Mining Educational Data to Analyze Students‟ Performance

Abstract

The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students’ performance and so on. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Present paper is designed to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. In this research, the classification task is used to evaluate student’s performance and as there are many approaches that are used for data classification, the decision tree method is used here.

By this task we extract knowledge that describes students’ performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counselling.

Introduction

The advent of information technology in various fields has lead the large volumes of data storage in various formats like records, files, documents, images, sound, videos, scientific data and many new data formats. The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations. Data mining techniques have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others.

Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about student’s performance and so on.

The main objective of this paper is to use data mining methodologies to study students‟ performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate student’s performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information’s like Attendance, Class test, Seminar and Assignment marks were collected from the student’s management system, to predict the performance at the end of the semester. This paper investigates the accuracy of Decision tree techniques for predicting student performance.

Literature Survey

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.

Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students. Mining in educational environment is called Educational Data Mining.

Han and Kamber describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Pandey and Pal conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer students will performer or not.

Hijazi and Naqvi conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother’s education and student’s family income were highly correlated with the student academic performance.

The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general.

Galit gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Pandey and Pal conducted study on the student performance based by selecting 60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India. By means of association rule they find the interestingness of student in opting class teaching language.

Ayesha, Mustafa, Sattar and Khan describes the use of k-means clustering algorithm to predict student’s learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

Bray in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socio-economic conditions.

Bhardwaj and Pal conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attribute, it was found that the factors like students‟ grade in senior secondary exam, living location, medium of teaching, mother’s qualification, students other habit, family annual income and student’s family status were highly correlated with the student academic performance.

Proposed System

In present day’s educational system, a student’s performance is determined by the internal assessment and end semester examination. The internal assessment is carried out by the teacher based upon students‟ performance in educational activities such as class test, seminar, assignments, general proficiency, attendance and lab work. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination.

*A. Data Preparations*

The data set used in this study was obtained from college database Initially size of the data is 50. In this step data stored in different tables was joined in a single table after joining process errors were removed.

*B. Data selection and transformation*

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table I for reference.

****

The domain values for some of the variables were defined for the present investigation as follows:

1.  **PSM –** Previous Semester Marks/Grade obtained in MCA course. It is split into five class values: *First – >60%, Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%*.
2.  **CTG –** Class test grade obtained. Here in each semester two class tests are conducted and average of two class test are used to calculate sessional marks. CTG is split into three classes: *Poor – < 40%, Average – > 40% and < 60%, Good –>60%.*
3.  **SEM –** Seminar Performance obtained. In each semester seminar are organized to check the performance of students. Seminar performance is evaluated into three classes: *Poor – Presentation and communication skill is low, Average – Either presentation is fine or Communication skill is fine, Good – Both presentation and Communication skill is fine.*
4.  **ASS –** Assignment performance. In each semester two assignments are given to students by each teacher. Assignment performance is divided into two classes: *Yes – student submitted assignment, No – Student not submitted assignment.*
5.  **GP -** General Proficiency performance. Like seminar, in each semester general proficiency tests are organized. General Proficiency test is divided into two classes: *Yes – student participated in general proficiency, No – Student not participated in general proficiency.*
6.  **ATT –** Attendance of Student. Minimum 70% attendance is compulsory to participate in End Semester Examination. But even through in special cases low attendance students also participate in End Semester Examination on genuine reason. Attendance is divided
7. into three classes: *Poor - <60%, Average - > 60% and <80%, Good - >80%.*
8.  **LW –** Lab Work. Lab work is divided into two classes: *Yes – student completed lab work, No – student not completed lab work*.
9.  **ESM -** End semester Marks obtained in MCA semester and it is declared as response variable. It is split into five class values: First *– >60% , Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%*.

*C. Decision Tree*

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

The three widely used decision tree learning algorithms are: ID3, ASSISTANT and C4.5.

*D. The ID3 Decision Tree*

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric - information gain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

*E. Measuring Impurity*

Given a data table that contains attributes and class of the attributes, we can measure homogeneity (or heterogeneity) of the table based on the classes. We say a table is pure or homogenous if it contains only a single class. If a data table contains several classes, then we say that the table is impure or heterogeneous. There are several indices to measure degree of impurity quantitatively. Most well known indices to measure degree of impurity are entropy, gini index, and classification error.



Entropy of a pure table (consist of single class) is zero because the probability is 1 and log (1) = 0. Entropy reaches maximum value when all classes in the table have equal probability.



Gini index of a pure table consist of single class is zero because the probability is 1 and 1-12 = 0. Similar to Entropy, Gini index also reaches maximum value when all classes in the table have equal probability.



Similar to Entropy and Gini Index, Classification error index of a pure table (consist of single class) is zero because the probability is 1 and 1-max (1) = 0. The value of classification error index is always between 0 and 1. In fact the maximum Gini index for a given number of classes is always equal to the maximum of classification error index because for a number of classes n, we set probability is equal to p=1/n and maximum Gini index happens at



while maximum classification error index also happens at



*F. Splitting Criteria*

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S, is defined as



Where *Values (A)* is the set of all possible values for attribute *A*, and *Sv* is the subset of *S* for which attribute *A* has value *v*. The first term in the equation for *Gain* is just the entropy of the original collection *S* and the second term is the expected value of the entropy after S is partitioned using attribute *A*. The expected entropy described by this second term is simply the sum of the entropies of each subset , weighted by the fraction of examples that belong to *Gain (S, A)* is therefore the expected reduction in entropy caused by knowing the value of attribute *A.*



And



The process of selecting a new attribute and partitioning the training examples is now repeated for each non terminal descendant node. Attributes that have been incorporated higher in the tree are excluded, so that any given attribute can appear at most once along any path through the tree. This process continues for each new leaf node until either of two conditions is met:

1. 1. Every attribute has already been included along this path through the tree, or
2. 2. The training examples associated with this leaf node all have the same target attribute value (i.e., their entropy is zero).

*G. The ID3Algoritm*

ID3 (Examples, Target\_Attribute, Attributes)

 Create a root node for the tree

 If all examples are positive, Return the single-node tree Root, with label = +.

 If all examples are negative, Return the single-node tree Root, with label = -.

 If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.

 Otherwise Begin

o A = The Attribute that best classifies examples.

o Decision Tree attribute for Root = A.

o For each possible value, *vi*, of A,

 Add a new tree branch below Root, corresponding to the test A = *vi*.

 Let Examples(*vi*) be the subset of examples that have the value *vi* for A

 If Examples(*vi*) is empty

 Then below this new branch add a leaf node with label = most common target value in the examples

 Else below this new branch add the subtree ID3 (Examples(*vi*), Target\_Attribute, Attributes – {A})

 End

 Return Root

Modules Of the Project

1. Admin Login :-In this module admin login into the application using username and password . If the login details are authenticated then admin get access to the application.
2. Manage Students:- In this Module admin manages students details like personal information, contact details, marks etc.
3. Data Selection and Transformation:- In this module admin selects the list of students whose information is use for the decision making. Those students data is arrange in required format.
4. Apply ID3 Decision tree Algorithm:- this algorithm is use to creates clusters. Clusters will be form based on the conditions specify during execution.
5. Measuring Impurity:- Given a data table that contains attributes and class of the attributes, we can measure homogeneity (or heterogeneity) of the table based on the classes. We say a table is pure or homogenous if it contains only a single class. If a data table contains several classes, then we say that the table is impure or heterogeneous. There are several indices to measure degree of impurity quantitatively.
6. Final Prediction Result:- The result will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination.
7. Conclusion

In this paper, the classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information’s like Attendance, Class test, Seminar and Assignment marks were collected from the student’s previous database, to predict the performance at the end of the semester.

This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination.

**Hardware and software requirements**

HARDWARE REQUIREMENTS:

* 1 GB RAM.
* 200 GB HDD.
* Intel 1.66 GHz Processor Pentium 4

SOFTWARE REQUIREMENTS:

* Windows XP, Windows 7,8
* Visual Studio 2010
* MS SQL Server 2008
* Windows Operating System

**Testing Technology**

System testing is a critical phase implementation. Testing of the system involves hardware devise and debugging of the computer programs and testing information processing procedures. Testing can be done with text data, which attempts to stimulate all possible conditions that may arise during processing. If structured programming Methodologies have been adopted during coding the testing proceeds from higher level to lower level of program module until the entire program is tested as unit. The testing methods adopted during the testing of the system were unit testing and integrated testing.

**UNIT TESTING:**

Unit testing focuses first on the modules, independently of one another, to locate errors. This enables the tester to detect errors in coding and logical errors that is contained within that module alone. Those resulting from the interaction between modules are initially avoided.

**INTEGRATION TESTING:**

Integration testing is a systematic technique for constructing the program structure while at the same time to uncover the errors associated with interfacing. The objective is to take unit-tested module and build a program structure that has been detected by designing. It also tests to find the discrepancies between the system and its original objectives. Subordinate stubs are replaced one at time actual module. Tests were conducted at each module was integrated. On completion of each set another stub was replaced with the real module.

**FUNCTIONAL TESTING:**

Functional testing is a technique in which all the functionalities of the program are tested to check whether all the functions that where proposed during the planning phase are full filled.

This is also to check that if all the functions proposed are working properly.

This is further done in two phases:

* One before the integration to see if all the unit components work properly
* Second to see if they still work properly after they have been integrated to check if some functional compatibility issues arise.

**PERFORMANCE TESTING:**

**Expected Result**

* The client should be able to connect to the server properly without any problems.
* The connection establishment between the mobile device and the server should take minimal time.
* The mobile device should be able receive data from the server uninterruptedly.
* Information provided by the application should be correct and as per the user’s need.
* **Observation**
* Connection can be established easily provided that the server is on.
* The connection with the server takes time as it uses Internet connection.
* Receiving data from the server takes time.
* Information coming from the database is correct.

**LOAD / STRESS TESTING :**

* **Expected Result**
  + Response time should be unaffected irrespective of the no of users.
  + The introduction of the newer clients should not make the server to work hap hazardously.
  + Continuous use of the server by different clients should not result into the server getting slowed down.
  + Response time should not be degraded if there is congestion in network.
* **Observation**
* The speed of transmission was fine even when the newer clients were getting added. The response of the server was satisfying even with the introduction of newer client.

**ASP.NET**

ASP.NET is more than the next version of Active Server Pages (ASP); it is a unified Web development platform that provides the services necessary for developers to build enterprise-class Web applications. While ASP.NET is largely syntax-compatible with ASP, it also provides a new programming model and infrastructure that enables a powerful new class of applications. You can migrate your existing ASP applications by incrementally adding ASP.NET functionality to them.

ASP.NET is a compiled .NET Framework -based environment.  You can author applications in any .NET Framework compatible language, including Visual Basic and Visual C#.  Additionally, the entire .NET Framework platform is available to any ASP.NET application. Developers can easily access the benefits of the .NET Framework, which include a fully managed, protected, and feature-rich application execution environment, simplified development and deployment, and seamless integration with a wide variety of languages.

**2.3 IDENTIFICATION OF THE PROJECT FROM THE STUDY**

**FEASIBILITY STUDY :**

The very first phase in any system developing life cycle is preliminary investigation. The feasibility study is a major part of this phase. A measure of how beneficial or practical the development of any information system would be to the organization is the feasibility study.

The feasibility of the development software can be studied in terms of the following aspects:

1.Operational Feasibility.

2.Technical Feasibility.

3.Economical feasibility.

4.Motivational Feasibility.

5.Legal Feasibility

**OPERATIONAL FEASIBILITY :**

The site will reduce the time consumed to maintain manual records and is not tiresome and cumbersome to maintain the records. Hence operational feasibility is assured.

**TECHNICAL FEASIBILITY :**

* At least 166 MHz Pentium Processor or Intel compatible processor.
* At least 512 MB RAM.
* 14.4 kbps or higher modem.
* A mouse or other pointing device.
* At least 50 GB free hard disk space.
* Microsoft Internet Explorer 4.0 or higher.

**ECONOMICAL FEASIBILTY :**

Once the hardware and software requirements get fulfilled, there is no need for the user of our system to spend for any additional overhead.

For the user, the web site will be economically feasible in the following aspects:

* The web site will reduce a lot of paper work. Hence the cost will be reduced.
* Our web site will reduce the time that is wasted in manual processes.
* The storage and handling problems of the registers will be solved.

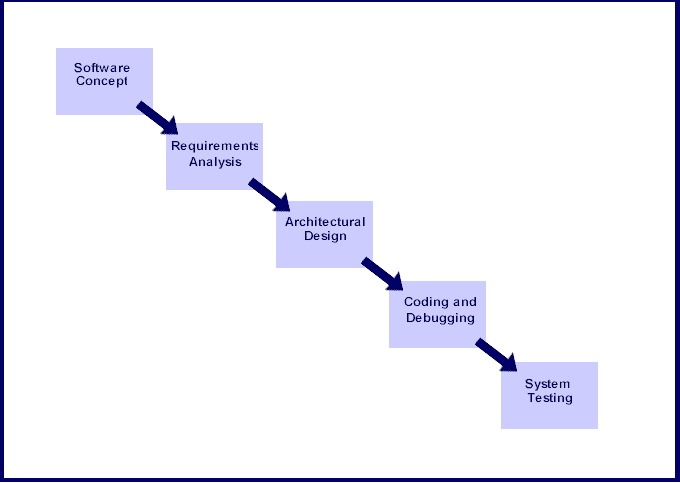
**LEGAL FEASIBILITY :**

The licensed copy of the required software is quite cheap and easy to get. So from legal point of view the proposed system is legally feasible.

**Steps Towards Implementation**

**System Development Life Cycle:**

The System Development Life Cycle is the process of developing information systems through investigation, analysis, design, implementation, and maintenance.  The System Development Life Cycle (SDLC) is also known as Information Systems Development or Application Development.



**🡺 Steps involved in the System Development Life Cycle :**

Below are the steps involved in the System Development Life Cycle.  Each phase within the overall cycle may be made up of several steps.

**Step 1:** **Software Concept**

The first step is to identify a need for the new system.  This will include determining whether a business problem or opportunity exists, conducting a feasibility study to determine if the proposed solution is cost effective, and developing a project plan.

This process may involve end users who come up with an idea for improving their work. Ideally, the process occurs in tandem with a review of the organization's strategic plan to ensure that IT is being used to help the organization achieve its strategic objectives.  Management may need to approve concept ideas before any money is budgeted for its development.

**Step 2:** **Requirements Analysis**

Requirements analysis is the process of analyzing the information needs of the end users, the organizational environment, and any system presently being used, developing the functional requirements of a system that can meet the needs of the users.  Also, the requirements should be recorded in a document, email, user interface storyboard, executable prototype, or some other form.  The requirements documentation should be referred to throughout the rest of the system development process to ensure the developing project aligns with user needs and requirements.

Professionals must involve end users in this process to ensure that the new system will function adequately and meets their needs and expectations.

**Step 3:** **Architectural Design**

After the requirements have been determined, the necessary specifications for the hardware, software, people, and data resources, and the information products that will satisfy the functional requirements of the proposed system

can be determined.  The design will serve as a blueprint for the system and helps detect problems before these errors or problems are built into the final system. Professionals create the system design, but must review their work with the users to ensure the design meets users' needs.

**Step 4:** **Coding and Debugging** Coding and debugging is the act of creating the final system.  This step is done by software developer.

**Step 5:** **System Testing** The system must be tested to evaluate its actual functionality in relation to expected or intended functionality.  Some other issues to consider during this stage would be converting old data into the new system and training employees to use the new system.  End users will be key in determining whether the developed system meets the intended requirements, and the extent to which the system is actually used.

**Step 6:** **Maintenance**

Inevitably the system will need maintenance. Software will definitely undergo change once it is delivered to the customer. There are many reasons for the change. Change could happen because of some unexpected input values into the system. In addition, the changes in the system could directly affect the software operations. The software should be developed to accommodate changes that could happen during the post implementation period.

There are various software process models like:-

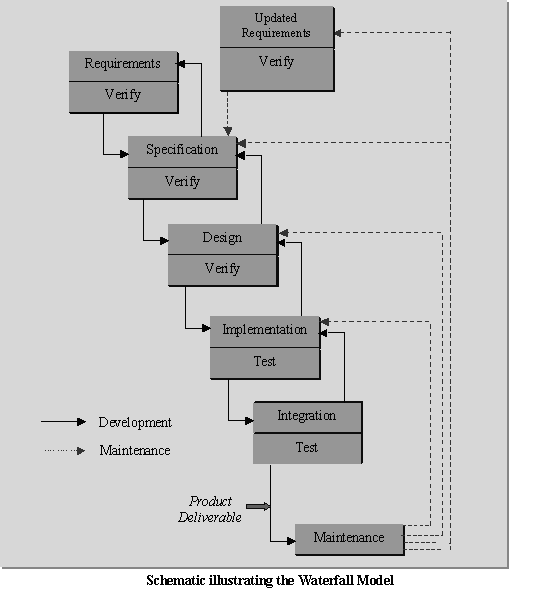
* Prototyping Model
* RAD Model
* The Spiral Model
* The Waterfall Model
* The Iterative Model

Of all these process models we’ve used the Iterative model(The Linear Sequential Model) for the development of our project.

**The Iterative model**

The waterfall model derives its name due to the cascading effect from one phase to the other as is illustrated in Figure1.1. In this model each phase well defined starting and ending point, with identifiable deliveries to the next phase.

This model is sometimes referred to as the linear sequential model or the software life cycle.



The model consists of six distinct stages, namely:

1.      In the **requirements analysis** phase

(a)    The problem is specified along with the desired service objectives (goals)

(b)   The constraints are identified

2.      In the **specification phase** the system specification is produced from the detailed definitions of (a) and (b) above. This document should clearly define the product function.

3.      In the system and software **design phase**, the system specifications are translated into a software representation. The software engineer at this stage is concerned with:

        Data structure

        Software architecture

        Algorithmic detail

        Interface representations

The hardware requirements are also determined at this stage along with a picture of the overall system architecture. By the end of this stage should the software engineer should be able to identify the relationship between the hardware, software and the associated interfaces. Any faults in the specification should ideally not be passed ‘down stream.

4.      In the **implementation and testing**phase stage the designs are translated into the software domain

        Detailed documentation from the design phase can significantly reduce the coding effort.

        Testing at this stage focuses on making sure that any errors are identified and that the software meets its required specification.

5.      In the **integration and system testing** phase all the program units are integrated and tested to ensure that the complete system meets the software requirements. After this stage the software is delivered to the customer [**Deliverable – The software product is delivered to the client for acceptance testing**.]

6.      The **maintenance** phase the usually the longest stage of the software. In this phase the software is updated to:

        Meet the changing customer needs

        Adapted to accommodate changes in the external environment

        Correct errors and oversights previously undetected in the testing phases

        Enhancing the efficiency of the software

Observe that feed back loops allow for corrections to be incorporated into the model. For example a problem/update in the design phase requires a ‘revisit’ to the specifications phase. When changes are made at any phase, the relevant documentation should be updated to reflect that change.

**Advantages of the Iterative Model:-**

        Testing is inherent to every phase of the Iterative model

        It is an enforced disciplined approach

       It is documentation driven, that is, documentation is produced at every stage

**Disadvantages of the Iterative Model:-**

The waterfall model is the oldest and the most widely used paradigm. However, many projects rarely follow its sequential flow. This is due to the inherent problems associated with its rigid format. Namely:

        It only incorporates iteration indirectly, thus changes may cause considerable confusion as the project progresses.

        As The client usually only has a vague idea of exactly what is required from the software product, this IM has difficulty accommodating the natural uncertainty that exists at the beginning of the project.

        The customer only sees a working version of the product after it has been coded. This may result in disaster any undetected problems are precipitated to this stage.

**4.User Manual**

**.net Framework:**

The .NET Framework is Microsoft's Managed Code programming model for building applications on Windows clients, servers, and mobile or embedded devices. Microsoft's .NET Framework is a software technology that is available with several Microsoft Windows operating systems. In the following sections describes , the basics of Microsoft .Net Frame work Technology and its related programming models.

C# is a language for professional programming. C# (pronounced C sharp) is a programming language designed for building a wide range of enterprise applications that run on the .NET Framework. The goal of C# is to provide a simple, safe, modern, object-oriented, highperformance , robust and durable language for .NET development. Also it enables developers to build solutions for the broadest range of clients, including Web applications, Microsoft Windows Forms-based applications, and thin- and smart-client devices.

**Microsoft SQL Server 2008**

Business today demands a different kind of data management solution. Performance scalability, and reliability are essential, but businesses now expect more from their key IT investment.

SQL Server 2008 exceeds dependability requirements and provides innovative capabilities that increase employee effectiveness, integrate heterogeneous IT ecosystems,and maximize capital and operating budgets. SQL Server 2008 provides the enterprise data management platform your organization needs to adapt quickly in a fast changing environment.

Benchmarked for scalability, speed, and performance, SQL Server 2008 is a fully enterprise-class database product, providing core support for Extensible Markup Language (XML) and Internet queries.

**Easy-to-use Business Intelligence(BI) Tools**

Through rich data analysis and data mining capabilities that integrate with familiar applications such as Microsoft Office, SQL Server 2008 enables you to provide all of your employees with critical, timely business information

tailored to their specific information needs. Every copy of SQL Server 2008 ships with a suite of BI services.

**Self-Tuning and Management Capabilities**

Revolutionary self-tuning and dynamic self-configuring features optimize database performance, while management tools automate standard activities. Graphical tools and performance, wizards simplify setup, database design, and performance monitoring, allowing database administrators to focus on meeting strategic business needs.

**Data Management Application and Services**

Unlike its competitors, SQL Server 2008 provides a powerful and comprehensive data management platform. Every software license includes extensive management and development tools, a powerful extraction, transformation, and loading (ETL) tool, business intelligence and analysis services such as Notification Service. The result is the best overall business value available.

Enterprise Edition includes the complete set of SQL Server data management and analysis features are and is uniquely characterized by several features that makes it the most scalable and available edition of SQL Server 2008 .It scales to the performance levels required to support the largest Web sites, Enterprise Online Transaction Processing (OLTP) system and Data Warehousing systems. Its support for failover clustering also makes it ideal for any mission critical line-of-business application.