ISSN:2249-5789
K Sagar et al, International Journal of Computer Science & Communication Networks,Vol 5(5),279-282

ID3 Based Decision Support System for Establishment of Coir Industry

K Sagar [#1], V.Valli kumari [#2]
[#1,#2] *Dept. of Computer Science and System Engineering, Andhra University, Visakhapatnam, Andhra* Pradesh
[#1] sagarkkommu@gmail.com, [#2] vallikumari@gmail.com

Abstract

This paper describes the development of decision support system for making a decision to establish a coir industry. The system shows a list of villages which are suitable for establishing an industry by considering input district and mandal. The system also provides, the decision establishing an industry based on user specification. ID3 algorithm is used as a machine learning algorithm. It generates the decision tree for making decision, based on training data.

Keywords- coir industry; machine learning; ID3; decision support system

1. Introduction

Establishment of any new organization such as an industry, in a geographical area, the decision and intelligence of human being are required as, the location has proper habitat factors for living or the location has less-possibilities to get affected by disasters. Sometimes, we may ignore or do not consider a few conditions or measurements, important to make decisions. The person alone may not have sufficient knowledge about the effecting factors in that area to make a decision, and they have to hire some professional. An intelligent decision support system (DSS) is required in such conditions, to provide a complete guidance.

DSS is defined as "an interactive, flexible, and adaptable computer based information system, specially developed for supporting the solution of a non-structured management problem for improved decision-making. It utilizes data, it provides easy user interface, and it allows the decision maker's own insights"[2]. There is a growing trend to provide managers with IS that can assist them in their most important task – making decisions[6]. All levels of management can get benefits with the usage of the DSS capabilities. The highest level of support is usually for middle and upper management. DSS has three fundamental components they are:

- Database management system (DBMS). A DBMS serves as a data bank for the DSS. It stores large quantities of data that are relevant to the class of problems for which the DSS has been designed and provides logical data structures (as opposed to the physical data structures) with which the users interact. A DBMS separates the users from the physical aspects of the database structure and processing. It should also be capable of informing the user about the types of data that are available and how to gain access to them.
- Model-base management system (MBMS). The role of MBMS is analogous to that of a DBMS. Its primary function is providing independence between specific models that are used in a DSS from the applications that use them. The purpose of an MBMS is to transform data from the DBMS into information that is useful for decision making. Since many problems that the users of a DSS will cope with may be unstructured, the MBMS should also be capable of assisting the user in model building.
- Dialog generation and management system (DGMS). The main product of an interaction with a DSS is insight. As their users are often managers who are not computer-trained, DSSs need to be equipped with intuitive and easy-to-use interfaces. These interfaces aid in model.

DSS needs a machine learning algorithm for utilizing the data and provide solution of a non-structured management problem for improved decision-making. In this DSS, ID3 decision tree algorithm is used as machine learning algorithm for making a decision to establish a coir industry.

In most of the countries like India, industries play a vital role in providing employability, resource utilization and income generation. Coir industry is one of those industries. It costs about Rs. 2 to Rs.5 lakh to set up a unit and each unit requires manpower of 8 to 15 labors a day. There is no dearth of market for coconut fiber and ropes, as it has a greater demand in the international market[5].

The coir fiber can be used in following products. They are coir yarn, mats, mattress, rugs, geo textiles, coir pith, garden articles, and diversified products such as umbrella, jewellery, bags, composite board and coir woods. The location of this industry depends on factors such as raw materials, water, manpower, area, transportation, market. By considering the above constraints, DSS is given a decision to establish coir industry.

2. Problem statement

Generally, to establish an industry in a village, needs to consider parameters such as availability of raw material, power consumed by the industry, water required[4]. So, entrepreneur has to gather all the details about these parameters and he/she has to visit all the villages (It time consuming and cost effective and taking few opinions from the domain experts). Sometimes taking opinions from the experts may lead to confusion for making decision or commit to false decision.

3. Proposed system

The proposed system is to replace the domain experts in the existing system and reduce the time taken by the entrepreneur to gather and analyzing the data. It replaces the expert by DSS. The DSS consists of two modules. These modules mainly deal with selection of an area to establish a coir industry by using the ID3 algorithm. The DSS is developed at mandal level for analyzing the gathered data and to provide a list of suitable villages for placing a coir industry and also it is developed at village level on entrepreneur requisites and provides decision about whether the particular village is suitable or not.

The DSS builds a knowledge base from village and industry databases which are crucial for decision making. The village database contains the information related villages such as population, marginal workers, raw material available, water availability, unused land. The industry database contains information about the minimum requirements of the industry.

By using the predefined procedures, training data is constructed from the above databases. The classification technique, ID3 algorithm has been applied on that training data to build a knowledge base for the DSS.

Fig.1 shows, the working flow of the proposed DSS. When the entrepreneur needs a decision, he has to fill the input form, which is provided by DSS. The DSS tests the input data based on knowledge base and generates class label for that input. Based on that class label, it gives the output to entrepreneur as suitable or not suitable.

4. ID3 Algorithm

A decision tree is a tree, and it is has branch nodes and leaf nodes. Each branch node represents a choice between a number of alternatives, and each leaf node represents a decision[1][3]. Decision trees are usually used to acquire information for the purpose of decision making. A user starts actions from the root node of the decision tree. From this node, users should recursively apply decision tree learning algorithm on each node to construct a decision tree. In this each branch node represents a possible scenario of decision and its outcome. Table I describes some of various parameters used in this algorithm.

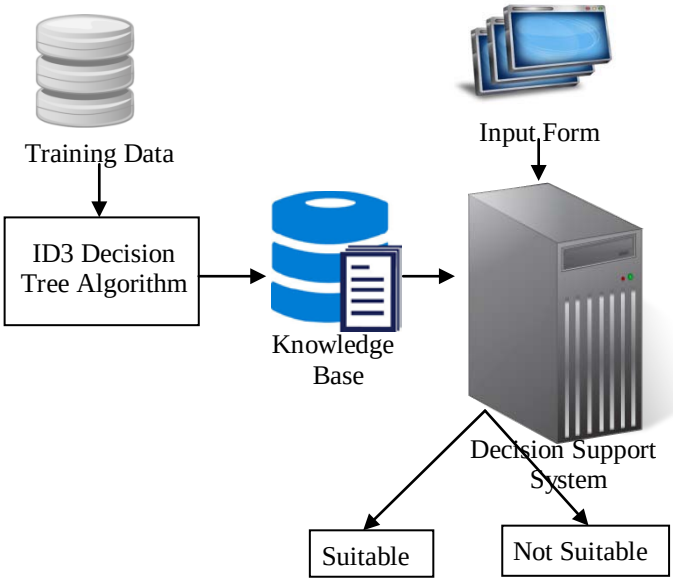

Fig. 1 Flow diagram of the DSS to establish an industry.

The domain values for some of the variables which are defined in this system are as follows.

Land – Availability of unused land in sq.km is obtained. Availability of land is divided in three classes: Min – below 0.9 sq.km, Mid – between 0.9 and 2.0 sq.km, Max – above 2.0 sq.km.

Water – Availability of water in TMCs is obtained. Availability of water is divided in three classes: Min – below 0.3 TMCs, Mid – between 0.3 and 0.7 TMCs, Max – above 0.7 TMCs.

Manpower – Availability of man power obtained. Availability of man power is divided in three classes: Min – below 150 persons, Mid – between 150 and 300 persons, Max – above 300 persons

RawMaterial – Raw material are obtained. Availability of raw material is divided in two classes: coconut and sugar cane.

Transportation – Availability of Transportation is obtained. Availability of transportation is divided in three classes: road, train, and marine.

EstablishCoir – Establish coir industry is obtained and it is declared as a class attribute. Yes - If all requirements are satisfied, N0 – if any requirement is not satisfied.

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan[7][1]. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given set to test each attribute at every tree node, in order to select an attribute that is most useful for classifying a given set. A statistical property called information gain is defined to measure the worth of the attribute.

Table I: Coir industry training data and its variables.

Variable	Description	Possible values
Land	Unused Land	{Min, Mid, Max}
Water	Water facility	{Min, Mid, Max}
ManPower	Man power	{Min, Mid, Max}
RawMaterial	Availability of raw materials	{Coconut, Sugar cane}
Transportation	Transportation availabilities	{Road, Train, Marine}
EstablishCoir	Decision to establish coir industry	{Yes, No}

a) Measuring Impurity

Given a data table that contains attributes and class of the attributes. Based on the classes of the table, algorithm can measure homogeneity (or heterogeneity) of the table. If it is pure or homogenous, it contains only a single class. If it is impure or heterogeneous, it means that it contains several classes. To measure the degree of impurity or entropy,

$$Entropy = \sum -P_j \log_2 P_j \quad (1)$$

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 the classes in the table have equal probability.

Calculate the entropy of *S* before calculating the information gain for *A* relative to *S*. Here, *S* is a set of 131 instances containing 90 *"No"*, and 41 *"Yes"*.

$$Entropy(S) = -P_{no} \log_2(P_{no}) - P_{yes} \log_2(P_{yes}) \quad \ldots(2)$$

$$= -(90/131)\log_2(90/131) - (41/131)\log_2(41/131)$$

Entropy (S) = 1.6372.

b) Information Gain

To determine the best attribute for a particular node in the tree, information gain is calculated. To calculate the information gain of the attribute **A**, relative to the set **S**,

$$Gain(S, A) = Entropy(S) - \sum_{V \in Values(A)} \frac{|S_V|}{|S|} Entropy(S_V) \quad \ldots (3)$$

Information gain is calculated for all the attributes. Table II describes the information gain of all the attributes of qualitative parameters.

From the Table II, RawMaterial has the highest gain, therefore, it is used as the root node as shown in Fig.2.

Table II: Information gain values of coir industry qualitative parameters.

Gain	Values
Gain(S, Land)	0.1412
Gain(S, Water)	0.1668
Gain(S, RawMaterial)	0.4988
Gain(S, Manpower)	0.0920
Gain(S, Transportation)	0.0402

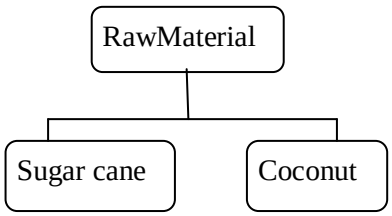

Fig. 2 A root for a given coir industry training data.

c) Procedure to build a decision tree

By observing the above calculations, the attribute RawMaterial is used to expand the tree. Then, remove the attribute RawMaterial from the samples of these sub-nodes and compute the Entropy and the Information Gain to expand the tree using the attribute with highest gain value. Repeat this process until the Entropy of the node equals null. At that moment, the node cannot be expanded anymore because the samples in this node belong to the same class.

5. Results

This coir industry DS System is trained by using ID3 decision tree algorithm with the coir industry training data. ID3 machine learning algorithm is simple and effective to classify the given input. This machine learning algorithm gives 90.07% accuracy.

Table III, describes the classification accuracy of ID3 and C4.5 algorithms, when these algorithms applied on the training data of coir industry using 10-fold cross validation is observed as follows,

Table III: Decision tree classifier Accuracy.

Decision tree algorithms	Correctly classified instances	Incorrectly classified instances	Unclassified classified instances
ID3	90.07%	9.16%	0.76%
C4.5	89.31%	10.69%	0%

The decision given by this system is most trust worthy. It is proved by observing the accuracy of the machine learning algorithms.

DSS developed for the villages in the East Godavari district. In East Godavari, there are 59 mandals and 1370 villages. It gives the suitable villages from the above villages. The entrepreneur

can filter suitable villages by choosing mandal in village by mandal window in DSS system shown in Fig.3. In this window entrepreneur choose district and mandal and submit the form for suitable villages. By taking input from the entrepreneur, DSS applies ID3 on input data then it gives the list of suitable villages as an output as shown in Fig.4.

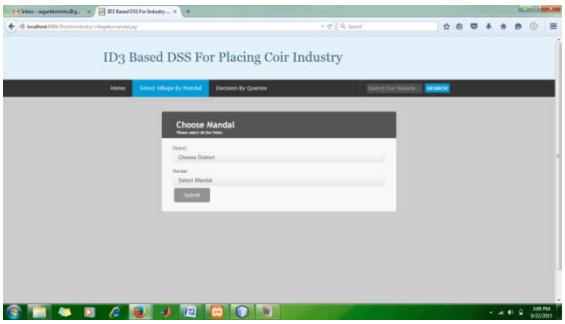

Fig.3. Mandal window in DSS

Along with the suitable villages the DSS shows the geographical data for every village. So, the user can choose the best from the suitable villages by looking into that data.

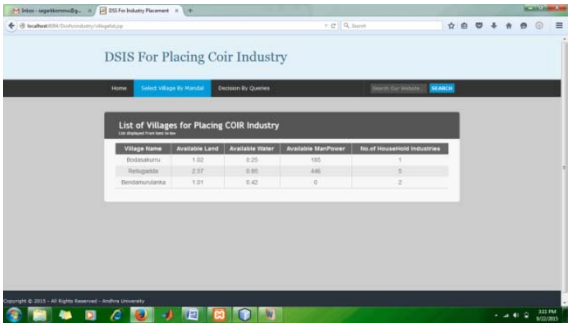

Fig.4. Suitable villages in chosen mandal.

If the entrepreneur has some constraints on parameters of coir industry and he/she wants to know placing an coir industry is possible or not, Then entrepreneur choosing Decision by Requisite window in DSS shown in Fig.5.

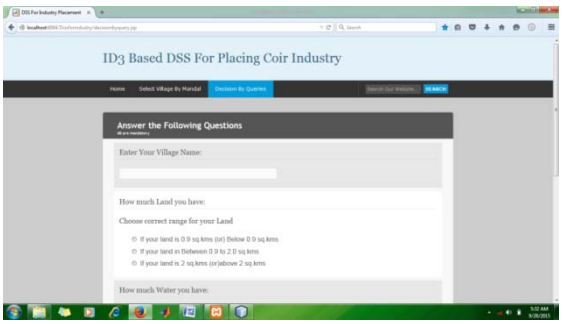

Fig. 5. Decision by Requisite window in DSS.

In this window entrepreneur has to give the constraints to DSS as input. The DSS applies ID3 on the given constraints and gives the suitability to placing coir industry as shown in Fig. 6.

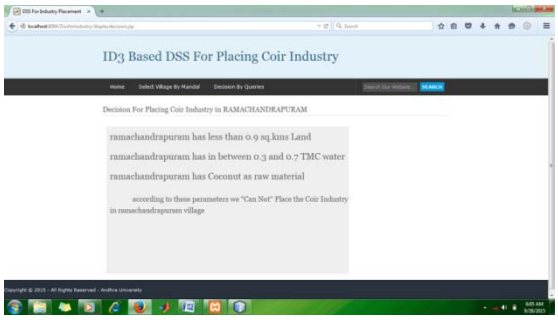

Fig. 6. Suitability of given constraints to establish coir industry.

6. Conclusion and future scope

The DSS for establishing a coir industry is implemented. It guides the user (entrepreneur) to make a decision to establish an industry in villages of East Godavari. This System is user friendly as it provides an interactive web interface which takes the user through a series of screens that provide the user with supporting decision, depending on the user selection such as mandal, industry in DSS and village name, raw material, transportation, water, area required in DSS by requisite.

This system is restricted to coir industry and East Godavari district. It can be extended by including several types of industries, and it can be extended to total state.

7. References

[1] T.Miranda Lakshmi, A.Martin, R.Mumtaj Begum, Dr.V.Prasanna Venkatesan.: "An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data", *I.J.Modern Education and Computer Science*, 2013, 5, 18-27.

[2] Udo Richard Franz Averweg.: "Decision-making support system: Theory and practice".

[3] K.Wisaeng.: "AComparison of Different Classification Techniques for Bank Direct Marketing",2013,IJSCE.

[4] http://www.yourarticlelibrary.com/industries/factors-influencing-the-location-ofindustries-geographical-and-non-geographical-factors/19695/

[5] http://www.thehindu.com/todays-paper/tp-national/tp-andhrapradesh/coir-units-making-good-business-in-east-godavari/article2849723.ece

[6] Randall E. Louw, Decision Support System.

[7] J.R.Quinlan, "*Induction of Decision Tree*", Journal of Machine learning", Morgan Kaufmann Vol.1, 1986, pp.81-106.