International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Decision Tree Based Information System for Improving School Facilities

J. Hareesh[1], V. Valli Kumari[2]

[1] M-Tech, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, Andhra Pradesh,

[2] Professor, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, Andhra Pradesh,

Abstract: ***Many Decision Support Systems have been designed and developed successfully in many areas like medical, business agriculture, production, marketing and etc. This project aims to design a Decision support Information system for Improving School Facilities (DSS-ISF) which is useful to District development officers, Block development officers, Village Sarpanch & parents for decision making at appropriate level. The decision regarding improvement of facilities in the schools may be further improved using this Decision support system. The work involves the generation of Decision tree classifier model from the data collected from District Information System for Education (DISE). The performance of proposed system is compared with manual analysis of decision making to come out at a conclusion that, decision making can be bettered and made more effective with the help of computer based Decision support systems. Moreover, Decision-making is made more efficient and systematic by using Decision support system.***

Keywords: Decision support Information system, District Information System for Education (DISE), school facilities, decision tree.

1. Introduction

As per the census 2011 [2], the literacy rate in India has risen to 73 per cent in comparison to 64.8 per cent in 2001. While male literacy rate stands at 80.9 per cent, which is 5.6 per cent higher than the previous census, the female literacy rate has been recorded at 64.6 per cent, an increase of 10.9 per cent since 2001. Although the literacy rate increases, the quality of education is not up to the mark [7]. The quality of education comes with the availability of the infrastructure and various facilities in the educational institutes (Schools, colleges, Universities etc.). Many of the educational institutes in India lack of basic infrastructure that they need.

While considering the schools in India, the condition of government schools in India is very poor. Except for two or three states, all the states of India have poor school facilities. More children are in school than ever before, but the quality of government schools has sunk to spectacularly low levels. India has long had a legacy of weak schooling for its young, even as it has promoted high-quality government-financed Institutes.

Even though many children attend schools, they remain ill-equipped [3]. There are many reasons for this failure. Firstly, there is an acute shortage of teachers in the schools. There is also longstanding neglect, insufficient public funding and responsibility, and a lack of motivation among some teachers to give special care to poor children. Even basic amenities like water and electricity are not being provided. There is a need for urgency and educational facilities need to be improved [8]. Awareness needs to be brought out regarding the free and compulsory education to all children up to the age fourteen is a constitutional commitment in India. Also through the Right of Children to Free and Compulsory Education Act [4], 2009, all children of the age of six to fourteen years have the right to free and compulsory education in a neighbourhood school till completion of elementary education. And the quality and quantity of the infrastructure needs in the government schools to be improved.

A decision support system [1] (DSS) is defined by L. Adelman as "interactive computer programs that utilize analytical methods, such as decision analysis, optimization algorithms, program scheduling routines, and so on, for developing models to help decision makers formulate alternatives, analyze their impacts, and interpret and select appropriate options for implementation". A computer Information System [5] (IS) is a system composed of people and computers that processes or interprets information. The term is also sometimes used in more restricted senses to refer to only the software used to run a computerized database or to refer to only a computer system. The basic differences between Decision Support System and Information system is, an information system may or may not be used for making a decision whereas the decision support system provided for deciding on the policy, planning or implementation.

Data Mining is the set of embeddable analytical methods that provide the capabilities to explore, summarize, and model the data. With the evolving nature of information and data, it is challenging to extract key information from the large data set, which is most valuable during decision-making. In order to develop systems for digging down the data storehouse, data mining techniques are useful at the very core of such systems.

2. Related Work and Background

Before 1995, decision-making regarding developing the school facilities, involves manual analysis of enormous amount of information recorded on papers. Manual analysis of this information is time consuming. Thus the task of developing a school based statistical system was initiated by National Institute of Educational Planning and Administration (NIEPA) during 1995 and a first version of the software named as District Information System for Education [6] (DISE) was released during the middle of

1995. The district level professionals were assisted in the establishment of Educational Management Information System (EMIS) units, trained in the data collection, computerization. DISE is the most comprehensive information system in the education sector and it collected schools data in more than 200 districts and 18 states.
Even though huge amount of data collected and stored in computers still there is no statistical tool or any intelligent application to perform analysis on data to acquire knowledge and store important decisions to improve school facilities. The traditional approaches need lots of manual analysis on data by the domain experts to derive conclusions and decisions about current status of school facilities.

2.1 Problems in Existing System

- **Ineffective utilization of information:** Government deals with enormous amount of data regarding schools. Such information can be utilized in an efficient way, like to help strategic planners in decision making. A large data warehouse may be build, on which various analytical techniques may be applied to get useful results. These results then may be used to help decision makers.
- **Slower analysis of information:** As discussed earlier, most of the information regarding schools in current system is maintained in computers. Even though it is very difficult to analyze records quickly to help decision makers.
- **Non- uniform developments:** The level of decision-making is less synchronized in the current system.

3. Proposed Method

Thus the problem clearly shows that there is need for a tool which performs analysis on data and to conclude decisions fast and accurately. The tools must be cost effective, user friendly as well as accurate.

This work proposes a Decision support Information system for Improving School Facilities (DSS-ISF) using the decision tree model which helps strategic planners in making better decisions regarding improving school facilities at District, Block and village level. The scope of the proposed system is taken as the schools in the schools in Andhra Pradesh.

The schools data of the state Andhra Pradesh is taken from the DISE. This raw data has been divided into data containing general information regarding the schools and the data containing the information of schools facilities. The following facilities of the school are considered for classification of schools.

Various basic and other facilities in schools considered for the analysis are:

- Student-Teacher Ratio
- Classroom facility
- Drinking water facility
- Toilet facility
- Library facility
- Playground facility
- Boundary wall facility
- Building status
- Medical checkups facility
- Hostel facility
- Other facilities (Electricity, computer lab, HM room, etc.)

Training data for the classification is generated by considering the priorities (weights) given to the facilities i.e. the basic facilities should have more weights compared to the other facilities. These priorities are given considering the school facility indicators and government standards.

Objectives of the proposed method

- To automate the manual information analysis this is time consuming, error prone and tedious.
- To make the system of budget allocation transparent and efficient in improving school facilities.
- To help the government and planning officers to make decisions on as how much attention a school needs for its development.

3.1 Data mining in DSS-ISF

In DSS-ISF, data mining techniques have been applied on schools data warehouse to draw useful results, which will form the basis for decision-making. Data mining techniques like decision tree and classification have been used in our system.

- **Decision tree (DS):** A decision tree is a predictive model that, as its name implies, can be viewed as a tree. Specifically each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their classification. The schools facilities are given as training data to the decision tree algorithm and it generates a tree having nodes as facilities and leaves as classes Excellent, Good and Poor.
- **Classification:** Refers to the data-mining problem of attempting to predict the category of categorical data by building a model based on some predictor variables. From this decision tree one can classify a school according to the availability of the facilities in that school.

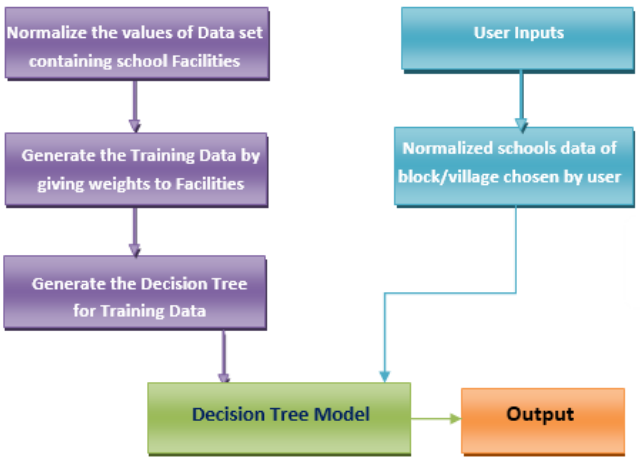

Figure 1: Classification of schools

Figure 1 shows how the classification of schools is done and

the steps are as follows.

The values of school facilities are normalized in to datasets
The school facility values from the raw data are taken and are normalized and formatted in to datasets.

Generation of training data
From the raw data a sample data was considered for generating the training data with class label. The school facilities in the data are divided in to basis facilities and other facilities.

Weights are assigned to each basic facilities and a rating value is calculated using the equation,
$Rating(x) = b_1x_1+b_2x_2+b_3x_3+b_4x_4+ \ldots. +b_nx_n$ where,
$b_1, b_2 \ldots b_n$ are weights and
$x_1 , x_2 , \ldots x_n$ are values of the various basic facilities of particular school.

Based on the rating values a class label is given to each school record thus training data is generated.

Decision tree ID3 algorithm is applied on the training data
The Decision tree ID3 algorithm is applied to the training data and the algorithm generates a decision tree.

Decision rules are generated from decision tree and are applied to the other school facility data for analysis
The decision tree generates decision rules and is applied on the other school facility values of schools in villages which are selected by the user for analysis.

3.2 Architecture of DSS-ISF

Figure 2 depicts the overall architecture of DSS-ISF and it consists of three components and each component is described as follows.

- **Schools' data warehouse:** The information of schools' like general information and facilities related information is gathered and is stored in a schools' data warehouse.

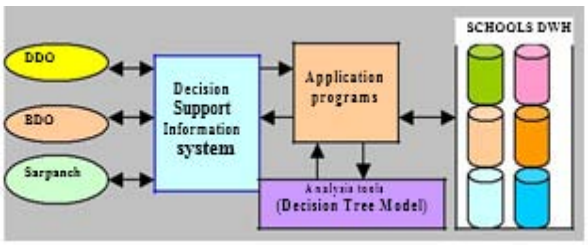

Figure 2: Components of DSS-ISF

- **Decision support information system:** This component of DSS-ISF is also called as user interface. This part of system aids decision makers in taking a decision. Three level namely District –level, Block level and village level decision Support is provided by DSS-ISF.
- **Analysis tools**: Component of DSS-ISF which works on the information stored in the schools data warehouse is called as analysis tool. Here there is a decision tree model which works on the school facilities. This tool summarizes the information and gives useful results, which are helpful to the decision makers.

4. Algorithm used for DSS-ISF

A machine researcher named J. Ross Quinlan in 1980 developed a decision tree algorithm known as ID3 (Iterative Dichotomiser). Later, he presented C4.5, which was the successor of ID3. ID3 and C4.5 adopt a greedy approach. In this algorithm, there is no backtracking. The trees are constructed in a top-down recursive divide-and-conquer manner.

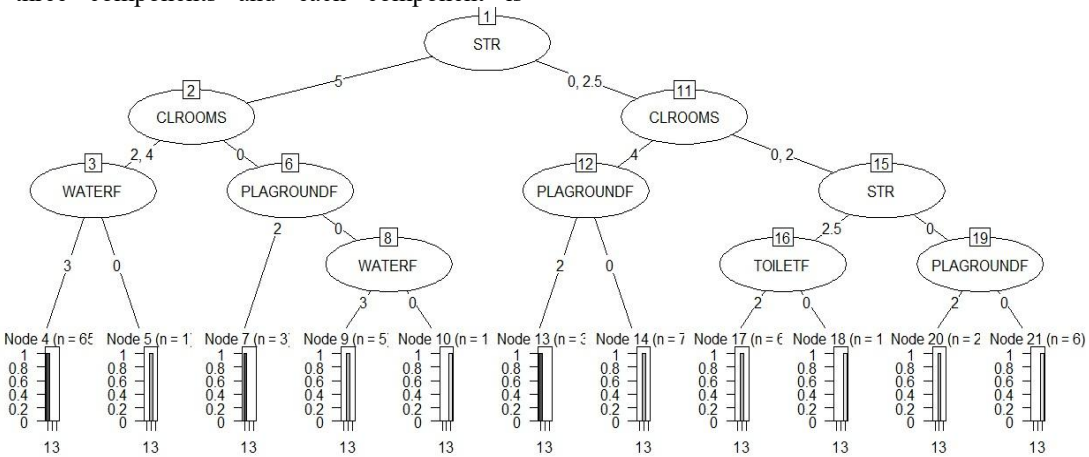

Figure 3: Decision tree for basic facilities in schools

Generation of Decision tree from training tuples of data partition DS.

Algorithm: Generating decision tree
Input:
Data Partition *DS*, which is a set of training tuples of various facilities present in a school and their associated class labels.
attribute_ list, the set of candidate attributes.
Attribute_selection_method, a method to find the splitting criterion that best partitions the *DS* tuples in to individual classes. This criterion includes a splitting attribute and either a splitting point or splitting subset.

Output:
A Decision Tree.
Method:
Create a node *N*;
if tuples in *DS* are all of the same class, *C* **then**
return *N* as leaf node labeled with class *C*;
if *attribute_list* is empty **then**
return *N* as leaf node with labelled with majority class in *DS*; | |majority voting
apply **Attribute_selection_method**(*DS*, *attribute_list*)
to find the best *splitting criterion*;
label node *N* with *splitting criterion*;
if *splitting_attribute* is discrete-valued **and** multiway splits allowed **then**
attribute_list= attribute_list – splitting_ attribute;
//remove splitting attribute
for each outcome j of *splitting criterion*
//partition the tuples and grow sub trees **for each** partition
let DS_i be the set of data tuples in *DS* satisfying outcome *j*; // a partition
if DS_i is empty **then**
attach a leaf labelled with the majority class in *DS* to node *N*;
else
attach the node returned by **Generate_Decision_Tree**(DS_j, *attribute list*) to node *N*;
endfor
return *N*;

The decision tree generated by applying ID3 algorithm on the basic facility values in the schools is shown in figure 3. STR, CLROOMS, WATERF, TOILETF, LIBRARYF and PLAYGROUNDF are the nodes of the tree, representing the basic facility values of the school. The values on the edges represent the normalized value of the respective basic facility. Finally the leaf nodes represent the class labels as excellent or good or poor to which particular school is classified.

5. Concluding Remarks

The performance of the DSS-ISF is much better than the current system of manual analysis regarding the time, risk of analyzing the information and making appropriate decisions and its calculated efficiency is ~98.2%. The performance of DSS-ISF may be further improved by considering few more decision factors like Demographic factor, Quality of facilities, literary rates, distance of the availability of the government school to the village etc. which also helps in making decisions regarding the need for start a new school in the village. Few other Decision support Systems (DSSs) for e-governance may be developed in near future such as

- DSS for population Control planning.
- DSS for Pension Planning schemes.
- DSS for Pay commission for state governments etc.

References

[1] Aravind Selwal "A Decision Support System for Village Economy Development Planning" he literacy rates in India from census data 2011: http://www.dailymail.co.uk/indiahome/indianews/article-2317341/Indias-literacy-rate-rises-73-cent-population-growth-dips.html
[2] An article on condition of government schools in India:http://theviewspaper.net/condition-of-government-schools-in-india/
[3] Ministry of Law and Justice "The Right of Children to Free and Compulsory Education Act, 2009, No. 35 of 2009, Ministry of Law and Justice, The Gazette of India, August 27, 2009, New Delhi"
[4] Definition of Information system: https://en.wikipedia.org/wiki/Information_system
[5] District Information system for Education (DISE) is the most comprehensive information system in the education sector: www.dise.in
[6] A. Abdul Salim "Educational Development at Micro Level Case Study of Two Villages in Kerala."
[7] Michael Ward "A report on Rural Education."