A

Major Project Report

on

PREDICTION OF STUDENTS PERFORMANCE USING DATAMINING APPROACH

Submitted in Partial Fulfillment of
the Requirements for the Degree
of

Bachelor of Engineering

in

Computer Engineering

to

North Maharashtra University, Jalgaon

Submitted by

Kalpesh P. Chaudhari
Riya A. Sharma
Rajeshwari J. Bari
Shreya S. Jha

Under the Guidance of

Mr. K. P. Adhiya

DEPARTMENT OF COMPUTER ENGINEERING
SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY,
BAMBHORI, JALGAON - 425 001 (MS)

2016 - 2017

SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY, BAMBHORI, JALGAON - 425 001 (MS)

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the major project entitled *Prediction of Students Performance using Datamining Approach*, submitted by

Kalpesh P. Chaudhari
Riya A. Sharma
Rajeshwari J. Bari
Shreya S. Jha

in partial fulfillment of the degree of *Bachelor of Engineering* in *Computer Engineering* has been satisfactorily carried out under my guidance as per the requirement of North Maharashtra University, Jalgaon.

Date: October 14, 2016
Place: Jalgaon

Mr. K. P. Adhiya
Guide

Prof. Dr. Girish K. Patnaik
Head

Prof. Dr. K. S. Wani
Principal

Acknowledgements

We hereby take this opportunity to express our heartfelt gratitude towards the people whose help is very useful to complete our project work on topic of **Prediction of Students Performance using Datamining Approach**. We would also greatly thankful to our Principal **Dr. K. S. Wani** , H. O. D. **Prof. Dr. Girish K. Patnaik** for providing precious help and advice. We would like to express our heartfelt thanks to our project guide **Mr. K. P. Adhiya** whose experienced guidance becomes very valuable for us. We would like to thank our parents for their support. Finally We would like to thank all my friends and well-wishers who have a very important role in nurturing and strengthening our career.

Kalpesh P. Chaudhari
Riya A. Sharma
Rajeshwari J. Bari
Shreya S. Jha

Contents

List of Tables

List of Figures

Abstract

The success of an educational institution can be measured in terms of quality of education it provides to its students. In education system, highest level of quality can be achieved by exploring the knowledge regarding redirection about students performance. Nowadays the lack of existing system to analyse and judge the students performance and progress is not being addressed. There are two reasons why this is happening. First, the existing system is not accurate to predict students performance. Second, due to the lack of consideration of some important factor that are affecting students performance. Predicting students performance is more challenging task because of large amount of data in educational database. This proposed system will help to predict students performance more accurately. For this the suitable datamining approach will be applied.In this approach, preprocessing step will be applied to raw dataset so that the mining algorithm will be applied properly. The prediction about students performance will help her/him to improve the performance.

Chapter 1

Introduction

There are increasing research interests in using data mining techniques in education. This emerging field called Educational Data Mining. It can be applied on the data related to the field of education. One of the educational problems that are solved with data mining is the prediction of students' academic performances. Prediction of students' performance is more beneficial for identifying the low academic performance students. Student retention is an indicator of academic performance and enrolment management of the university. The ability to predict a students performance is very important in educational environments. Students academic performance is based upon different factors like social, personal, Psychological and other environmental variables. A very promising tool to achieve this objective is the use of Data Mining. Data mining techniques are used to discover hidden patterns and relationships of data, which is very much helpful in decision-making. Now-a-days educational database is increased rapidly because of the large amount of data stored in it. The loyal students motivate the higher education systems; to know them well is by using valid management and processing of the students' database. Data mining approach provides valid information from existing student to manage relationships with upcoming students[1].

Introduction chapter will introduce the work, It will focus exactly on what is the area of project and explains what is actually be done in this work. All ideas about project work are cleared here. The sections of the chapter are organized as follows. Section 1.1 presents background. Section 1.2 presents Problem Denition. Scope of the project is discussed in section 1.3. Section 1.4 presents objective of the project. Finally, summary of the chapter is given in the last section.

1.1 Background

Before starting of the proposed system some necessary information related to the EDM is given below.

1.1.1 Electronic Learning (E-Learning)

E-learning, as one of the main areas of e-services, has undergone intensive development as an inevitable result of the recent proliferation of internet technology. Traditional learning means restrict the student to certain learning methods, at a particular place and time whereas e-learning services create wider ranges for organizations and individuals who involved in learning and teaching. These environments facilitate the delivery of large parts of education through the use of tools and materials that are accessible directly to the learners home or office and at any time. In addition, the advancements in technology, which are used to enhancing the interactivity and media content of the web and the increasing quality of delivery platforms, create an ideal environment for the expansion of e-learning systems[2].

1.1.2 Data Mining

Data mining is the process of discovering various models, derived values and summaries from a given collection of data. It is important to realize that the problem of discovering or estimating dependencies from data or discovering new data is only one part of the general experimental procedure used by engineers, scientists and others who apply standard steps to draw conclusions from data. The overall process of finding and interpreting patterns and models from data involves the repeated application of the following steps[2]:
1. Understand the application domain, the relevant previous knowledge and the goals of the end-user (formulate the hypothesis).
2. **Data Collection:** Determining how to find and extract the right data for modeling. First, we need to identify the different data sources are available. Data may be scattered in different spreadsheets, files, and hard-copy (paper) lists.
3. **Data integration:** Integration of multiple data cubes, databases or files. A big part of the integration activity is to build a data map, which expresses how each data element in each data set must be prepared to express it in a common format and record structure.
4. **Data selection:** First of all the data are collected and integrated from all the various sources, and we select only the data which useful for data mining. Only relevant information is selected.
5.**Pre-processing:**The Major Tasks in Data Pre-processing are: Cleaning, Transformation and Reduction.

- **Data cleaning:** Also called data cleansing. It deals with errors detecting and removing from data in order to improve the quality of data. Data cleaning usually includes fill in missing values and identify or remove outliers.

- **Data Transformation:** Data transformation operations are additional procedures

of data pre-processing that would contribute toward the success of the mining process and improve data-mining results. Some of Data transformation techniques are Normalization, Differences and ratios and Smoothing.

- **Data Reduction:** For large data sets, there is an increased likelihood that an intermediate, data reduction step should be performed prior to applying data-mining techniques. While large datasets have potential for better mining results, there is no guarantee that they will produce better knowledge than small datasets. Data reduction obtains a reduced dataset representation that is much smaller in volume, yet produces the same analytical results.

6. **Building the model:** In this step we choose and implement the appropriate data-mining task (ex. association rules, sequential pattern discovery, classification, regression, clustering, etc.), the data mining technique and the data mining algorithm(s) to build the model.

7. **Interpretation of the discovered knowledge (model /patterns):** The interpretation n of the detected pattern or model reveals whether or not the patterns are interesting. This step is also called Model Validation/Verification and uses it to represent the result in a suitable way so it can be examined thoroughly.

8. **Decisions / Use of Discovered Knowledge:** It helps to make use of the knowledge gained to take better decisions.

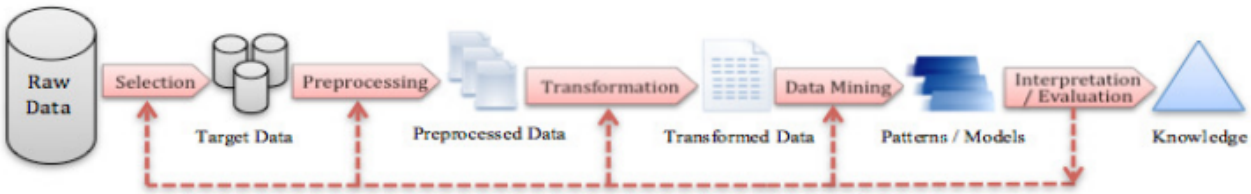

Figure 1.1: Datamining Process

1.1.3 Educational Data Mining

There are increasing research interests in using the data mining techniques in education. This emerging field called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments. The data can be collected from operational and historical data reside in the databases of educational institutes. The student data can be academic or personal, and it can be collected from e-learning systems. Educational Data Mining uses different techniques such as Decision Trees, Nave Bayes, Neural Networks, K- Nearest neighbor, and many others.

The knowledge is hidden among the educational data set and it is extractable through data mining techniques. The EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. Knowledge discovered by EDM algorithms can be used to help teachers to manage their classes, understand their students learning processes and reflect on their own teaching methods. Evaluation is a systematic process of collecting, analysing and interpreting evidences of students progress and achievement in cognitive areas of learning for the purpose of taking a variety of decisions[3][1].

1.2 Problem Definition

Measuring of academic performance of the student is challenging task. This academic performance depends on factors like personal, socio economics, physiological, environmental variables and educational history. For this we apply various datamining steps such as datapreprocessing, datamining techniques such as classification, association and clustering. Using varios datamining algorithms, pattern evaluation, knowledge representation and performance prediction are applied on collective database of above mentioned factors. This is more useful for identifying weak students and the identified students can be individually assisted by the educator so that their performance is better in future.

1.3 Scope

The scope of this system is to identify the factors influencing the performance of students in final examinations and find out a suitable data mining algorithm to predict the grade of students so as to a give timely and an appropriate warning to students those who are at risk. In the present investigation, a survey cum experimental methodology was adopted to generate a database and it was constructed from a primary and a secondary source. The obtained results from hypothesis testing reveals that type of school is not influence student performance and parents' occupation plays a major role in predicting grades. This work will help the educational institutions to identify the students who are at risk and to and provide better additional training for the weak students.

1.4 Objectives

The main objectives are[3]:

- Identication of highly inuencing predictive variables on the academic performance of students.
- Find how Student can improve his/her Performance.
- Predict the grade at academic examination.

1.5 Summary

In this chapter, an introduction of the project topic with the background and problem denition has been discussed. Scope and objective of the project has been also discussed in this chapter. Next chapter presents Literature survey of the project.

Chapter 2

System Analysis

Systems analysis is the study of sets of interacting entities, including computer systems analysis. The development of a computer-based information system includes a systems analysis phase which produces or enhances the data model which itself is a precursor to creating or enhancing a database. There are a number of different approaches to system analysis. When a computer-based information system is developed, systems analysis would constitute the development of a feasibility study, involving determining whether a project is economically, socially, technologically and organizationally feasible.

The sections of the chapter are organized as follows. Section 2.1 presents literature survey. Proposed system of the project is discussed in section 2.2. Section 2.3 describes feasibility study. Risk analysis of the project is presented in section 2.4. Section 2.5 presents Project scheduling. Effort allocation topic is covered in section 2.6. Finally, summary of the chapter is given in the last section.

2.1 Literature Survey

- Paris et al. compared the data mining methods accuracy to classifying students in order to predicting class grade of a student. These predictions are more useful for identifying the weak students and assisting administration to take remedial measures at initial stages to produce excellent graduate that will graduate at least with the second class upper.

- Rathee and Mathur apply ID3, C4.5 and CART decision tree algorithms on the educational data for predicting a students performance in the examination. All the algorithms are applied on the internal assessment data of the student to predict their academic performance in the final exam. The efficiency of various decision tree algorithms can be analyzed based on their accuracy and time taken to derive the tree. The

predictions obtained from the system have helped the tutor to identify the weak students and improve their performance. C4.5 is the best algorithm among all the three because it provides better accuracy and efficiency than the other algorithms[2][4].

- Kortemeyer and Punch applied data mining classifiers as a means of comparing and analyzing students' use and performance who have taken a technical course via the web. The results show that combining multiple classifiers leads to a significant accuracy improvement in a given data set. Prediction performance of combining classifiers is often better than a single classifier because the decision is relying on the combined output of several models[4].

- Zhu et al. proposed SAMME algorithm using a multiclass exponential loss function. The numerical experiments have indicated that AdaBoost.MH in general performs very well, and SAMMEs performance is comparable with that of the AdaBoost.MH, and sometimes slightly better. Also, they discussed the computational cost of SAMME. SAMME generates only one Kclass classifier in each iteration; thus, it is K times faster than AdaBoost.MH[3].

- Bakar et al. proposed an agent based data classification approach, and it is based on creating agent within the main classification process. They show the use of agent within the classification theory, which will help to improve classification speed, and maintain the quality of knowledge. The proposed agents are embedded within the standard rule application techniques. The result shows the significant improvements in classification time and the number of matched rules with comparable classification accuracy[3][4].

2.2 Proposed System

Universities are confronted with a severe competition among each other, trying to attract the most appropriate students who will successfully pass through the university educational process, and making eorts to cope with student retention. University management is very often forced to take quickly important decisions, and therefore timely and high quality information is needed. This new emerging eld, called Educational Data Mining (EDM), concerned with developing methods that extract knowledge from data come from the educational context.The data can be collected from historical and operational data reside in the databases of educational institutes. The student data can be an academic. The preprocess is carried out on the information collected for deletion of unwanted information. Based on the rule student performance is being predicted[4].

2.3 Feasibility Study

A Feasibility study is an evaluation and analysis of the potential of the proposed project which is based on extensive investigation and research to give full comfort to the decisions makers. Feasibility studies aim to objectively and rationally uncover the strengths and weaknesses of the existing business or proposed venture, opportunities and threats as presented by the environment, the resources required to carry through, and ultimately the prospects for success[3].

- Economical Feasibility In our project, Asp.net language is easy to used and easy to learn. it is easily available. thats why this system is economically beneficial to anyone.
- Operational Feasibility Our project do not cause any harmful impact on the environment. So it is operationally feasible.
- Technical Feasibility ASP.net required for project is easily available. We have used sql server as back end which is easily available So the project technically feasible.

2.4 Risk Analysis

Risk analysis and management are a series of steps that help a software team to understand and manage uncertainty. Many problems can plague a software project. A risk is a potential problem it might happen, it might not. But, regardless of the outcome, its a really good idea to identify it, assess its probability of occurrence, estimate its impact, and establish a contingency plan should the problem actually occur. Everyone involved in the software process managers, software engineers, and customers participate in risk analysis and management. Before embarking on the project it is necessary to review all of the risks that might be involved in it. These risks have been documented before the coding of the project started.

The majority of the risk components lie under the categories[3].

- Project Risks
- Business Risks
- Technical Risks

Project Risks	Yes	No
Will the project meets the requirement for what it is intended to do?	Yes	
Business Risks		
Will the project satisfy the business needs in the organization?	Yes	
· Technical Risks		
Will the project technically support all requirements?	Yes	

2.5 Effort Allocation

Identification of project, requirements gathering and study of existing system accounts 10 percent of effort. 15 percent of effort is normally applied to data modeling and coding. Identification of functional and non-functional requirements, testing accounts 5 percent of project effort. Designing requires 30 percent of effort.

2.6 Summary

In this chapter, literature survey, proposed system and feasibility study of the project has been described. Risk analysis and effort allocation has been also presented in this chapter. Next chapter presents system requirement specifications.

Chapter 3

System Requirement Specification

Understanding user requirements is an integral part of information systems design and is critical to the success of interactive systems. It is now widely understood that successful systems and products begin with an understanding of the needs and requirements of the users. As specified in the ISO 13407 standard (ISO, 1999), user-centered design begins with a thorough understanding of the needs and requirements of the users. The benefits can include increased productivity, enhanced quality of work, reductions in support and training costs, and improved user satisfaction.

The sections of the chapter are organized as follows. Section 3.1 presents hardware requirements. Software requirements of the project is discussed in section 3.2. Section 3.3 describes Functional requirements. Non-Functional requirements of the project is presented in section 3.4. Finally, summary of the chapter is given in the last section.

3.1 Hardware and Software Requirements

- *Hardware Requirements*

Hardware requirements are given as follows[4]:

- Processor: Intel Core i3(Any Generation).
- RAM: 1 GB.
- Hard Disk Drive: 1 TB.

- *Software Requirements*

Software requirements are given as follows[4]:

- OS: Windows
- Microsoft Visual Studio 2015

- Microsoft .NET Framework 4.6
- Microsoft SQL Server 2015

3.2 Functional Requirements

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that dene what a system is supposed to accomplish[4].

1. **Admin:**
   The Admin is the one who handle all the information related to student, subject or teacher. Admin has all the permissions to rome in the system in any module at any time for any task. Admin can create new account of student or teacher. Delete account of student or teacher. Admin can upload the final Examination Result to the system. Admin maintains the system properly like checks the connectivity between client and server. Admin can generate the overall report which shows the graph of good students, average students, poor students.

2. **Teacher:**
   Teacher is another important person of the system. There may be more than one teacher in this system. Teacher fills the attendance, ISE marks and audit points of the student which helps in the prediction of students performance.Teacher can also generate the report which shows subjectwise report.

3. **Student:** Student has no authority to made any change in the system. Student can fill Personal details. Student can only view its performance report. Students can improve its performance by suggestions given by system.

4. **Databse:**Database contains encrypted data, data from student,teacher and data from admin.

3.3 Non Functional Requirements

A non-functional requirement is a requirement that species criteria that can be used to judge the operation of a system, rather than specific behaviors. This should be contrasted with functional requirements that dene specific behavior or functions. The plan for implementing functional requirements is detailed in the system design. The plan for implementing nonfunctional requirements is detailed in the system architecture.

3.4 Summary

In this chapter, Hardware and software requirements, functional and nonfunctional requirements are discussed in brief. The next chapter discuss about the system design and architecture of the system.

Chapter 4

System Design and Architecture

System design provides the understanding and procedural details necessary for implementing the system recommended in the system study. Design is a meaningful engineering representation of something that is to be built. It can be traced to a customers requirements and at the same time assessed for quality against a set of predefined criteria for good design. In the software engineering context, design focuses on four major areas of concern: data, architecture, interfaces and components.

The sections of the chapter are organized as follows. Section 3.1 presents system architecture. System Design is presented in Section 3.2. The Summary is presented in the last section of the chapter.

4.1 Conceptual Model and Architecture

This system identify the factors which affects on the students performance and helps students to understand their learning status through the predicted performance using the performance prediction model and students current learning usage data and views this performance to students with suitable messages to motivate and regulate their own behavior. The teacher also can view the predicted students' performance and help earlier in identifying weak students who need special attention and take the required procedure, such as sending an suggestion[3][2].

The first step is historical students data selection and pre-processing. The database has been considered as 3/3 initially in which 2/3 of the data after pre-processing has been used for training data to build the performance model and the remaining 1/3 data is used for testing purpose to evaluate the model.

Also, it selects and pre-processes the current students data to predicts the current students data and store the results with the current students' data and the best rules for performance optimization in the **Students Performance Information DB** to be ready for **Performance prediction Agent** to predict the current students' performance to view

their current progress.

Rules Generator: Traverses each path of the highest weight tree model to generate the tree rules and store it with its performance class in **Model and Rules DB**.

Model and Rules DB now contains the Model and the generated rules. There is no need to run the complete process of model building and evaluation again, and the system will provide the Model and Rules to **Performance Prediction Agent** directly to predict the current students performance.

Messages DB: Contains Adaptive Messages according to students learning status and current performance and it can affect students learning behaviors and achievements.

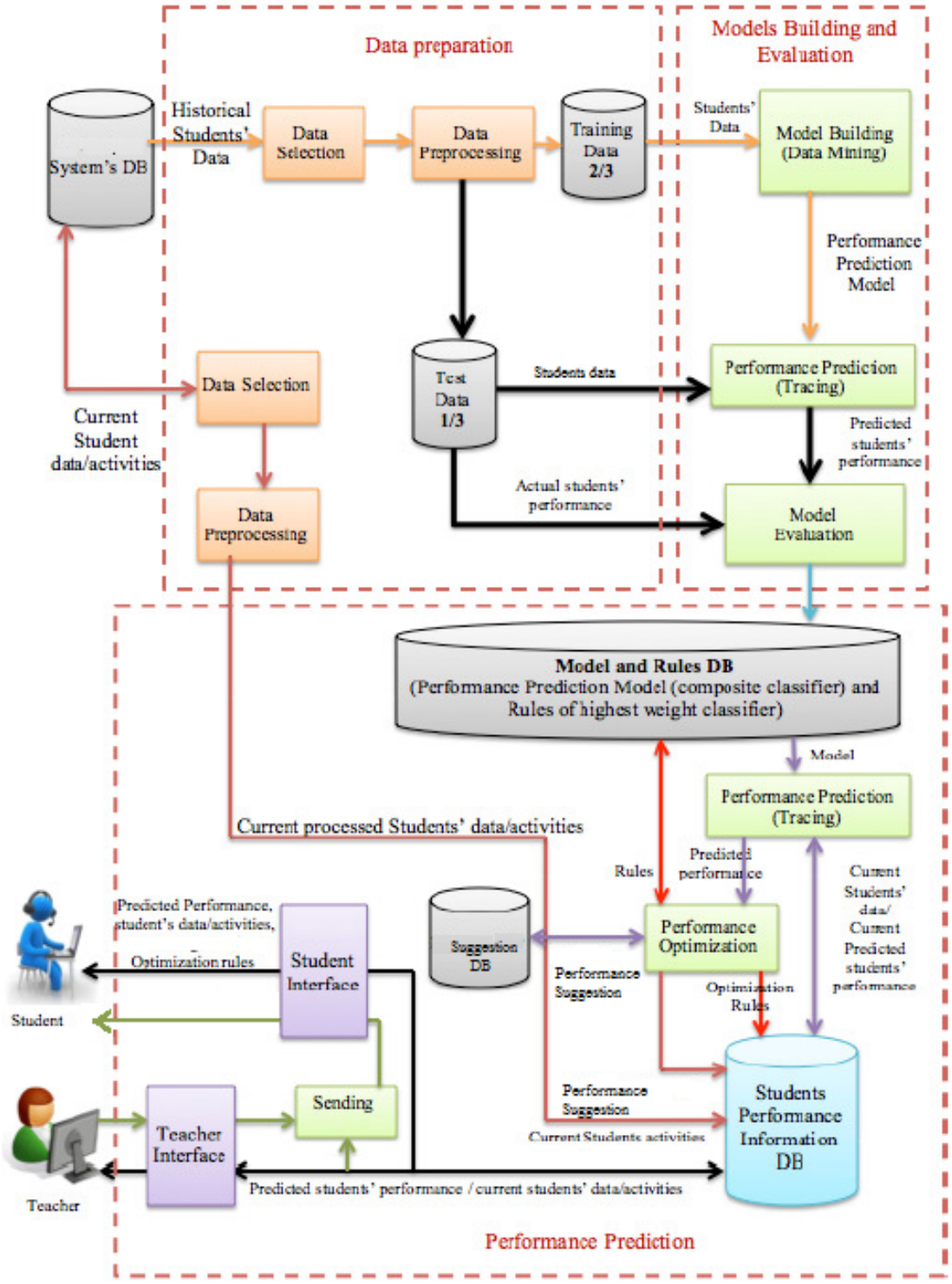

Figure 4.1: Architecture of System

4.2 System Design

The UML Diagrams for the proposed system are given below:

- *Usecase Diagram:*

Figure 4.2: Usecase Diagram

- *Sequence Diagram:*

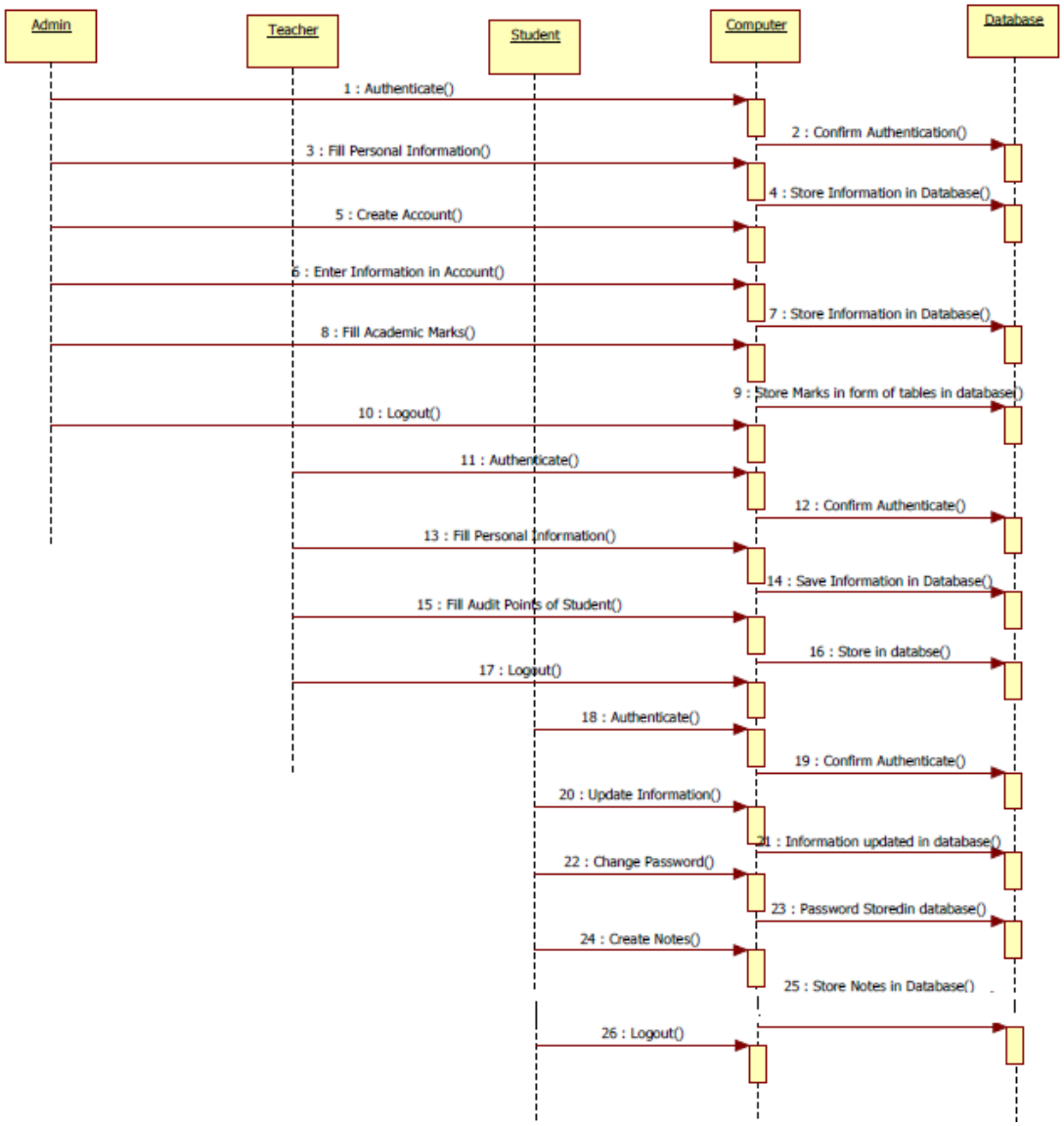

Figure 4.3: Usecase Diagram

- *Class Diagram:*

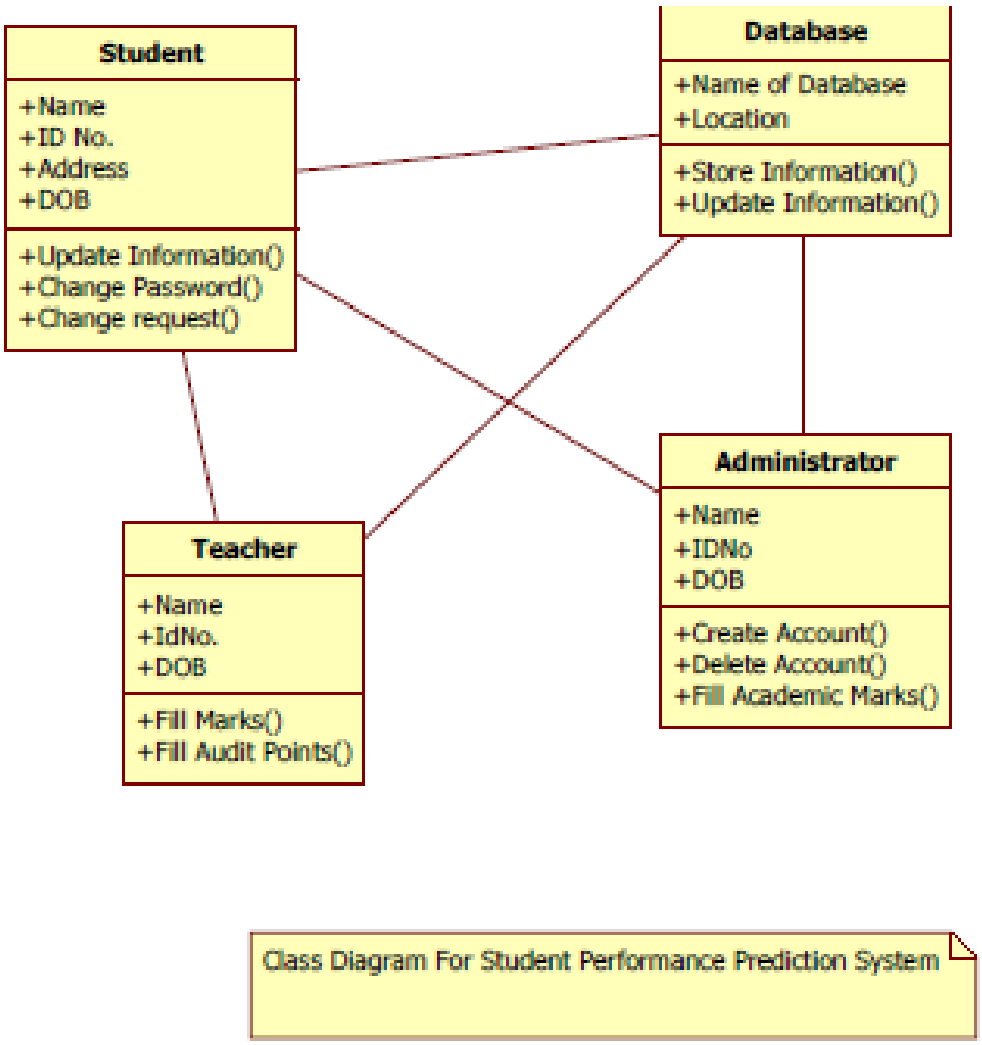

Figure 4.4: Class Diagram

- *Statechart Diagram:*

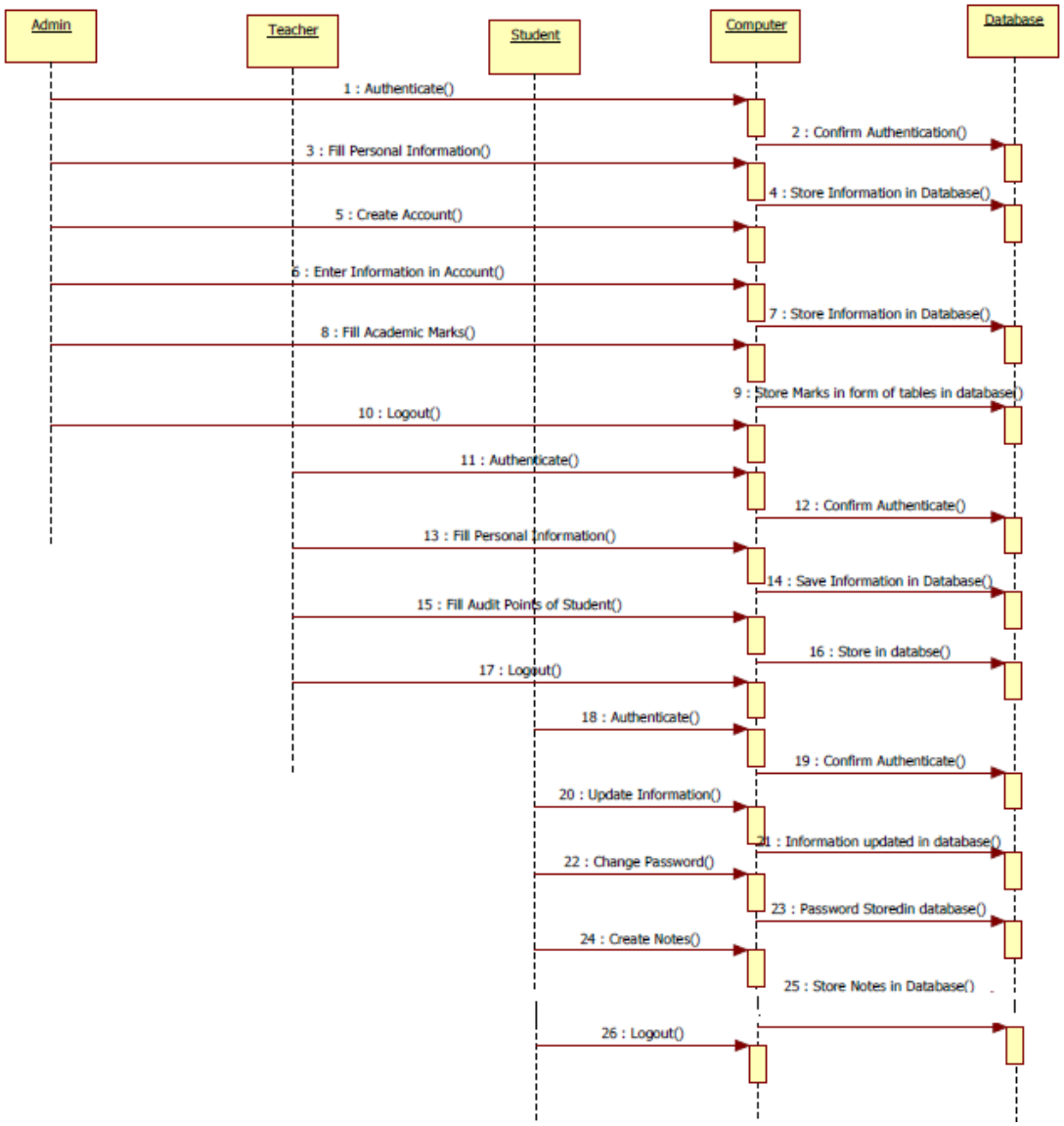

Figure 4.5: Statechart Diagram

4.3 Summary

In this chapter, the Conceptual Model and Architecture is described in brief and the UML diagrams are presented. In next chapter the Conclusion and Future Work is presented.

Chapter 5

Conclusion and Future Work

It showed how useful datamining can be used in higher education particularly to improve graduate students performance.

The experiment can be extended with more distinctive attributes to get more accurate results, useful to improve students learning outcomes.

Bibliography

[1] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," *International Journal of Computer Science and Management Research*, vol. 1, 2012.

[2] K. V. J.K. Jothi Kalpana, "Intellectual performance analysis of students by using data mining techniques," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, 2014.

[3] V. Ramesh, "Predicting student performance: A statistical and data mining approach," *International Journal of Computer Applications*, vol. 63, no. 8, 2013.

[4] D. A. M. Dr. Abdullah AL-Malaise and M. Alkhozae, "Students performance prediction system using multi agent data mining technique," *International Journal of Data Mining and Knowledge Management Process*, vol. 4, no. 5, 2014.