

Counsel System for Effective Farming using Data Mining Algorithm

T. Rangunthar¹, S. Selvakumar², G. Ilamurugan³

^{1,3}Dept. of Computer Science, Sri Sai Ram Institute of Technology, Chennai

²Dept. of Computer Science, G.K.M College of Engineering and Technology, Chennai

Abstract- Agriculture is one of the sectors where technology and advancements in technologies have not been utilized to the fullest potential. In this work, we took the issue of selecting a crop for farming on an agricultural land. If there is more than one option to plant a crop at a time using limited resources, then selection of crop is a puzzle. This paper proposes a method named Crop Counsel Method to solve crop selection problem, and maximize net yield rate of crop over season. The selection of crop is going to be based on profit percent achieved per hectare of land based on the given resources calculated on data mining techniques. For classification of data sets and association between datasets we have used decision tree induction and Apriori. It can subsequently lead to economic growth of farmers and our country.

Keywords: Farming; Counseling system; Data Mining; Decision tree induction; Apriori

1. Introduction

Agriculture has always been the backbone of our country. In India almost 60% of the people are dependent on agriculture. So it becomes an important aspect to advance the agricultural practices so that the farmers could earn more profit and get technological help for their living.

Many farmers are unaware of the results they are going to get when they sow a certain crop. So they follow a trend or go by trial method for sowing the crop. This is an inefficient method as it doesn't guarantee success neither gives a value which will determine the probability of success. Therefore this paper proposes a crop recommendation system for selection of crops for certain parameters based on profitability of the crop for the current season. The system is going to be a two phased project. In one phase we will take all the input from the user, then a crop based on his/her requirement is recommended to him. In second phase we assume that the user has very little information about the parameters and so the system generates an automatic list of crops. In both the phases the last steps would be prediction of rainfall and sorting of crops based on profit rates for that year.

Data sets are the foundation of the recommendation system. The first step of Crop Counsel Method (CCM) is to collect data as far as possible from the Internet. We download data from government websites, farming websites and channels, and so on. After the data collection, we structure it into a table format with all the necessary parameters listed. From this table we classify the data using decision tree

induction algorithm. In decision tree induction, we use information gain method to find the priorities of various parameters for making a tree. Now the data's are split into the sub-trees based on priority values and the data is classified.

2. Literature survey

From the research article [1-2], the researcher expresses that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages. There are several applications of Data Mining techniques in the field of agriculture. The researchers implemented [3] K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied [4] for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed [5] using Support Vector Machines. Soil profile descriptions were proposed [6] by the researcher for classifying soils in combination with GPS based technologies. They were applied K-Means approach for the soil classification. In a similar approach, crop classifications using hyper spectral data was carried out [7] by adopting one of the data mining approach i.e. Support Vector Machines. One of the researchers used [8] an intensified fuzzy cluster analysis for classifying plants, soil and residue regions of interest from GPS based color images. In [1], researchers used MLR and Density based clustering techniques for rice production and compared it with actual production. They are finding out the variation in predicted and actual production of rice for the region. Prediction of rainfall is done using [9] linear trend and average comparison.

3. Methodologies

The data collected for rainfall prediction is for last 20 years of predicted and actual rainfall in Tamil Nadu state across the four seasons. The crop data are taken into following variables
soil type, duration, water needed, seasons sowed, budget, profit per hectare. Duration is given in no. of days, water requirement is given as millimeters and profit is calculated on market price of the crop for that season.

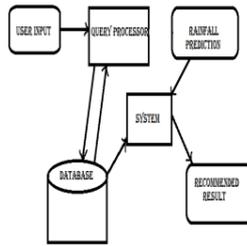


Fig 1. Overview of counsel system process

3.1 Classification

For the classification of the data collected we are going to use decision tree induction. Decision tree learning uses a decision tree as a predictive model which maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

In this method we calculate information gain. Information gain is the amount of information that's gained by knowing the value of the attribute, which is the entropy of the distribution before the split minus the entropy of the distribution after it. The largest information gain is equivalent to the smallest entropy.

| Farmer Id | Rice | sugarcane | Wheat | Coconut |
|-------------|------|-----------|-------|---------|
| FarmeridFI1 | x | x | | x |
| FarmeridFI2 | | x | x | |
| FarmeridFI3 | x | x | | |
| FarmeridFI4 | x | | | x |
| FarmeridFI5 | | x | x | |
| FarmeridFI6 | | x | | x |

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$

The attribute with highest value will serve as root of the tree and similarly different nodes will be placed in the tree. So any new query will be first classified based on the decreasing order of information gain

The data table has attributes soil types, duration, water requirement, season and crops for these values. For the decision tree induction the information gain for all these are found out and then the tree is constructed.

Table 1. Farmer Training Dataset

3.2 Prediction of rainfall

The prediction of the rainfall is done based upon the previous year's values. The data is compared for all the years season by season and using a linear trend function a range of values for the current season can be predicted.

a) Algorithm for prediction of actual rainfall

Inputs: Rainfall data.

Output: Forecasted Rainfall

Start

Step 1: Accept rainfall data from the rainfall database

Step 2: Calculate average rainfall from the rainfall data

Step 3: Repeat Steps (3.1 to 3.4) for requested periods

Step3.1: Estimate rainfall using Linear trend

Step3.2: If (Estimated rainfall \geq (Average rainfall + 10% Average rainfall)) then Rainfall status = Above Normal

Step3.3: If (Estimated rainfall \leq (Average rainfall - 10% Average rainfall)) then Rainfall status = Below Normal

Step3.4: If (Estimated rainfall lies between above and below normal limit) then Rainfall status = Normal

Step 4: Display the Rainfall status.

Stop

Stop

b) Algorithm for prediction of normal rainfall

Inputs: Rainfall data.

Output: Forecasted Rainfall Data

Start

Step 1: Accept the rainfall data from the rainfall database made by admin.

Step 2: Predict the normal rainfall by linear trends

Stop

3.3 Prediction of Crop Match

What are association rules? Association rule learning is a data mining technique for learning correlations and relations among variables in a database. What's an example of Apriori. Let's say we have a database full of farmer crop cultivation. You can think of a database as a giant spreadsheet where each row is farmer database and every column represents different crop cultivation.

By applying the Apriori algorithm, we can learn the grocery items that are cultivated together by association rules. You can find those items that tend to be cultivated together more frequently than other items — the ultimate goal being to identify the crop which cultivated together, For example:

You can probably quickly see that rice + coconut seem to frequently occur together. These are called 2-itemsets. With a large enough dataset, it will be much harder to "see" the relationships especially when you're dealing with 3-itemsets or more. That's precisely what Apriori helps with, you might be wondering how Apriori works? Before getting into the nitty-gritty of algorithm, you'll need to define 3 things:

The first is the size of your item set. Do you want to see patterns for a 2-itemset, 3-itemset etc. The second is your support or the number of transactions containing the item set divided by the total number of transactions. An item set that meets the support is called a frequent item set.

The third is your confidence or the conditional probabilities of some item given you have certain other items in your item set. A good example is given rice in your cultivation, there is a 67% confidence of having coconut also in your cultivation. Through apriori method we can identify the best friends crop for cultivation from the training dataset which give better profit in cultivation.

3.4 Foretelling

For recommendation we use a combined result of prediction and profit calculation sorted in decreasing order. The absolute difference between required water and rainfall is calculated and the crops are further shortlisted. Among these crops profit of all the crops are calculated based upon the market price of the crop.

The crops giving most profit per hectare of land are recommended to the user. In case the user wants to grow a crop out of their wish the profit value changes based on the resources to be acquired by that crop. The user is provided with whole set of data about the crops and its requirements.

$$Profit = market\ rate \times yield\ per\ hectare - cost\ of\ production$$

4. Results

The following results are obtained from all the modules

4.1 Classification

For the classification technique we calculated information gain for all the parameters and the result obtained are as below.

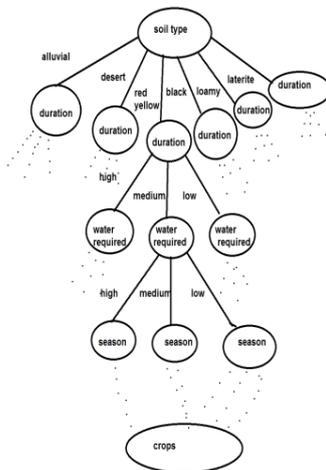


Fig 2. Decision tree Induction for crop, rainfall, soil and water
 soil type=1.19
 duration=1.14
 water requirements=1.12
 season=0.705

This means that selection of a crop is mainly based on soil type, then the duration the crop is taking, then the water requirements of the crop and at last the season of the crop.

4.2 Prediction of Rainfall

The prediction of rainfall module was taken from a research paper [9], which concludes it to be accurate with actual parameters by 14%. The module is used to get the predicted value of rainfall from the database.

For the current year the forecasted value is found out to be
 Summer=146.52 mm
 Winter=19.55mm
 Monsoon=299.595mm
 Post monsoon=364.235mm

These values are an approximation of forecasted rainfall of the year. Once the actual value is received the database will be updated and the forecast will be based on those values.

4.3 Foretelling

A resultant crop set was found, among the crops set the crops giving higher returns were listed in descending order and the top two are recommended.

Also any random selection of crop is entertained by the user and the details regarding the same are displayed.

In phase 1 the values obtained are accurate and are easily matched.

In phase 2 the values are submitted in form of a range and which are in turn mapped to values and similar results are obtained.

In both cases selection of crop which are not recommended is also shown, it provides the freedom to select a crop based on the user wish and needs.

5. Conclusion

A methodology for recommending a crop based on some given parameters is developed. The results are based on

rainfall prediction and market price of the crop. This system is useful for farmers who find it difficult to select a crop and hence find it difficult for to earn profit. This is done using two phases making it more users friendly. The system can also be used to determine all the needs and requirement for a crop and so can be helpful in reducing heavy debts incurred to farmers thereby improving economic condition of both the farmer and the region.

REFERENCES

[1] D Ramesh, B Vishnu Vardhan "ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES"
 [2] G Ruß, "Data Mining of Agricultural Yield Data : A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages : 24-37.
 [3] Jorquera H, Perez R, Cipriano A, Acuna G, "Short Term Forecasting of Air Pollution Episodes", In: Zannetti P (eds) Environmental modeling , WIT Press, UK, 2001
 [4] Rajagopalan B, Lall U, "A K-Nearest Neighbor Simulator for Daily Precipitation and Other Weather Variables", Wat Res Res 35(10), 1999, pages: 3089-3101.
 [5] Tripathi S, Srinivas V V, Nanjundiah R S, "Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach", J Hydrol, 2006, pages: 621-640.
 [6] Verheyen K, Adriaens D, Hermy M, Deckers S, "High-resolution continuous soil classification using morphological soil profile descriptions", Geoderma Vol.101, 2001, pages: 31-48.
 [7] M Trnka, "Projections of Uncertainties in Climate Change Scenarios into Expected Winter Wheat Yields", Theoretical and Applied Climatology, vol. 77, 2004, pages : 229-249
 [8] Meyer G E, Neto J C, Jones D D, Hindman T W, "Intensified Fuzzy Clusters for Classifying Plant, Soil and Residue Regions of Interest from Color Images". Computer Electronics Agric Vol. 42, 2004, pages: 161-180.
 [9] Dayanand. G. Savakar, Dr. Venkatesh, Dr. Anami.B.S, R. R. Rohidekar "Rainfall Prediction based on Rainfall Statistical Data" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 3

