



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: III Month of publication: March 2019

DOI: <http://doi.org/10.22214/ijraset.2019.3436>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Optimum Crop Prediction using Data Mining and Machine Learning techniques

Prof. Amol Pande¹, Sarvan Purohit², Shruti Jadhav³, Krisha Shah⁴

¹Head of Department, Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai.

^{2,3,4}Datta Meghe college of Engineering, Navi Mumbai

Abstract—In this paper, we develop a web based application system that will assist farmers in selecting the crop that is most suitable for his land. This system is better than other systems as it considers various parameters like soil color, soil moisture, soil ph value, season, rainfall, temperature all at once unlike other systems which consider only two or three parameters but not all at once. Based on all these parameters the system will predict the best crop for the farmer using machine learning approach, decision tree (ID3 algorithm).

Keywords—parameters, machine learning, decision tree, ID3 algorithm

I. INTRODUCTION

The primary occupation of India is agriculture. Agriculture is the broadest economic sector and is the main source of income for many people in India. Agriculture is dependent on many climate and economy factors. These factors include soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall. But the problem in today's farming culture is lack of knowledge about the climatic trends and patterns, scientific knowledge about the soil that will benefit the crop yield. Optimum crop prediction system aims at taking all these factors into consideration and provide a central platform for farmers to analyse a optimum solution for their crop yield. This system will focus on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using ID3 algorithm.

II. EXISTING SYSTEM

All the existing systems consider only a single parameter or at the most two to three parameters and based on these two three parameters they give the prediction of crop yield. Also many of the existing systems just provide the information about crops and not the prediction about crop yield. Hence a system that provides not only information about crops to the farmers but also provide the prediction of the best suitable crop for a particular land needs to be developed.

III. PROPOSED METHODOLOGY

A. Data Collection and Data Filtering:

Data used in this system is collected on ground reference basis and the dataset was filled manually considering all the combinations of the parameters. Then data filtering was done based on the conditions required by the crops for cultivation. The rows with the condition that were not required by any of the crop were removed. The Table 1 shows the parameters on which crop prediction will be done and the Figure 2 gives the instance of the database used in the system.

| |
|---|
| Following are the parameters of the table : |
| Soil_color |
| Soil_moisture |
| Soil_ph |
| Season |
| Rainfall |
| Irrigation |
| Crops(Target Attribute) |

Table1: Parameters

| Gr. No. | Soil Colour | Soil Moist | Soil PH | Season | Rainfall in mm | Irrigat ⁿ %/hA | Temp erest ⁿ °C | Crops ? |
|---------|-------------|------------|--------------------|--------|----------------|---------------------------|----------------------------|---------|
| 01 | Black | dry | Neu | Kharif | very high | Low | Normal | wheat |
| 02 | Black | Med | Neu | Rabi | Medium | Normal | Normal | wheat |
| 03 | Black | wet | Neu | Summer | High | Low | Hot | Rice |
| 04 | Black | V.dry | Acid | Rabi | Low | Low | Hot | Jawar |
| 05 | R/Y | Med | Neu | Kharif | Very Low | Normal | Hot | Jawar |
| 06 | R/Y | wet | Acid | Rabi | High | Normal | Normal | Rice |
| 07 | R/Y | V.wet | Neu | Kharif | Very High | Normal | Normal | Rice |
| 08 | R/Y | Dry | Neu | Summer | Low | Low | Hot | Mangoes |
| 09 | R/Y | Med | Acid | Summer | Medium | Low | Hot | Mangoes |
| 10 | Laterite | Med | Neu | Kharif | High | Low | Normal | Rice |
| 11 | Laterite | wet | Neu | Summer | V.High | Low | Hot | Rice |
| 12 | Black | wet | Alkal ⁿ | Kharif | V.Low | High | Normal | wheat |
| 13 | Black | Dry | Alkal ⁿ | Rabi | Medium | Normal | Normal | wheat |
| 14 | Laterite | V.wet | Alkal ⁿ | Kharif | High | Low | Hot | Rice |
| 15 | Laterite | V.wet | Alkal ⁿ | Rabi | Very High | Low | Normal | Rice |
| 16 | Laterite | Dry | Neu | Summer | Low | Normal | Hot | Mangoes |
| 17 | R/Y | Dry | Neu | Kharif | Low | Normal | Hot | Jawar |
| 18 | R/Y | Med. | Acid | Rabi | Very low | High | Normal | Jawar |
| 19 | Laterite | Dry | Acid | Summer | Very low | Normal | Hot | Mangoes |
| 20 | Black | Med | Neu | Rabi | Very low | High | Normal | wheat |
| 21 | Black | wet | Neu | Kharif | Medium | Normal | Normal | wheat |
| 22 | Black | Dry | Neu | Rabi | Low | High | Hot | wheat |
| 23 | Black | Med | Acid | Kharif | Low | High | Hot | Jawar |
| 24 | R/Y | V.dry | Neu | Kharif | Very low | High | Hot | Jawar |
| 25 | R/Y | Med | Alkal ⁿ | Summer | Medium | Low | Hot | Mangoes |
| 26 | R/Y | Med | Alkal ⁿ | Summer | V.Low | High | Normal | Rice |
| 27 | Laterite | Med | Neu | Rabi | Low | High | Normal | Rice |
| 28 | Laterite | Dry | Neu | Summer | Low | Normal | Normal | Mangoes |
| 29 | Laterite | Dry | Neu | Summer | V.Low | V.High | Hot | Mangoes |
| 30 | Laterite | Dry | Alkal ⁿ | Summer | V.Low | High | Hot | Mangoes |

Figure 2: Instance of database used

B. Entropy Calculation:

Entropy in ID3 gives the amount of uncertainty in the dataset. Entropy is calculated using the formula:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

where,

S is the current dataset

i is set of classes in S

p is proportion of the number of elements in class to the number of elements in set S

Using this formula entropy is calculated for every attribute of the above mentioned table.

C. Calculation of Gain:

Information gain indicates how much uncertainty in dataset S was reduced after splitting the dataset S on the attribute A. Information gain is calculated using the formula:

$$IG(S, A) = H(S) - \sum_{t \in T} p(t)H(t) = H(S) - H(S|A).$$

where,

H(S) is the entropy of set S

T is subsets created from splitting set by attribute

Using the above formula information gain calculation is shown in figure 3:

| Gain wrt (S) | Value |
|---------------|--------|
| Soil colour | 0.7414 |
| Soil moisture | 0.6569 |
| Soil PH | 0.2118 |
| Season | 0.6911 |
| Rainfall | 0.6962 |
| Irrigation | 0.1081 |
| Temperature | 0.2131 |

Figure 3: Gain calculation

D. Decision Tree formation

Decision tree is one way to display the algorithm that has conditional statements. Root node of the decision tree is selected as the attribute having maximum gain. Considering the above example, gain of soil colour is maximum. Hence soil colour is selected as the root node for the above dataset.

Similarly all the child nodes are determined and we get the decision tree as shown in Figure 4:

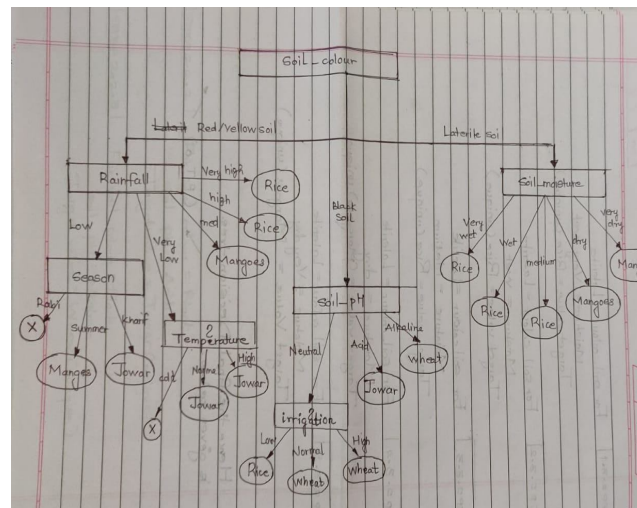


Figure 4: Decision tree

E. Error Percentage:

Training error is the error that you get when the training model is run on the dataset which was used to train the model that is the training dataset. Testing error arises when the training model is run on the datasets to which it is not exposed previously. This helps to measure the accuracy of the system. If the training error is low and testing error is high, you are likely overfitting your training data.

IV. IMPLEMENTATION

In the above example, ID3 is applied just to 30 tuples. However the dataset used in the system is based on the climatic conditions of Maharashtra and it consists of 2683 tuples. Out of these 2683 tuples, some of the tuples are used as training

dataset and remaining tuples are used as testing dataset. After applying ID3 algorithm to this dataset, we get a rule base. This rule base would be provided as output rules at the backend. The user has to enter the “district” where his land is located as input. This input will follow the rule base and based on this rule base and input from user the crop will be predicted. These rules are used to classify the crops to a particular district of Maharashtra.

Figure 5 shows the rule base for fruits:

```
[13 Bananas / 150 Grapes / 30 Mangoes / 150 Oranges]

| Rainfall = High: [6 Bananas / 0 Grapes / 10 Mangoes / 150 Oranges]
| | Season = Kharif: [0 Bananas / 0 Grapes / 0 Mangoes / 66 Oranges]
| | Season = Rabi: [6 Bananas / 0 Grapes / 0 Mangoes / 80 Oranges]
| | | Soil_ph = Acidic: [0 Bananas / 0 Grapes / 0 Mangoes / 28 Oranges]
| | | Soil_ph = Alkaline: [0 Bananas / 0 Grapes / 0 Mangoes / 26 Oranges]
| | | Soil_ph = Neutral: [6 Bananas / 0 Grapes / 0 Mangoes / 26 Oranges]
| | Season = Summer: [0 Bananas / 0 Grapes / 10 Mangoes / 0 Oranges]
| | Season = kharif: [0 Bananas / 0 Grapes / 0 Mangoes / 4 Oranges]
| Rainfall = Low: [0 Bananas / 76 Grapes / 10 Mangoes / 0 Oranges]
| | Irrigation = High: [0 Bananas / 44 Grapes / 0 Mangoes / 0 Oranges]
| | Irrigation = Low: [0 Bananas / 0 Grapes / 6 Mangoes / 0 Oranges]
| | Irrigation = Medium: [0 Bananas / 32 Grapes / 4 Mangoes / 0 Oranges]
| | | Soil_color = Black: [0 Bananas / 16 Grapes / 0 Mangoes / 0 Oranges]
| | | Soil_color = Laterite: [0 Bananas / 16 Grapes / 2 Mangoes / 0 Oranges]
| | | Soil_color = Red_yellow: [0 Bananas / 0 Grapes / 2 Mangoes / 0 Oranges]
| Rainfall = Medium: [7 Bananas / 74 Grapes / 10 Mangoes / 0 Oranges]
| | Season = Kharif: [0 Bananas / 38 Grapes / 0 Mangoes / 0 Oranges]
| | Season = Rabi: [7 Bananas / 0 Grapes / 0 Mangoes / 0 Oranges]
| | Season = Summer: [0 Bananas / 36 Grapes / 10 Mangoes / 0 Oranges]
| | | Irrigation = High: [0 Bananas / 20 Grapes / 0 Mangoes / 0 Oranges]
| | | Irrigation = Low: [0 Bananas / 0 Grapes / 6 Mangoes / 0 Oranges]
| | | Irrigation = Medium: [0 Bananas / 16 Grapes / 4 Mangoes / 0 Oranges]
```

Figure 5: Rule base for fruits

Similarly all the the other rule base are obtained an applied to backend. The Figure 6 shows the training and testing error analysis of every entity along with the depth of the decision tree.

Figure 6: Error analysis

| Entity | Depth | Error(train) | Error(test) |
|------------|-------|--------------|-------------|
| Fruits | 3 | 0.034985 | 0.210526 |
| | 4 | 0.029155 | 0.252632 |
| | 5 | 0.023324 | 0.294737 |
| Vegetables | 3 | 0.214575 | 0.231034 |
| | 4 | 0.183536 | 0.248276 |
| | 5 | 0.164642 | 0.251724 |
| Cash crop | 3 | 0.165680 | 0.259804 |
| | 4 | 0.127219 | 0.323529 |
| | 5 | 0.106509 | 0.352941 |
| Grains | 3 | 0.029155 | 0.252632 |
| | 4 | 0.185656 | 0.239914 |
| | 5 | 0.221780 | 0.197342 |
| Overall | 3 | 0.546256 | 0.743066 |
| | 4 | 0.446822 | 0.735766 |
| | 5 | 0.406345 | 0.728672 |

Based on this table, the following can be concluded:

- As the depth increases the Error(train) decreases thereby stating that the training is improved due increase in depth of the decision tree and vice-versa.
- As the depth increases the Error(test) increases because of the repetition of the attribute in the tree leading some false counts in test cases and vice-versa. (only exception is overall entity)
- Error rate of individual entity is less as compared to the overall entity, which states that due to increase in the number of target attribute the accuracy of the decision making tree is degraded.
- Error rate at depth=4 acts as the Average error rate for depth=3 and depth=5,Hence we can say that depth=4 acts as the average case in decision making and prediction.

So, depth=4 is an ideal state to be considered in decision tree generation for our system.

V. CONCLUSION

Use of technology in agriculture can change the situation of decision making and farmers can yield in better way. Optimum crop prediction system with the help of ID3 algorithm and data mining will help the farmer to make the best choice for crop to be cultivated. Also this system will provide the farmer with prediction of many crops and not just one crop. With this prediction, he can not only grow the crop of his choice but also utilise the space between the plants to grow other plants which he can afford to grow in the space between the plants. This will prove more profitable for the him. When the hard work of the farmer is supported by the analytical knowledge of machine learning technology, it will be beneficial for him to achieve economic balance in the society.

REFERENCES

- [1] Impact of Climate Change on Yields of Major Food Crops in India: Implications for Food Security Pratap S. BIRTHAL, Md. Tajuddin Khan, Digvijay S. Negi and Shaily Agarwal Agricultural Economics Research Review Vol. 27 (No.2) July-December 2014 <http://ageconsearch.umn.edu/bitstream/196659/2/1-PS-Birthal.pdf>
- [2] Predictive ability of machine learning methods for massive crop yield prediction Alberto Gonzalez-Sanchez, Juan Frausto-Solis, Waldo Ojeda-Bustamante Vol 12, No 2 (2014). <http://dx.doi.org/10.5424/sjar/2014122-4439>
- [3] A Survey on Data Mining Techniques in Agriculture M.C.S.Geetha Assistant Professor, Dept. of Computer Applications, Kumarakuru College of Technology, Coimbatore, India. Vol. 3, Issue 2, February 2015 <http://ijid.informaticspublishing.com/index.php/ijid/article/download/110845/77859>
- [4] Enhanced Crop Yield Prediction and Soil Data Analysis Using Data Mining Ms.P. Kanjana Devi Research Scholar in PG Dept. of Comp. Science Kamalam College of Arts and Science, Anthiyur, Udumalpet. Volume 4, Issue 6, December, 2016. http://www.ijmcs.info/current_issue/IJMCS161204.pdf
- [5] Analysis of agriculture data using data mining techniques: application of big data. Majumdar et al. J Big Data (2017). <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0077-4>
- [6] Prediction of Crop Yield using Machine Learning Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L Professor, Dept. of Computer Engineering, VESIT college, Maharashtra, India. Volume: 05 Issue: 02 | Feb-2018. <https://www.irjet.net/archives/V5/i2/IRJET-V5I2479.pdf>
- [7] Prediction of crop yield using regression analysis V. Sellam and E. Poovammal Department of Computer Science and Engineering, SRM University, Kattankulathur – 603203, Tamil Nadu, India. Indian Journal of Science and Technology, Vol 9(38), 10.17485/ijst/2016/v9i38/91714, October 2016 https://www.google.co.in/url?sa=t&source=web&rct=j&url=http://www.indjst.org/index.php/indjst/article/viewFile/91714/73940&ved=2ahUKEwj3OMTf_ajeAhWCE4gKHZlqAvQOFjAJegOIBRAB&usq=AOvVaw2yebzhok-7wzNS9jta93te&csid=1540724532864
- [8] Agriculture Analysis for Next Generation High Tech Farming in Data Mining P. Vinciya, Dr. A. Valarmathi IJARCSSE, vol. 6, Issue 5, 2016. <https://drive.google.com/file/d/1SpjYcns9fYhNnXlYJvtLaJ60FycmpxD8/view?usp=drivesdk>
- [9] Agricultural production output prediction using Machine Learning Classification techniques Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty Department of Computer Science and Engineering BRAC University, Dhaka, Bangladesh (IEEE - 2017). <https://ieeexplore.ieee.org/document/8016196>
- [10] Analysis of soil behavior and prediction of crop yield using Classification approach Monali Paul, Santosh K. Vishwakarma, Ashok Verma, GGITS, Jabalpur (2015) <https://ieeexplore.ieee.org/document/75461>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)