CROP ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES

[1]Mr.Sandeep Kumar Hegde, [2]Rajalaxmi Hegde,[3]Thanmaya.S, [4]Kiran.N, [5]Vivek Kumar
[1]Assistant Professor, [2]Assistant Professor, [3]Student, Student, [4]Student, [5]Student
[1]Computer Science and Engineering,
[1]NMAM Institute of Technology, Karkala, India

Abstract: This paper solves the problem which arises in the production of crops by analyzing the various factors using data mining techniques. This system gathers information about the crops that are cultivated from the different place around the states. Farmers have the accessibility to take help of this system to predict the best crop available and the ongoing crop details so that they can have maximum profit. This system uses various criteria such as place, population, crop type, soil type, stock, current requirement, season, no of farmers cropping the same crop, crop duration etc. to predict the crop for farmers.

***Index Terms–*: Crop prediction, ID3 algorithm, Profit prediction, Crop management, Stock management, Area management, Farmer management, Data sets.**

I. INTRODUCTION

A plant or animal product that is harvested and grown extensively for the benefits and profits can be referred to as crop. Crops can be referred to the harvested parts or the harvest itself in a more refined state. Majority of the crops are cultivated in agriculture and aquaculture. Prediction means that timely advice to predict the crop future productivity and analysis made to help farmers to have the maximum crop yield in terms of production. In agricultural problems, Data Mining is used extensively. To analyze large data sets and establish useful pattern and classification in the data set is analyzed by Data Mining. Extracting the information from the set of data and transforming it into an understandable structure for further use is Data Mining process. Farmers are the backbone of our country. They live in villages and cultivate varies crops under different categories such as vegetables, flowers, dry fruits, and fruits. Sometimes, it is hard for farmers to get back the invested money from the crop because of the market value of the crop when it comes to the market. Farmers will not know the market value until they bring their crop to the market which makes the farmer struggle a lot. This crop rates can be controlled if there is a system which helps the farmers to know about the crops planted around his location with the prediction to suggest the farmers about the crop to be cultivated to maximize profit.

This will help to ensure food security of common folk and a fair earning for the farmer. Unpredictable climate and market can create a major impact on food security. On the other hand, many farmers suffer a huge loss for the very same reason. Every farmer tries to predict the market and act accordingly. But with limited information, there is a higher chance that they might come to a wrong conclusion. This leads to uneven production in terms of variety in the market. Our system gathers all the data about the crops that are being cultivated around different places. Farmers have the Farmers have the accessibility to take help of this system to predict the best crop available and the ongoing crop details so that they can have maximum profit. This system uses various criteria such as place, population, crop type, soil type, stock, current requirement, seasons, no of farmers cropping the same crop, crop duration etc. to predict the crop for farmers. We believe that our proposed system will help in planting decisions to farmer increase their potential for revenue. Also, it will help even the production of crops in terms of variety. This system will help to achieve the appropriate food reserve as well. It will help to plan the strategy to meet the demand when there is a low yield in production. Smart agriculture using the data mining technique can help farmers to increase their productivity to a noticeable extent. Weather forecast data obtained from reliable sources will help to determine the temperature and rainfall which will help to determine the crop productivity to a great extent. Our system will integrate all the data collected and then predict by analyzing those data collected.

II. RELATED WORK

The model to enhance the accuracy of precipitation prediction is explored. Linear Regression and Map scale back rule are used for prediction of the weather. Solely weather knowledge is employed for prediction.[1]. Focuses on several soil parameters like dampness, nitrogen, pH and additionally compared accuracy is conferred. Yield prediction is completed by Naïve Bayes, Apriori rule. The accuracy achieved [2]. Crop parameters and Climate data are used for prediction. In this J48 , random forests, Naïve Bayes, SVM are enforced. Soil parameters aren't considered[3]. To scale back extra plant food usage throughout cultivation of lands and to extend the soil vitality is planned by the author. Techniques of mining are used in farming and tested soil aspects to get reports that facilitate call support to gather procedures, crop rotation, and fertilizer needs. Here conditions like climate could be varied. correct prediction won't occur[4]. During this, there's a comparative study between DBSCAN and AGNES. MLR, DBSCAN, and Agnes are used. for every crop formula springs on an individual basis [5]. The building of a decision network for preciseness farming is planned by prediction rule. Hadoop Distributed FileSystem (HDFS) is employed. Here prediction is suitable for a crop of the given soil parameters however not the yield[6]. Images are captured. The cloud technology is used for communication. Image capturing is done by the use of IOT and for location tracking GPS. Image process techniques are focused[7]. The epitome of a decision network for rice crop yield prediction. A GUI is generated using the language Java and Net Beans tools for the farmer's ease and facilitate the choice manufacturers. Prediction of this paper is merely for rice crop[8]. In this work, a system is conferred using the mining techniques to forecast the soil dataset category being

analyzed. This category will specify crop yields as predicted. K-NearestNeighbour and Naïve Bayes strategies are used.ID3 will be used for continuous and categorical inputs however K-Nearest Neighbour is used for continuous price inputs[9]. Irrigation and crop sort parameters are most well-liked. Apriori and k means that rule is inexplicit. Intensive Computationally, particularly once the dimensions of the growing knowledge set[10].

Table 1.1: Literature Survey

Publications and Authors	Used techniques	Achieved parameters	Constraints
1. 2015, Thool.R.C	The algorithm used is LinearRegression and Map Reduce for weather forecasting.	This model improves the accuracy of rainfall forecasting	Weather data is used for forecasting
2.2016,Hemageethaa.N	The yield prediction is done byAppriori and Naïve Bayes algorithm	Focuses on several soil parameters like dampness, nitrogen and compared accuracy is conferred.	The accuracy achieved is 77%
3.2016,Isakki.P	Naïve Bayes, J48, random forests, ANN, SVM are implemented.	Crop parameters and Climate data are used for prediction	Parameters like soil are not considered
4.2015,M.P.Singh	Techniques of mining are used in farming and tested soil aspects to get reports that facilitate call support to gather procedures, crop rotation, and fertilizer needs. Here conditions like climate could be varied. correct prediction won't occur	To scale back extra plant food usage throughout cultivation of lands and to extend the soil vitality is planned by the author	Accurate prediction is not done. Variable climate condition
5. 2016,Chandra.N	MLR, DBSCAN, and Agnes are used	In this, there is a comparative study between DBSCAN and AGNES. MLR, DBSCAN, and Agnes are used.	Every crop has a different formula.
6.2015,SwetaBhattachrya	Hadoop Distributed FileSystem (HDFS) is used.	The building of a decision network for preciseness farming is planned by prediction rule. Algorithm	prediction is suitable for a crop of the given soil parameters however not the yield
7.2015,Jayasakthi.J	Image capturing is done by the use of IOT and for location tracking GPS	The cloud technology is used for communication	Image processing Techniques are focused more.
8. 2016,Nikita Gandi	. A GUI is generated using the language java and Net Beans tools for the farmer's ease and facilitate the choice manufacturers.	The epitome of decision network for rice crop yield prediction.	This papers prediction is only for rice crop.
9. 2015,Ashok Varma ,Monali Paula, Santosha K. Vishwa Karma,	K-NearestNeighbour and Naïve Bayes methods are used.	In this work, a system is conferred using the mining techniques to forecast the soil dataset category being analyzed. This category will specify crop yields as predicted.	K-Nearest Neighbour is used for continuous value inputs but ID3 can be used for continuous and categorical inputs.
10.2014,Geetha.R	Apriori and k means algorithm is implied.	Irrigation and crop type parameters are preferred.	Computationally intensive, especially when the size of the data set grows.

III. PROPOSED SYSTEM

In this segment, We will present the abstract model of our system and systematically explain the different aspects of it. Our system is divided into two interfaces. One being the farmer side and other being the admin. Here the farmer side presents crop prediction by analyzing the data given by the farmer. Admin side has the control of farmers database. We use the data collected to predict the necessary crop to be grown in a certain region by using the Id3 algorithm. The detailed explanation of these concepts is presented using the Data flow diagram shown in figure 1.

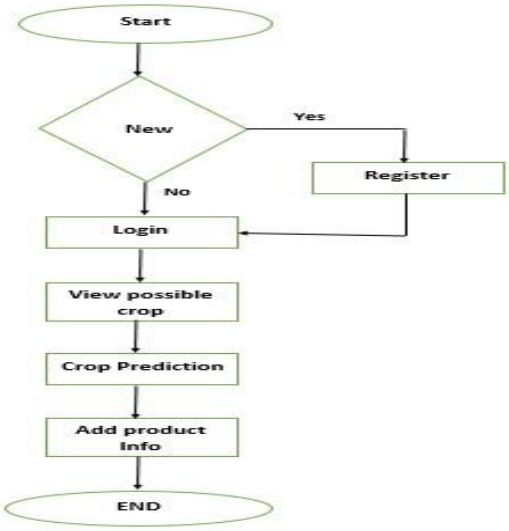

Fig 1: Data Flow Diagram of Farmer End

In figure 1 we have created a platform which will help the farmer to find the best and suitable crops for themselves. First, they will have to login into their account and if they don't have their own existing account then they will have to create one. Once they have successfully created an account, then they can access the feature of the platform to their use. Once the farmer logins into their account, they would be redirected towards their homepage from where they will have the option to go into those different pages, they very much desire. They have the option to view possible crops available for them to put into use for them. They will have the accessibility to view the crops along with other features which makes them the best suitable for it like the soil type, temperature, humidity, duration, water level. Then they have the option to add their opinion into the database about the different crops. Thus, when they do, it will help other farmers to have searched for their possible crops. Then they can log out from their respective account.

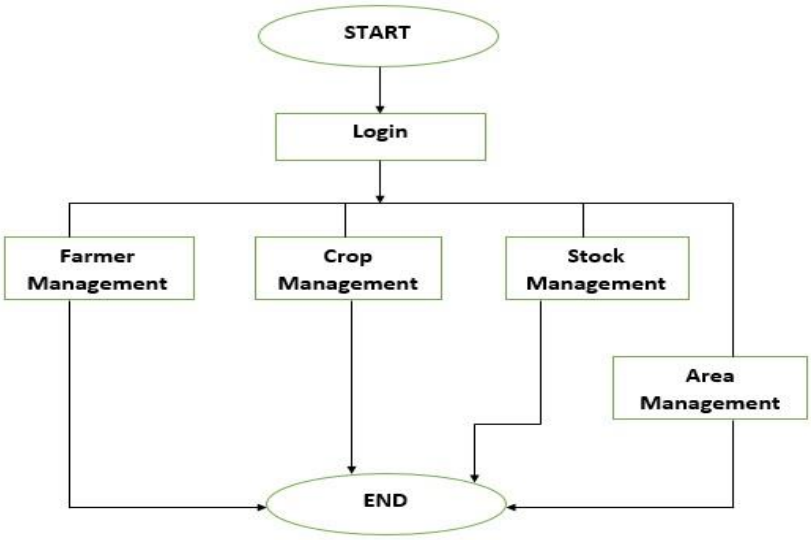

Fig 2: Data Flow Diagram of Admin End

As shown in figure 2 admin module, first the admin will log in using his credentials and then will have all the access to all the database and access to operations on the platform. He has access to the farmer database, where he can add, delete or update the farmers. Then he will have the access to crop database where he will have the access to the add, delete and update the different types of crops details. He will also have the access to manage the location database where he can add, delete or update specific information such as population, soil type and availability of water, etc. He has access to update different crops availability in stock management. Availability of crops helps the farmer to predict the crop market so that they can predict the prices.

Pseudo Code:

ID3 (Instances, Result, Parameters)
An admin root node for the farmer side can be created as the tree
If all Instances match with the existing database, then return that node to the root with label = +.
If Instances does not match with the existing database, then return that node to the root with label = -.
If the Parameters that are used to predict is returned empty, then return those nodes which have more the occurrence of the Result from the database of Instances
else
A -> Here this variable has values of Parameters that represents all the collected data stored in the database as Instances
A= This variable is stored with decision Parameter that is matched with the admin root node
Among the different probable values of A which is represented by v_i,
A new tree branch has to be added below the admin root node to check whether A= v_i,
Let Instances(v_i) have a subgroup of Instances with value v_i for values of A
If the values in Instances (v_i) is vacant
Then a new leaf node is added to the admin root node which has more occurrence of the Result from the database of Instances
else a new subtree ID3 branch is added below the admin root node,
ID3 (Instances (v_i), Result, Parameter – {A})
End
Return admin root node

MODULES:

Admin

- Login- In this the admin has the accessibility to login so that he can have access to various database and operation.
- Crop management – To add/delete/update different types of crops details to our system
- Famer management – To add / delete /update farmers
- Location management – To add/delete/update location-specific information such as population, soil type, availability of water etc.
- Stock management – To update different crop data in stocks

Farmer

- Registration- In the farmer database, who don't have the account, they can register to access the features of the platform.
- Login- Those farmers who have an existing account in the database, they can log in into the account and access the features in it.
- Crop registration – To register/update crop details whatever they will grow
- Crop prediction – Prediction of crop based on population, availability and season.
- Crop information – To know about information about every crop.

IV. RESULTS

The different algorithm has different predicting ability and with different predicting ability, we get a different result and plays a major role in selecting the right algorithm to implement. ID3 algorithm has the highest predicting ability and thus has an advantage over others when it comes to the selection of the algorithm. It has an accuracy of 95% in predicting the data. Then there are other algorithms which do the work like Random forests but it's the ability to predict the right data is less than that of ID3 algorithm. It has an accuracy of 83.5%. We also have Naïve Bayes algorithm which we may choose while predicting the crop and in Naïve Bayes, we have an accuracy of 88.2%. Well, Naïve Bayes have the most accuracy when we compare it with others but didn't consider ID3. Then we have SVM algorithm which also helps in predicting the crop. It has an accuracy of 86.7%. We may also choose the KNN algorithm which has an accuracy of 82.8%. If we look at all the algorithm and then we can see that the ID3 algorithm is the best algorithm when compared to others and therefore we choose the ID3 algorithm.

Table 4.1: Prediction accuracy of the different algorithm

Algorithm	Accuracy
ID3	95%
Random forests	83.5%
Naïve Bayes	88.2%
SVM	86.7%
KNN	82.8%

Table 4.1 above shows the predicting accuracy of the different algorithm available.

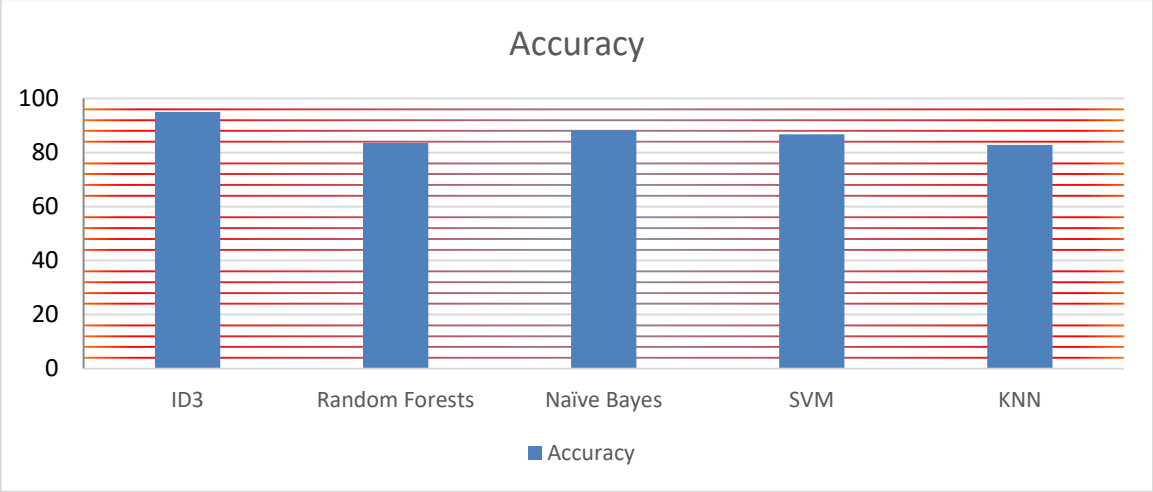

Fig 3: Accuracy Graphs of different algorithms

V. CONCLUSION

Our planned system presents a way to pick a sequence of crops to be planted over the season. This technique could improve improves the monetary standing of a farmer by predicting the simplest crop which will be cropped. The planned technique resolves choice of crop supported prediction yield rate influenced by parameters (e.g. weather, soil type, population, stock, crop sort and current crop). It takes ongoing crop details, stock details, plantation days and space as input and finds a crop whose production returns the most profit to the farmer.

VI. ACKNOWLEDGMENT

The satisfaction that accompanies the completion of any task would be incomplete without the mention of all the people, without whom this endeavor would have been a difficult one to achieve. Their constant blessings, encouragement, guidance, and suggestions have been a constant source of inspiration. First and foremost, my gratitude to my project guide, Mr. Sandeep Kumar Hegde for his constant guidance throughout the course of this project and for the valuable suggestions. I also take this opportunity to express a deep sense of gratitude to the project coordinators for their valuable guidance and support. I acknowledge the support and valuable inputs given by, Dr.Uday Kumar Reddy the HOD, CSE, NMAMIT, Nitte. My sincere thanks to our beloved principal, Dr.Niranjan N Chiplunkar for permitting us to carry out this project at our college and providing us with all the needed facilities. Finally, thanks to staff members of the Department of CSE and our friends for their honest opinions and suggestions throughout the course of our project.

REFERENCES

1.Bendre, M. R., Thool, R.C., Thool, V. R., “Big Data in Precision Agriculture: Weather Forecasting for Future Farming”, 1st International Conference on Next Generation Computing Technologies, pp.744-750, 2015.

2. Hemageetha, N., "A survey on application of data mining techniques to analyze the soil for agricultural purpose", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.3112-3117, 2016.

3. Sujatha, R., Isakki, P., “A study on crop yield forecasting using classification techniques”, International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp.1-4, 2016.

4. Rakesh Kumar, M.P. Singh, Prabhat Kumar, and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", Vel Tech RangarajanDr.Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 – 8 May 2015. pp.138-145

5. Ankalaki, S., Chandra, N., Majumdar, J., “Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Special Issue 2, pp.25-29, 2016.

6. Kushwaha, A.K., SwetaBhattachrya, “Crop yield prediction using AgroAlgorithm in Hadoop”, International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 5- No2, pp.271-274, 2015.

7. Gayatri, M.K., Jayasakthi, J., Anandha Mala, G.S., “Providing Smart Agricultural Solutions to Farmers for better yielding using IoT”, IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), pp.40-43, 2015.

8. Niketa Gandhi, Leisa J. Armstrong, OwaizPetkar “Proposed Decision Support System (DSS) for Indian Rice Crop Yield Prediction “.International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016) 13 978-1-5090-0615-1/16 2016 IEEE

9. Monali Paul, Santosh K. Vishwakarma, Ashok Verma.” Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach” 978-1-5090-0076-0/15 2015 IEEE

10. Fathima, G.N., Geetha, R., “Agriculture Crop Pattern Using Data Mining Techniques”, International Journal of Advanced Research in Computer Science and Engineering, Vol. 4, Issue 5, pp.781-786, 2014.