the university.
- (ii) Cut off mark is set by the school management.

(iii) Candidates who UTME score exceeds the UTME cut off mark are short listed.

(iv) Direct entry student must:

- a) Possesses Ordinary National Diploma (OND) or Higher National Diploma (HND) or Nigerian Certificate in Education from a reputable polytechnic or colleges of Education.
- b) HND holders must have at least lower credit.
- c) OND holders must have upper credit.
- d) NCE holders must have upper credit.

(v) Students pay school fees and document themselves

The above description describes the admission process into the various undergraduate programmes run by the university.

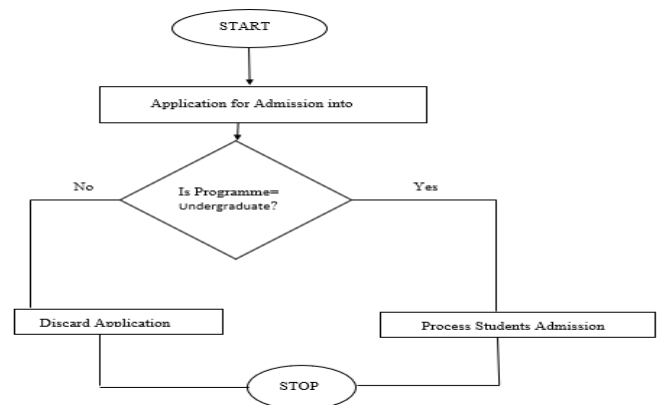


Figure 1: System flowchart of Admissions Application

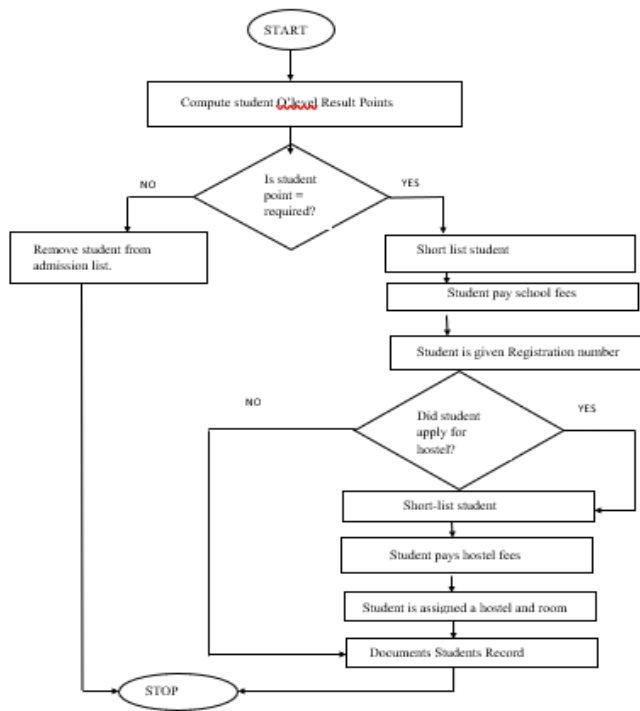


Figure 2 : System flow chart of student admission and registration

EXAMS AND RECORD PROCEDURE

The result and transcript of students are vital documents that are derive from various departments and finally kept at the exams and records department of the university. This unit uses the following technique in collating and documenting students’ results

1. Lecturers forward a copy of the result courses they handled for a semester to exams and records in the department.
2. These results are documented and forwarded to the Faculty exams and record.
3. The results are then vetted by vetting officers/senate.
4. A student request for a transcript through the bursary, and the department is notified.
5. The Bursary raises the student transcript, and forwards it to the academic unit for documentation and then forwards it back to the department.
6. The transcript generated is cross checked with the one at exams and record.

7. The transcript is recomputed at exams and record.
8. The recomputed transcript is dispatched to the address requesting it.

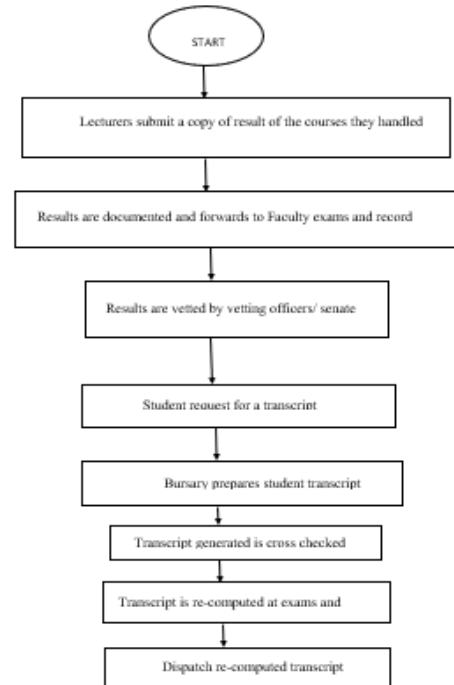


Figure 3: Flow chart of exams and transcript computation

The faculty exams and record department also prepare departmental lecture time table, examination time table and final result statement.

The process of result statement is carried out thus,

- a) Faculty forwards the final result of the students at the end of their study to the university senate.
- b) The senate vets the result to make sure there are no students who have over stayed in the school and no form of fraud in the result computation process.
- c) Senate sends back the approved and vetted result to the department.
- d) The department forwards a copy of the approved result to the exams and records.

- e) Exams and records section prepare a statement of result for each student in the result sheet.
- f) Students sign the duplicate copy of the statement of result and keep the original copy.
- g) Exams and record officer documents the duplicate copy signed by the student.
- h) Exams and record officer prepare a certificate to be awarded to the student on the next convocation.

ANALYSIS OF THE PROPOSED SYSTEM

The following stages are used in object-oriented analysis and design methodology to analyze and design a system within its boundary.

1. Preliminary Investigation (initial survey) phase

This is the initial project investigation. The objective is to identify what system is required, what should be its broad scope, tentative cost, time and other resources requirements. Once this broad and tentative information is collected, the decision is made whether to proceed developing the system or not. The goal of system investigation phase is to answer the following questions:

- a. What is the business problem?
- b. Is it a problem or an opportunity?
- c. What are the main causes of the problem?
- d. Can the problem be solved improving the current information system?
- e. Is a new information system needed?
- f. Is this a feasible information solution to this problem?

2. Problem Analysis Phase

This stage provides the analyst sufficient opportunity to understand thoroughly the problem, opportunities and/or directives that triggered the project. This stage

is primarily concerned with system owners and system users view of the existing system.

3. Requirement (use case) Analysis Phase

This stage defines the business requirement for the new system. Or is the process of discovering, analyzing, defining and documenting the business requirements. It involves measuring each functional and non-functional requirement, technical, operational and transitional requirements. This phase answers the question

- a. what do users need and want from the existing system?

To answer this question, the following activities must be embarked upon at this place:

- (i) **Define use cases.**

A use case represent a class of functionality provided by the system or actions that the user might carry out in order to complete a specific task.

- (ii) **Analyze use cases (requirements).**

4. Decision Analysis Phase

Decision making is a framework that helps project managers solves wide range of decision-making problems. Any decision analysis process is based on three main policies, called the 3c's principle.

I. **Consistency** – is an important aspect of decision analysis process for similar kinds of problems and opportunities to enable consistency in decision making over time.

II. **Comprehensiveness** – decision analysis process should include a comprehensive assessment and analysis of the business situation. Missing or incomplete information can lead to incorrect decisions.

III. **Continuity** – the value of decision analysis will significantly diminish if it is done only on discrete situation during the course of the project. Decision analysis is an unremitting process of constructing and refining decisions during a course of a project.

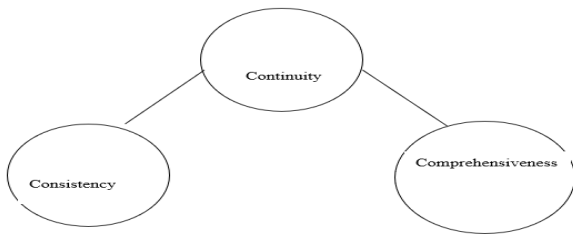


Figure 4 : the 3c' principle of decision making

5. Design Phase

The purpose of the design phase is to devise a solution for the problem specified by the requirement. This stage is the first step in moving away from problem domain to solution domain. This document is similar to a proposal or plan for the solution, and is used later at the implementation, testing and maintenance stage.

The design activity is often divided into two separate stages.

- a) **System design** which is sometimes also called a top-level design. This aims to identify the modules that should be in the system, the specification and how they interact with each other and the desired results. At the end of the system design, all the major data structure, file format, output format as well as the major modules in the system and specification are decided.
- b) **Detail design** the internal logic of each of the modules specified in the system design is decided. During this stage further details of the structures and algorithmic design of each module is specified.

6. Implementation and Testing Phase

System implementation is the development, installation and testing of system components and delivery of that system into production. (Bentley *et al.*).

System implementation has several major activities. There are five major tasks in this phase; coding, testing, installation, documentation and training as in

Figure 5. The purpose of this phase is to convert the physical system specifications into working and reliable software and hardware, document the work that have been done and provide help for the existing and future users.

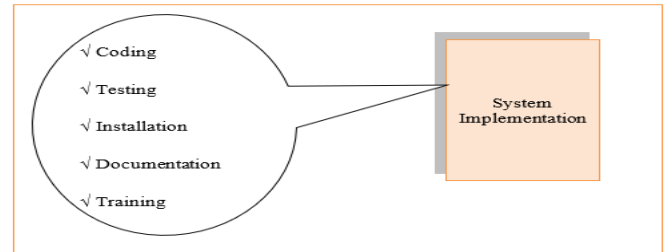


Figure 5 : Five Activities in the System Implementation Phase

a) Coding

Coding is the process whereby the physical specification that is created in the preceding stages is turned into working computer codes by the programmer team.

b) System testing

Testing is the process of running the program with a set of data to check for the correctness of the program (Eze, 2007.)

c) Installation

Installation is the process of moving away from an existing system to the new, enhanced or present system.

i. Direct installation

Direct installation is also referring to abrupt cut-over installation. Direct installation is where an existing system is turned off and the new system is turned on.

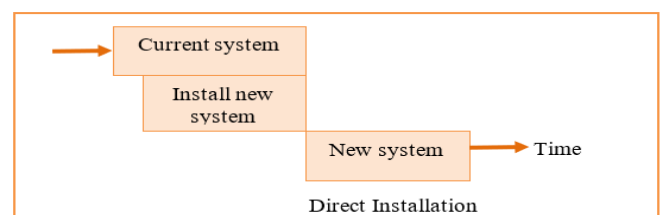


Figure 6 : Direct installation

ii. Parallel Installation

Parallel installation is when both system; the existing and new system are operating at the same time until the management decide that the existing system can be turned off as in Figure 7.

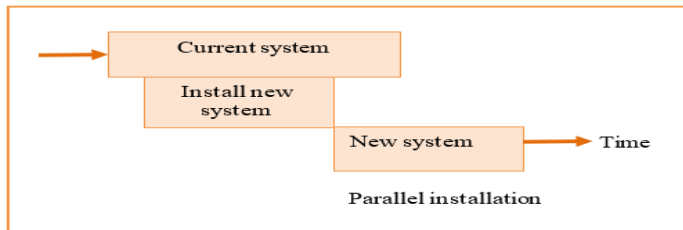


Figure 8 : Parallel installation

iii. Single Location Installation

Single location also refers to location or pilot installation. This approach is where the organization is trying out to use a new system at one location and then use the experience in deciding how the entire system should be deployed throughout the organisation as in Figure 9.

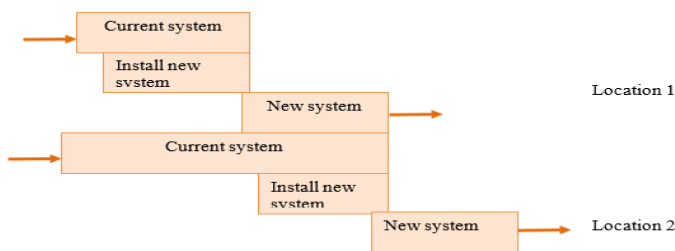


Figure 9 : Single-location installation

iv. Phase Installation

Phase installation is similar with single-location but the difference is the new system is installed in functional components.

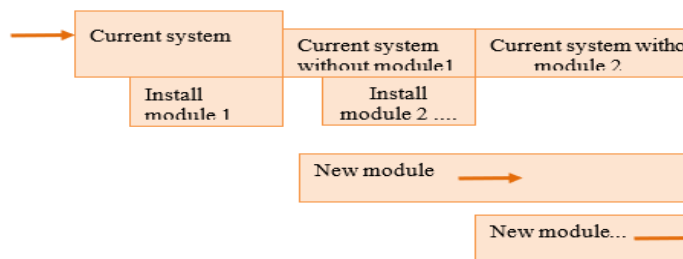


Figure 10 : Phased installation

Installation may also require a system acceptance test plan. System acceptance test is a final opportunity where end users, management and information system operation decide either to use or reject the

system. System acceptance test is a test performed on the final system wherein users conduct verification, validation and audit tests (Bentley *et. al.*, 2007).

7 Documenting the System

Every system development is unique and needs unique documentation. There are two basic types of documentation; system documentation and user documentation.

i. System documentation

System documentation records detailed information about system’s design specification, its internal workings and its functionality (Hoffer *et. al.*, 2005).

ii. Internal documentation: Internal documentation is part of the program source code or is generated at compiled time.

iii. External documentation: external documentation is a system that includes the outcome of structures diagramming techniques such as data flow entity-relationship diagrams.

b. User documentation

As stated earlier, system documentation is intended for system maintenance programme; user documentation is intended for the user of the system. There are few types of user documentations;

i. **Reference guide**- consists of exhaustive list of system functions and command.

ii. **Quick reference guide**- provides essential information about operating the system in a short and concise format.

iii. **User’s guide**- provides information on how users can use a computer system to perform and accomplish some specific tasks.

OBJECT-ORIENTED ANALYSIS AND DESIGN TERMINOLOGY

I. **Actor:** this represents anything (such as users or roles) that needs to interact or interface with to exchange information.

Use case: a use case describes a series of activities or actions perform by a system to yield visible results.

Use case Modelling: a use case model is a view of a system that emphasizes the behaviour as appears to the outside users.

Polymorphism: is the attribute of being able to as assign more than one meaning to a thing in different contexts specially to allow an entity such as a variable, a function or an object to have more one form.

Message Passing: it is done when one object hooks to one or more other object methods to request information or action. Messages are intrinsic elements of unified modelling language interaction diagrams.

Inheritance: It explain methods or attributes defined in one object class and can be reused in another object class. Inheritance is the method that permits new classes to be created out of the existing classes by extending and refining its capabilities.

Class: It is a set of an object that shares the same features and methods behaviour.

Behaviour: are things that an object can perform which correspond to functions that act on their features (data).

Attribute: are data that represents features or characteristics of interest about an object. The characteristics of an instance of a class are called attributes that are implemented as variable inside a class when defining software entity.

Object: is anything that is capable of being seen, touched or sensed and about which users store data (features) and combine behaviour.

Unified modelling language: It is a set of modelling tools or equipment used to specify a software system in terms of objects.

Encapsulation: It is a conceptual independence where everything in the product that relates to the portion of the world is modelled by that object is found in the object itself. (Stephen, 1996) defines encapsulation as the packing of several items together into one unit.

USE CASE DIAGRAM

The use case diagram represents the interaction among actors and use cases and equally specifies the communication between a user and the system. It defines the environment and its interaction with the use cases of the proposed systems.

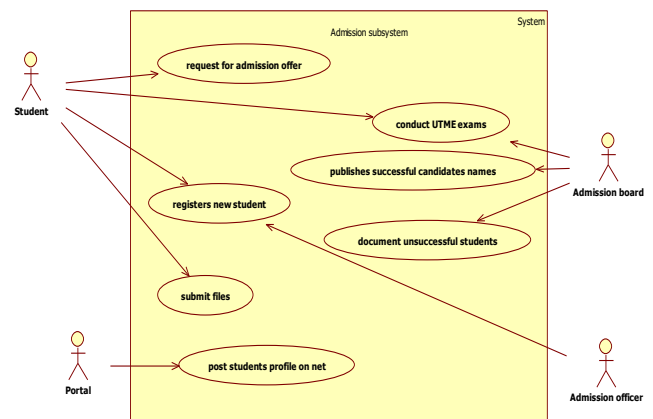


Figure 11: A use case diagram of Admissions subsystem

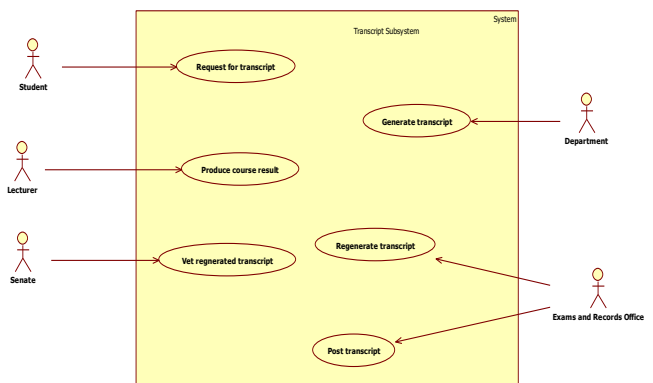


Figure 12 : a use case diagram of transcript subsystem

Table 1 : use case analysis table for generating information admission use case

USE CASE NAME	REGISTER NEW STUDENT	
USE CASE DESCRIPTION	This use case describes the procedure for student registration. On Completion the student is given a matriculation number.	
Typical courses of events	<p>Actor action</p> <p>STEP 1: this use case is initiated when a student prints his/her letter of admission from the university portal.</p> <p>STEP 2: student submits the admission offer to the admission office.</p>	<p>System response</p> <p>STEP 4: the admission officer verifies the student particulars with the list of admitted student from the admission board.</p> <p>STEP 5: student pays schools fees.</p> <p>STEP 6: accepts student form and issues student matriculation number.</p>
ALTERNATE COURSE	STEP 3: if the student records are not found in the original list, the student arrested and investigated.	
PRE-CONDITION	Registration can only be carried out only by the students who have met the requirements and criteria for admission.	

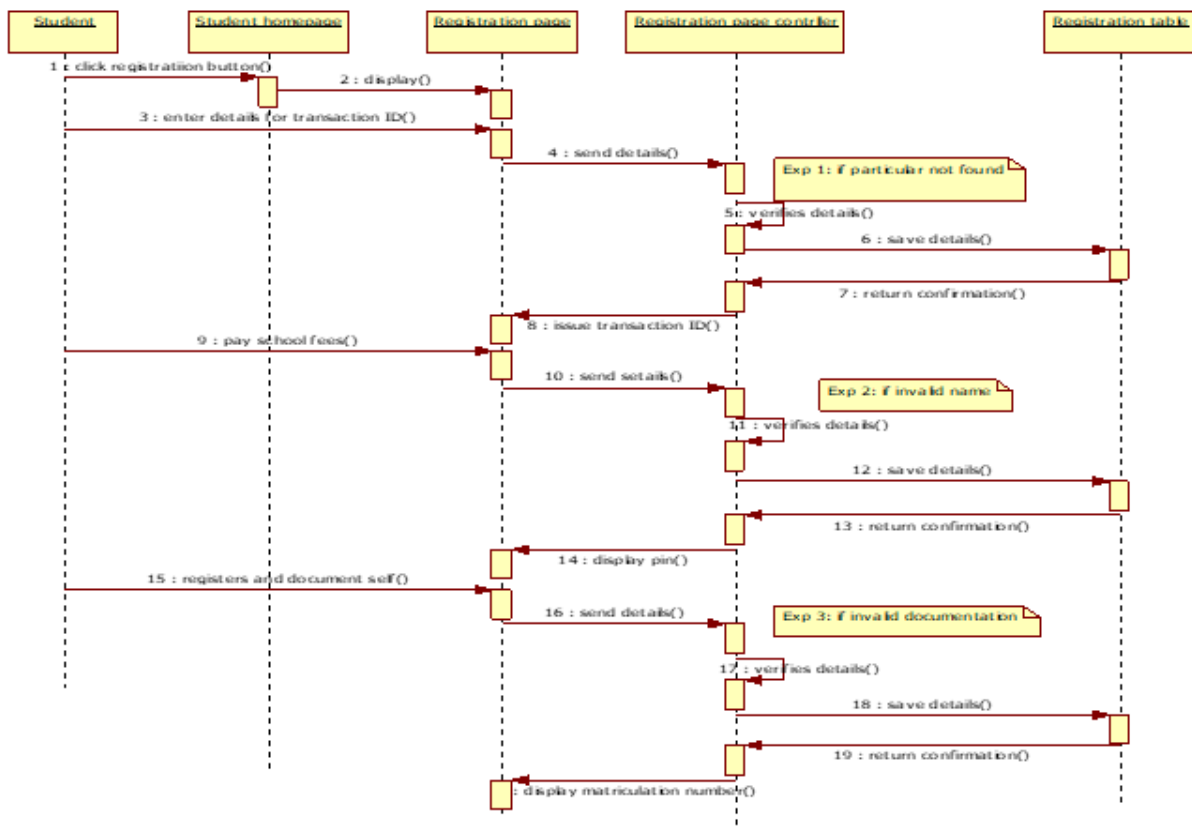


Figure 13: analysis sequence diagram for student registration

Table 2: Use case analysis for generating student transcript

USE CASE NAME	GENERATE TRANSCRIPT	
ACTORS	STUDENTS, EAXAMS AND RECORD OFFICER AND DEPARTMENT	
DESCRIPTION	This use case describes the process of generating students transcript	
Typical course of Action	<p>Actor action</p> <p>STEP 1: this use case is initiated when a student apply for his/her transcript.</p> <p>STEP 6: the use case terminates when the generated transcript is received by those demanding it.</p>	<p>System response</p> <p>STEP 2: student personal academic records are validated</p> <p>STEP3: student file is retrieved at exams and record and transcript request is forwarded to the department</p> <p>STEP 4: transcript is generated for the student.</p> <p>STEP 5: the transcript is vetted and then sent to where it is being requested.</p>
ALTERNATE COURSE	STEP 7:if student file is not found, the admission record is checked to see if the student was properly admitted into the university and if so, new file is generated.	
PRE-CONDITION	Transcript can only be raised for bonafide student of the university who have Complete their undergraduate studies.	
POST-CONDITION	transcript is sealed and sent to its destination	

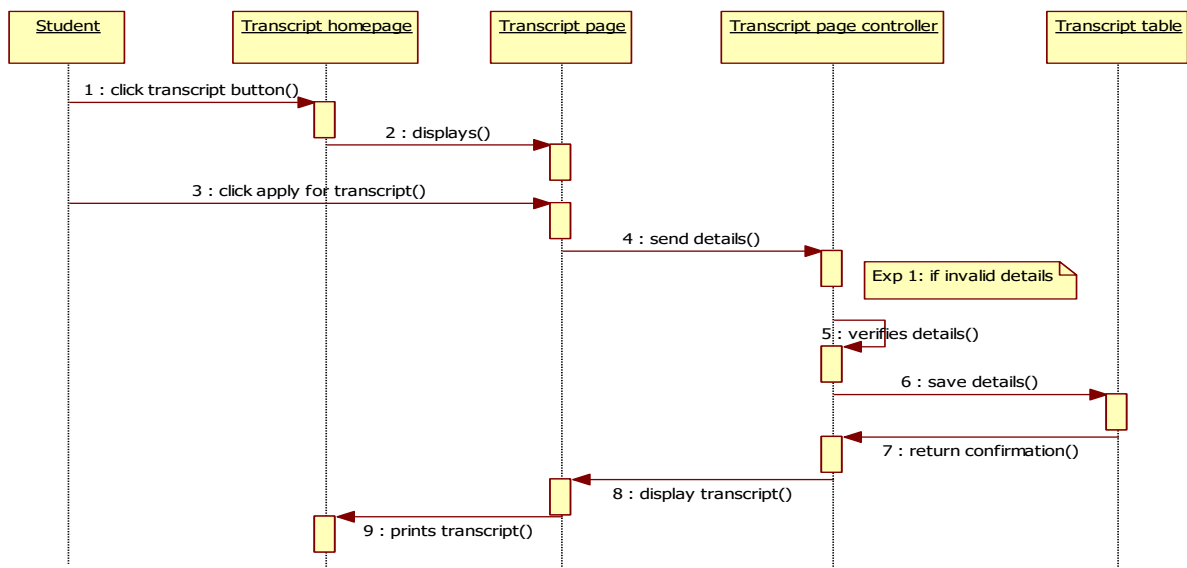


Figure 14 : analysis sequence diagram for transcript generation

FINDING AND IDENTIFYING OBJECTS

The next stage of this analysis is finding and identifying business objects. The business objects that are identified at the course of this research are mentioned below:

- Lecturers
- Students
- Admission officer
- Admission list
- Destination
- Transcript
- Exams and record officer
- Faculty
- Department
- Session
- Admission record
- Result
- Student file
- Portal
- University senate
- School fees
- Academic program

PROBLEMS IDENTIFIED WITH THE PRESENT SYSTEM

The problems that are identified in the existing system are enumerated below:

- I. Poor performance evaluation
- II. Lack of data mining algorithm
- III. Unstable admission procedures and statistics
- IV. Records still stored in files and excel format

ADVANTAGES OF THE PROPOSED SYSTEM OVER THE PRESENT SYSTEM

The proposed system has better and great advantage than the existing system because it provides the following facilities:

- I. Provide algorithm for mining and extracting data
- II. Provide a stable admission procedure
- III. Provide data warehouse infrastructure
- IV. Load, clean and extract data in warehouse

LIMITATIONS OF THE PROPOSED SYSTEM

- i. The proposed system requires costly and expensive computing equipment.
- ii. It requires too much work load in building it.
- iii. Requires extensive research laboratory ready with internet connections to source information.
- iv. The proposed system requires a lot of fund to begin.
- v. It requires a lot of brain storming ideas.

JUSTIFICATION OF THE NEW SYSTEM

- i. The proposed system will help the institution to find the relationship between results of courses and the lecturers that taught in different sessions.
- ii. The proposed system will help in finding the relationship between results and admission procedure.
- iii. The proposed system will provide a better performance evaluation

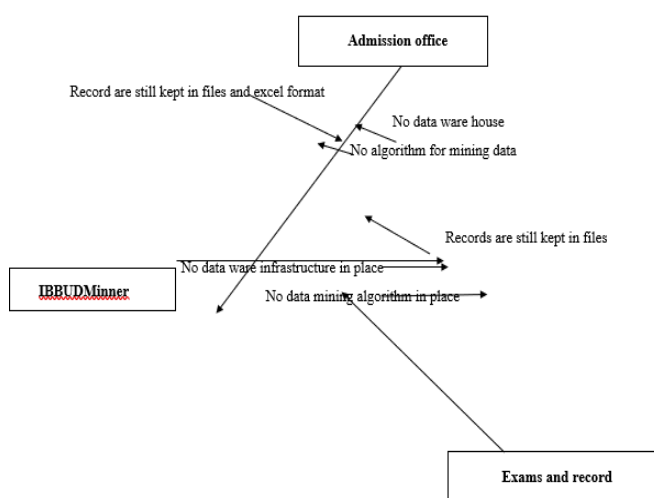


Figure 15 : Diagram that shows the weaknesses identified in the present system

iv. The new system will help in providing stable and normal admission procedures and statistics.

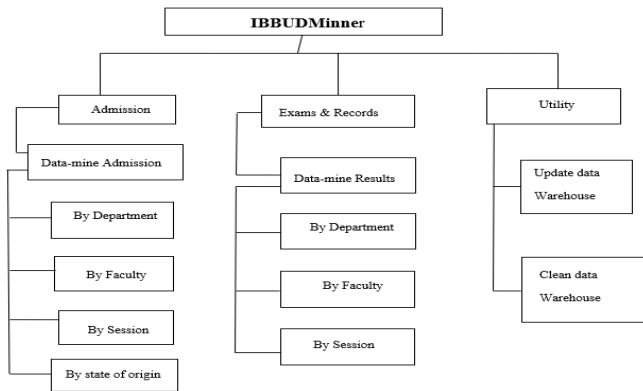


Figure 16: High level model of the proposed system

Summary

The research investigated on analysis of data mining model for successful implementation of data warehouse in tertiary institution in Ibrahim Badamasi Babangida University Lapai, Niger State. The literature review focused on several works done on data mining techniques/models by other researchers. The method used in data collection was interview which was carried out by the researcher and it was validated by experts before it was administered and the methodology used was object-oriented analysis and design to propose the new system.

III. CONCLUSION

Using a particular methodology to analyze a given system in a dynamic world is the major activity in software development. The researcher applied all the required methods, techniques and procedures to investigate and analyze the present system in order to proffer solutions to the identified problems. Finally, the proposed system tends to provide inter-relationship amongst these two departments in the institution for appropriate decision making and proper management.

IV. REFERENCES

[1]. Anthony Danna, Oscar H. Gandy, (2002). Journal of business ethics: All that glitters is not gold Digging Beneath the surface of data mining.

[2]. Cesar V, Javier B, liela S and Alvaro O (2009). Recommendation in Higher Education using data mining techniques, In Proceedings of the Educational Data Mining, conference.

[3]. Fadzilah S. and Abdoulha M., (2009). Uncovering hidden information within University's Student enrolment data using Data Mining, in proceedings of the third Asia international conference on modelling & simulation conference, IEEE Computer Society. M E. Fayad SJSU-compE L09546 fall (2002).

[4]. Muslihah W., Yuhanim Y., Norshahriah W, Mohd Rizal M., Nor Fatimah A., and Hoo Y. S., (2009). Predicting NDUM Student's Academic Performance Using Data Mining Techniques, In Proceedings of the Second International Conference on Computer and Electrical Engineering, IEEE Computer Society.

[5]. Naeimeh D., Mohammad S., and Mohammad B., (2004). A New Model for Using Data Mining Technology in Higher Educational Systems in Proceedings of the IEEE Conference.

[6]. Naeimeh D., Mohammad B., and Somnuk P., (2005). Application of Enhance Analysis Model for Data Mining Processes in Higher Educational System, In Proceedings of the ITHET 6th Annual International Conference, IEEE.

[7]. Naeimeh D., Somnuk P., and Mohammad B., Vol. 7, No. 1, pp. 31-54. (2008). Data Mining Application in Higher Learning Institutions, Informatics in Education.

[8]. Nguyen N., Paul J., and Peter H., pp. 7-12, (2007). A Comparative Analysis Techniques for Predicting Academic Performance in Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference.

[9]. Pathom P, Anongnart S, and Prasong P. (2008). Comparisons of Classifier Algorithms: Bayesian Network, C4.5, Decision Forest and NBTree for Course Registration Planning Model of Undergraduate Students, Sripatum University

Chonburi Campus, Office of Computer Service,
Chonburi Thailand, IEEE.

- [10].Ramaswami M., and Bhaskaran R., Vol. 7, Issue 1, No. 1, (2010). CHAID Based Performance Prediction Model in Educational Data Mining, IJCSI International Journal of Computer Science Issues.
- [11].Tissera R, Athauda I, and Fernando C.(2006). Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining, ICIA, IEEE.
- [12].Witten, I, Frank E.(2000).Data Mining. Academic Press.
- [13].Yiming M, Bing L, Ching W., Philip Y., and Shuik L., (2000).Targeting the Right Students using Data Mining, ACM.

Cite this article as :

Amit Mishra, "Analysis of Data Mining Model for Successful Implementation of Data Warehouse in Tertiary Institutions (A Case Study of Irahim Badamasi Babangida University) ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 305-316, September-October 2022. Available at doi : <https://doi.org/10.32628/CSEIT228531>
Journal URL : <https://ijsrcseit.com/CSEIT228531>