

Data Mining Techniques for College Information System

Ravina.H.Popli

Department of Computer Science & Engg
HVPM COET Amravati, Maharashtra, India

Dr. Anjali Raut

Department of Computer Science & Engg
HVPM COET Amravati, Maharashtra, India

Abstract: Data mining is the technique to extract useful and necessary data from large database, data-warehouse or from big data. Data mining is been a rapid research technique in all sectors to extract useful data quickly and fastly from a big database. Every sector is moving in a zone to collect and retrieve data digitally such as educational sector, industrial sector, government sector etc. Decision tree is a graph where branching condition plays a great role. This article is basically designed from a student perspective as to how retrieve data from a medium level database as for example a college database which consist of information and data necessary for a college.

Keywords: Data Mining, Decision tree, Information System.

Introduction:

Data mining is a technique to extract useful information from a large amount of database. Data mining is a term derived by concatenating two terms “data” and “mining”. Data is collection of useful information and mining in its general term is a geographical process to extract useful and valuable mineral from the earth. Thus data mining is a technique to retrieve valuable and useful information from the database. A database is a collection of information that is organized so that it can be easily accessed, managed and updated.

An information system is any organized system for the collection, organization, storage and communication of information. College information system is a collection of data which is required for smooth functioning of a college. In this article we will use decision tree as a data mining technique to extract useful and valuable data from the database storing information related to college.

Literature survey:

Nowadays everything is going “digital” in every sector for example industrial sector, educational sector, government sector, banking sector, health and insurance sector and so on. Previously everything was handwritten including registration, maintenance of details, searching, sorting etc. with these there was several issues like it was a huge hectic for employees to maintain and update the records. These old written technique was made digitalized first by using file processing systems which again has several drawbacks such as data

redundancy and data isolation and updation problem which lead to certain data anomalies. To overcome the drawbacks of file processing system database system were designed. A database is collection of useful and related data. Use of database is the most successful way of storing data in computer. Every sector is moving to digital by storing data in the database. In this paper we are basically focusing on storing data in educational sector basically considering college. A college is a place where higher education is provided to students and having some course period in years. In college data is collected from every student in every year which can be stored in database. Since college includes various sectors such as library, accounting etc data collected is huge and in tremendous amount from the students. there should be a way to extract data from such a huge college database. So to extract useful information from database we use data mining. There are certain techniques to perform data mining and we will discuss how each technique can be used to perform data mining.

College information system:

A college information system is a system for collection, organization, storage and communication of information related to entire college. Every detail is included such as student data, Faculties data, working staff data, accounting details, library details and so on. Almost all data gathered in college is inter-related to each other.

Data Mining:

Data mining is the process to extract useful information and data from the database.

In Data Mining we basically extract useful knowledge from the data base. Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. There are certain steps of extracting knowledge from database such as:

- 1) Learning the application domain.
- 2) Gathering and integration of data
- 3) Cleaning and preprocessing data.
- 4) Reducing and projecting data
- 5) Choosing the functions or techniques of data mining

- 6) Choosing the data mining algorithm
- 7) Perform data mining
- 8) Evaluating and analysis of data
- 9) Use of discovered knowledge.

There are various data mining techniques such as:

- 1) Association
- 2) Clustering
- 3) Classification
- 4) Prediction
- 5) Sequential patterns
- 6) Decision Tree

In this paper we will basically focus on data mining techniques which can be used to perform data mining on college database

- 1) **Association:** Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns.

Association rules can be created for college database such as

- A student who issue the library card also issues books available in library then we can say that when a person issues books from library then he has or will issue library card.
- A student who want to appear for exam will always fill the exam form then we can say that when student fill exam form then he will appear for exam

- 2) **Classification:** Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data items into groups. Association rules can be created for college database such as

- We can apply classification in the application such as for finding those students who got placed in the company so we can create two groups

one as placed students group and second as unplaced student group

- Again we can apply classification in the application for finding the yearly result for the students who got passed so we can again create two groups for passed students and failed students.

- 3) **Clustering:** Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

Clustering technique can be used in college database such as:

- It can be used for book management in the library. In a library, there is a wide range of books on various topics available. The challenge is how to keep those books in a way that readers can take several books on a particular topic without hassle. By using the clustering technique, we can keep books that have some kinds of similarities in one cluster or one shelf and label it with a meaningful name and store in database. If readers want to grab books in that topic, they would only have to go to that shelf instead of looking for the entire library or database of library.

- 4) **Prediction:** The prediction, as its name implied, is one of a data mining techniques that discovers the relationship between independent variables and relationship between dependent and independent variables. Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

Prediction can be used in college information system as

- The prediction analysis technique can be used to predict the students who will be placed in certain companies depending on the overall performance of the students

- It can also used to predict the future carrier choice of student depending on their interest and their growth in particular sector or field
- 5) **Sequential Patterns:** Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period. Often used over longer-term data, sequential patterns are a useful method for identifying trends, or regular occurrences of similar events. Sequential Patterns can be designed for college information system such as
- With certain database from previous year activities commonly performed by students can be recorded and new ideas can be given to students to improve the performance of the students in that activity
 - With certain record from the database students can be given new ideas to change the common way of studying activity

6) **Decision Trees:** Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees can handle both categorical and numerical data. It can be used in conjunction with other data mining techniques Decision tree can be constructed for certain data in college information system such as

- Decision tree for result analysis for students

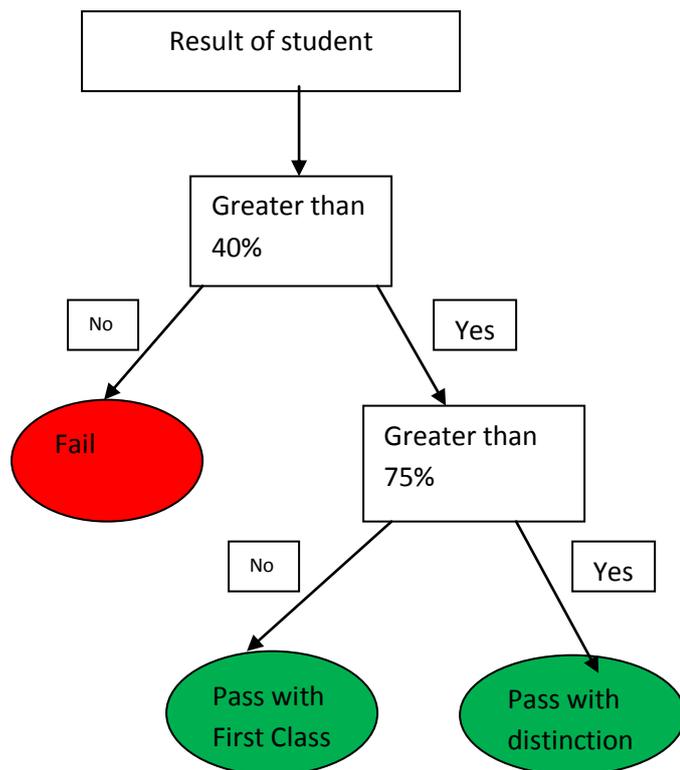


Fig: Result analysis of student using Decision tree

This decision tree is used to analyze the result of the student whether he is fail or pass and also pass with first class or pass with distinction.

Conclusion:

Data mining is the process of extracting useful information from large database. There are various data mining techniques such as Association, Classification, Decision trees, Clustering, Prediction and sequential patterns which can be used and are very useful in extracting data from college information system in various ways described above. Thus by using various data mining techniques we can extract useful information and knowledge from larger databases such as college information system which includes huge data of students, library etc.

References:

- 1) www.wikipedia.com
- 2) www.ibm.com/developerworks/library/ba-data-mining-techniques/
- 3) <http://www.zentut.com/data-mining/data-mining-techniques>
- 4) Paper on data mining techniques and application
- 5) Book of data mining concepts and techniques by Jiawie Han

Author Details:

- 1) **Miss Ravina.H.Popli** student of ME 1st Year (CSE) at HVPM COET Amravati, Maharashtra (India)
- 2) **Dr A.B.Raut** Head of Department(HOD) Department of Computer Science & Engineering HVPM COET Amravati, Maharashtra.