DOI: 10.21917/ijsc.2020.0303

NOVEL REGION SPECIFIC DECISION SUPPORT SYSTEM FOR CROP SELECTION AND CULTIVATION

P. Brindha and K. Kavitha
Department of Computer Science, Mother Teresa Women's University, India

Abstract
The prime concern of any country is Agriculture. Every nation has to feed its population by making strong policy support for Agricultural production. This paper deals with providing decision support for the crop to be selected for cultivation based on several influencing parameters. Even though there are many decision support systems available for Agriculture, there is a lack in region specific ones. The proposed system aims to overcome the aforementioned issue. For this purpose, the system considers climatic data from the Government of India web portal, Tamil Nadu Agricultural University portal and reports. The precision data collected from the fields will be given as inputs to the proposed system. The crops to be selected for cultivation are based on the historical data and guidelines from the Tamil Nadu Agritech portal. The accuracy of the proposed decision support system is assessed by getting feedback from the farmers.

Keywords:
Accuracy, Agriculture, Crop Selection, Decision Support System, Historical Data

1. INTRODUCTION

The agricultural produce of any nation has to be increased to meet its growing population. This is a crucial aspect that has to be addressed by India, and it is a very viable proposal as India is a land of agriculture. Agriculture plays a vital role in the economy of the nation. Majority of the population depend on Agriculture as their primary means of livelihood. The Indian farmers belong to small, medium and large scale category, wherein 90% of the farmers in Tamil Nadu fit into small and marginal category [2].

Talking about increasing the agricultural produce, crop selection is one of the major decisions influencing the productivity factor. The crop should be selected based on numerous parameters like soil condition, change in weather, rainfall and water availability, pest infestation data etc. If an unsuitable crop variety is selected for sowing, it will definitely affect the yield and business value of agriculture. Consequently, the farmers, the backbones of the nation, will face huge loss and this may eventually lead to the scarcity of food.

Restricting its scope to the Madurai region in Tamil Nadu, this work strives to contribute some findings that would help solve the aforesaid issue. In this regard, the objective of this research work is to collect the agricultural parameters, to identify the parameters influencing the crop growth, and to develop a Decision Support system for crop selection for cultivation. In addition to the traditional way of collecting data from agricultural fields, data acquired through precision agriculture [7] is more accurate and easy to incorporate for different analytics.

The precision data acquired from the fields are considered as the prime inputs for this research work. The inputs are then analysed with the various data collected from the agricultural website and reports [1]. These data are usually called as historical data. By considering all the related parameters, data mining techniques are employed to arrive at the decisions. The obtained decisions are then evaluated to prove their suitability.

The entire research paper is organized as follows: section 2 briefly describes the existing works that prompted the conduct of this research. Section 3 explains the phases of the proposed framework. Section 4 presents the metrics used for the evaluation as well as the results; and section 5 concludes the paper with final remarks and directions for future research.

2. REVIEW OF LITERATURE

Every research work is motivated by several existing works. This section will elaborate the existing works that motivated the proposed research work.

Rupnit et al. [3] proposed an AgroDSS. The AgroDSS is implemented as a cloud based service. It provides decisions related to improving the business value of Agriculture, support in avoiding risks, and helps in better planning. The authors have presented an elaborate case study with respect to pests in both orchards and fields. This paper mainly focuses on the business perspective.

The Cropsyst [6] is started in early 1990s. This is a simulation model used to analyze the crop productivity and the impacts of the environment. This Cropsyst is evolved to incorporate yearly crop simulations as well as rotations. It has taken into consideration the various factors, and produces different other supporting simulators for watershed, weather etc. This is evolving day by day to adapt the changes. The economic risk analysis is also done using this simulator.

Lindblom et al. [8] addressed the drawbacks of the existing AgriDss. They recorded that the lack of user centric design and participatory approach will be a great hindrance in implementing the existing decision support systems for agriculture. Also, they highlighted the advantages of using precision agriculture. According to them, the agricultural knowledge and innovation system should be the kernel of the agricultural decision support system. Then only it will address the growing changes and accommodate sustainability issues in farms.

In [4], a recommender system is proposed for nutrient management. The horticulture crop tomato is chosen for experimental study. The data acquired using different sensors and the impacts of nutrient deficiency caused by different fungal and other infections are studied in this research work. The authors considered parameters related to growth of the crop and yield as their major criteria. More yields without proper nutrient value will be a loss to the consumer of the yield. This is addressed in this research work. Here the terrace garden alone is taken into account.

A DSS for Precision agriculture [12] is developed for the smart irrigation project sponsored by the United State Department of Agriculture. The required information is gathered through the field sensor, transmitted through WSN and the decisions are received by the farmers using mobile devices. The decisions will be sent through email or text messages. The data cloud is enabled by NSF.

The prediction based agro advisory system [11] is proposed to predict the remedies for pest related issues. The pest surveillance data, expert advice, historical data, climate and weather parameters are collected to find the suitable advice to the farmers. The feedbacks from the stakeholders are then considered to refine the advices. Machine learning algorithms are employed for predictions.

Abedinpour et al. [10] proposed Aquacrop model. It is simulated for kharif Maize crop. The model is developed to predict the yield of the Maize crop in semi-arid environment in India by considering two parameters. They are: irrigation type, and level of nitrogen. Different water treatments and nitrogen levels were studied for a period of one year and used to calibrate the yield for the next year.

In [9], the impact of climate change on pest and disease infestation in rice is studied. The infestation of pest is influenced by several parameters like minimum temperature, wind speed, rainfall, etc. The major pests of the paddy crop are analyzed and its impacts are clearly visualized in the research work.

Though there are varieties of Decision Support systems available in the literature, the lack of region specific decision support systems for assisting the stakeholders has propelled this research work to existence. The following section elaborates the proposed framework.

3. PROPOSED FRAMEWORK

The proposed research work is sub divided into several phases for reducing the complexity. The phases of the proposed work are depicted in the Fig.1. There are four phases namely, Data collection phase, Preprocessing phase, Decision support phase, and Outcome and Evaluation phase.

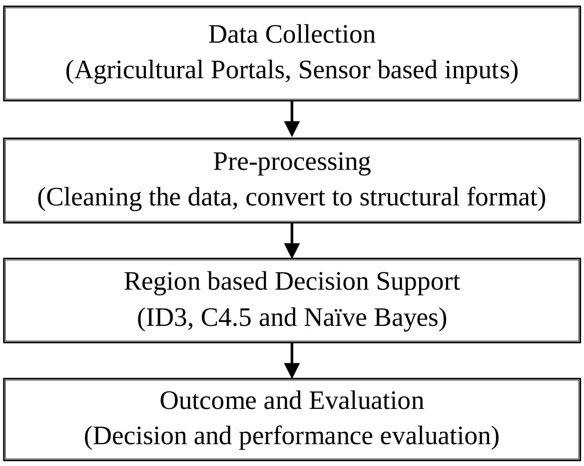

Fig.1. Region based Decision Support for Crop selection and Cultivation

Here for drawing decisions about region specific crop selection for cultivation, three algorithms are used. The performance and suitability of the three algorithms are analyzed in the final phase. The phases of the proposed work are explained in the following section.

3.1 DATA COLLECTION PHASE

As far as agriculture is concerned, the data used for making decisions are available from different sources. Moreover, the parameters tend to change according to the change in our environment. So, information related to weather, rainfall, crop, soil, optimal condition, water treatment, fertilizers, season-specific crop varieties, pest infestation and remedy, etc. should be collected.

As all data could not be obtained from single source, it has to be assimilated from the respective sources and prepared towards enabling decision making. Several sensors such as Soil moisture sensor FC-28 are also used to collect the real time field related parameters. In this phase the data is collected from the sources mentioned and passed for the next preprocessing stage.

3.2 PREPROCESSING PHASE

The data obtained from different sources are mostly in different formats. Some of these data may be structured, while the others may be unstructured. The inputs from expert systems and reports are usually in text form. Occasionally there may be missing values. The data should be cleaned before being provided for decision support. In the preprocessing phase, the invalid data is removed, and the unwanted data is filtered and converted into a structured format. Similarly, the existing conditions of the agricultural field are obtained using sensors. Some of the inputs are taken from inbuilt sensors in the mobile phones. This phase will remove the invalid data, filter the unwanted data and convert them into a structured format. This phase will perform the cleaning and pre-processing work. The output of the phase will consist of set of datasets used for making decisions.

3.3 DECISION SUPPORT PHASE

Decision tree learning is one of the predictive modelling approaches used in statistics, data mining, and machine learning.

The present weather and soil conditions are given as inputs to the proposed system. These inputs are obtained through sensors. The decision support system is developed using three decision tree algorithms. The ID3, C4.5 and Naïve Bayes algorithms are used to generate the decisions on crop selection for cultivation.

3.3.1 ID3 Algorithm Steps:

ID3 [20] stands for Iterative Dichotomiser 3 and is named so because the algorithm iteratively (repeatedly) dichotomizes (divides) features into two or more groups at each step. It was invented by Ross Quinlan and uses a top-down greedy approach to build a decision tree.

Step 1: Calculate the Information Gain of each feature.

Step 2: Considering that all rows don’t belong to the same class, split the dataset S into subsets using the feature for which the Information Gain is maximum.

Step 3: Make a decision tree node using the feature with the maximum Information gain.

Step 4: If all rows belong to the same class, make the current node as a leaf node with the class as its label.

Step 5: Repeat for the remaining features until we run out of all features, or the decision tree has all leaf nodes.

3.3.2 C4.5 Algorithm Steps:

Step 1: C4.5 is collection of algorithms for performing classifications in machine learning and data mining. It develops the classification model as a decision tree [20].

Step 2: Check if algorithm satisfies termination criteria

Step 3: Computer information-theoretic criteria for all attributes

Step 4: Choose best attribute according to the information-theoretic criteria

Step 5: Create a decision node based on the best attribute in step 3

Step 6: Induce (i.e. split) the dataset based on newly created decision node in step 4

Step 7: For all sub-dataset in step 5, call C4.5 algorithm to get a sub-tree (recursive call)

Step 8: Attach the tree obtained in step 6 to the decision node in step 4

Step 9: Return tree

3.3.3 Naive Bayes Algorithm Steps:

Naive Bayes [21] is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms.

Step 1: Calculate the prior probability for given class labels

Step 2: Find Likelihood probability with each attribute for each class

Step 3: Put these value in Bayes Formula and calculate posterior probability.

Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

3.4 OUTCOME AND EVALUATION PHASE

The decisions obtained are evaluated in this phase by furnishing the decisions obtained to the farmers for validation. The farmers are randomly selected from villages in the Madurai district. The recommendations of the farmers are taken into account for evaluation. In addition to this the execution time is also considered for evaluation.

4. RESULTS AND DISCUSSION

The details about data being collected and considered for the proposed work are described in this section. The proposed system is implemented using Weka tool. The performance metrics and the results are analyzed in this section.

4.1 DATA COLLECTION

The data used for the decision support system are drawn from different web portals listed below. Different parameters are considered from these inputs.

Table.1. Data Collected Sources

Reference	Remarks
[18]	Paddy Expert system
[19]	Rice statistics
[14]	Production and Irrigation statistics
[15]	Guidelines for Rice
[16]	Weather Archive
[17]	Water related data
[13]	Business value of Agriculture – Exports by India

4.2 SCOPE OF THE STUDY

The scope of the research work is restricted to the city Madurai in Tamil Nadu. Since the parameters need to be considered for decision support in crop selection for cultivation is large, the scope of the study is restricted to Madurai. The parameters considered for the experimental study includes weather, soil parameter, guidelines from the Tamil Nadu agriculture portal, global land temperature, fertilizers, type of sowing adapted, sensor data, etc. The refined dataset used to draw decisions using the aforementioned three decision tree algorithms contains 1036 instances.

4.3 PERFORMANCE EVALUATION METRICS

The performance of the proposed work is analyzed using metrics like accuracy and execution time. The decisions obtained are validated with a few set of small and medium category farmers around the Madurai district. The crops suggested by the proposed DSS are submitted to the farmers for evaluation. Based on their remarks, the accuracy is calculated using the confusion matrix.

Predicted values	Actual Values: Positive	Actual Values: Negative
Positive	TP	FP
Negative	FN	TN

Fig.2.Confusion Matrix

The accuracy of the decisions drawn from the proposed research work is analyzed using the following formula [5].

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

where,

True positive (TP): Number of correct crop selection that are suitable for cultivation.

False positive (FP): Number of incorrect crop selection that are suitable for cultivation.

True negative (TN): Number of correct crop selection that are not suitable for cultivation.

False negative (FN): Number of incorrect crop selection that are not suitable for cultivation.

The confusion matrixes of the algorithms used are listed in Fig.3 – Fig.5.

Predicted values	**Actual Values** Positive	Negative
Positive	875	72
Negative	10	79

Fig.3. Confusion Matrix of NB Algorithm

Predicted values	**Actual Values** Positive	Negative
Positive	918	36
Negative	5	77

Fig.4. Confusion Matrix of C4.5 Algorithm

Predicted values	**Actual Values** Positive	Negative
Positive	892	62
Negative	10	72

Fig.5. Confusion Matrix of ID3 Algorithm

Similarly, the efficiency of any proposed system depends upon the time taken for execution. It is described in the following section.

4.4 PERFORMANCE ANALYSIS

Every Decision support system is accepted by the stakeholders only if it produces good decision support. The accuracy of the proposed work is analyzed for the three algorithms used to provide the decisions. The accuracy level of the C4.5 is higher than that of the others. It is shown in Fig.6.

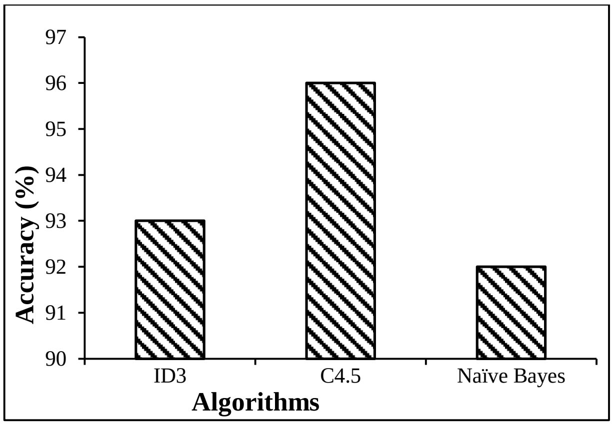

Fig.6. Accuracy

The observations does not mean that C4.5 always offers better accuracy when compared to that of ID3 and Naïve Bayes. Depending upon the domain and area of suitability the performance of the decision tree algorithms vary. For the domain considered in this research work, C4.5 is best suitable and hence offers better accuracy.

The Fig.7 depicts the analysis of preprocessing time; execution time and overall time taken for getting the decisions form the proposed system. The time taken for preprocessing the data is constant. Whereas the time taken for execution varies with respect to the algorithm employed. Similarly, the percentage of time spent for execution is visualized using the following figure. The C4.5 is consuming least percentage of execution time for the respective domain.

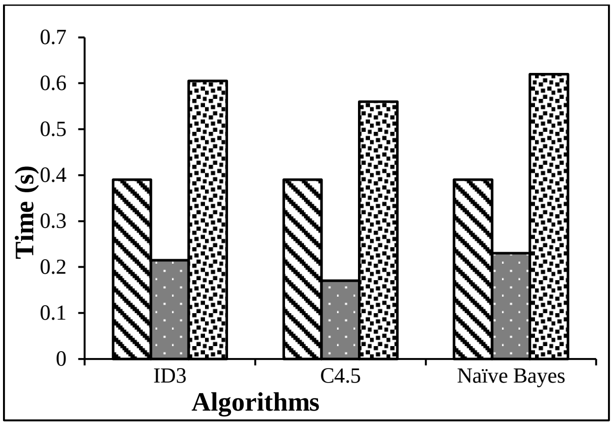

Fig.7. Execution Time (s)

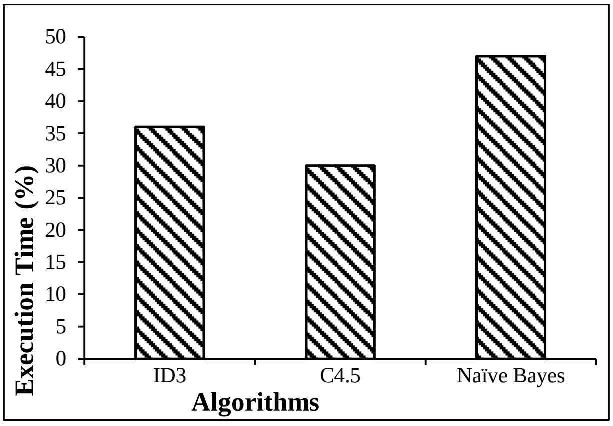

Fig.8. Percentage of Time taken for Execution

5. CONCLUSION

In this paper, a technique is devised to support the farmers in making quick and wise decisions with respect to the changing environmental conditions. In the process, this research work has made the following contributions:

- Agricultural parameters are collected from different sources for the Madurai region;
- The information collected are validated and used for decision making;
- A novel region specific decision support system for crop selection and cultivation is designed.
- The validity of the decision obtained are tested through the agricultural practitioners.

The parameters used to draw the decisions are gathered from various online agricultural portals. The scope of the study is restricted to the Madurai city. The data acquired using sensors are fed as inputs to the system. Three decision tree algorithms are employed for implementing the proposed system. The decisions obtained are analyzed in terms of accuracy and execution time. C4.5 is best suitable for the aforesaid domain, which is a modern farming technique that uses research data of soil characteristics, soil types, crop yield data collection, and suggests the farmers the right crop as per their region-specific parameters.

This reduces the wrong choice on a crop and increases in productivity. The crop suggested for cultivation will produce greater yield. The system can be further extended to include automated smart farm. The inputs from the experts can also be incorporated with feedback from the agricultural practitioners. This study can further be extended to other regions and other crops also.

REFERENCES

[1] R. Wanjari, K.G. Mandal, P. Ghosh, T. Adhikari and N.H. Rao, "Rice in India: Present Status and Strategies to Boost Its Production Through Hybrids", *Journal of Sustainable Agriculture*, Vol. 28, No. 2, pp. 19-39, 2006.

[2] K. Thiyagarajan and R. Kalaiyarasi, "*Status Paper on Rice in Tamilnadu*", Rice Knowledge Management Portal Publisher, 2010.

[3] R. Rupnik, M. Kukar, P. Vracar and Z. Bosnic, "AgroDSS: A Decision Support System for Agriculture and Farming", *Computers and Electronics in Agriculture*, Vol. 161, pp. 260-271, 2019.

[4] G. Yogeswari and A. Padmapriya, "Recommender System for Nutrient Management Based on Precision Agriculture", *International Journal of Recent Technology and Engineering*, Vol. 8, No. 4, pp. 1-12, 2019.

[5] Claudio Stockle, Marcello Donatelli and Roger Nelson, "CropSyst, A Cropping Systems Simulation Model", *European Journal of Agronomy*, Vol. 18, No. 3-4, pp. 289-307, 2003.

[6] G. Yogeswari and A. Padmapriya, "Precision Data Acquisition and Analysis for Nutrient Management of Tomatoes", *Asian Journal of Computer Science and Technology*, Vol. 8, No. 2, pp. 20-23, 2019.

[7] J. Lindblom and M. Ljung, "Next Generation Decision Support Systems for Farmers: Sustainable Agriculture through Sustainable IT", *Proceedings of 11th European Symposium on International Farming Systems Association*, pp. 1-6, 2014.

[8] R. Ruba Mangala and A. Padmapriya, "Visualizing the Impact of Climatic Changes on Pest and Disease Infestation in Rice", *International Journal of Recent Technology and Engineering*, Vol. 8, No. 3, pp. 8413-8421, 2019.

[9] M. Abedinpour, A. Sarangi, T.B.S. Rajput, M. Singh and T. Ahmad, "Performance Evaluation of Aqua Crop Model for Maize Crop in a Semi-Arid Environment", *Agricultural Water Management*, Vol. 110, pp. 55-66, 2012.

[10] R.R. Mangala and A. Padmapriya, "Prediction Based Agro Advisory System for Crop Protection", *Proceedings of International Conference on Intelligent Data Communication Technologies and Internet of Things*, pp. 1-7, 2019.

[11] Venkatalakshmi Balakrishnan, "Decision Support System for Precision Agriculture", *International Journal of Research in Engineering and Technology*, Vol. 3, No. 19, pp. 849-852, 2014.

[12] Guidelines for Rice, Available at: http://www.knowledgebank.irri.org/decision-tools/rice-doctor, Accessed on 2020.

[13] Production and Irrigation Statistics, Available at: https://visualize.data.gov.in/all_visualization, Accessed on 2020.

[14] Water Related Data, Available at: https://www.indiawaterportal.org/datafinder, Accessed on 2020.

[15] Paddy Expert Systems, Available at: http://agritech.tnau.ac.in/expert_system/paddy/Index.html, Accessed on 2020.

[16] Cropstaticsrice, Available at: https://www.farmer.gov.in/cropstaticsrice.aspx, Accessed on 2020.

[17] Yaser Sakkaf, "Decision Trees: ID2 Algorithm Explained", Available at: https://towardsdatascience.com/decision-trees-for-classification-id3-algorithm-explained-89df76e72df1, Accessed on 2020.