

Predictive Analysis of Crop Yield using Machine Learning Algorithm

Nitin Shettar¹ Dr. B. M. Sagar²

¹M. Tech Student ²Head of the Department

^{1,2}Department of Information Science & Engineering

^{1,2}RV College of Engineering®, Bengaluru-560059, India

Abstract— Agriculture has highly significant role in the growth of Indian economy that pressurizes it to seek modern scientific approaches in order to increase the yield. Problems of limited availability of resources like land, labor etc. have made it to impart technology to revitalize the economy of the country. Machine learning has emerged with multipronged opportunities to aid data intensive agri-technologies domain. This paper proposes framework that makes use of C 5.0, SVM, Naive Bayes machine learning algorithms to predict the agricultural productions as low, medium and high yield. In addition to this we are calculating accuracy and time taken as metrics for the particular yield.

Keywords: Machine Learning Algorithm, Agricultural Production

I. INTRODUCTION

Agriculture is the main economical source of India and more than 82% of farmers belong to agricultural domain. Quality and production of crops is decreasing now a day due to unsuitable climatic conditions to grow the crops. Achieving high yield of various crops is becoming a challenge to the farmers. Food security and high yield can be obtained by Crop Yield Prediction which can help the farmers to improve their yield and they can predict their yield before farming. Each crop cultivation varies that means different crops will use different climatic condition, different type of soil. Different regions are suitable to grow particular crop and through crop yield prediction farmers will get to know about parameters (rainfall, temperature, region, season, area) suitable to grow their crop and get high yield. IoT technology can reduce the cost and improve the profitability of traditional farming.

Reap proposition is starting at now in perspective on data assembled in field based rural examinations that get trim execution under the accumulation of conditions (eg., soil quality and trademark conditions). However, crop execution data assembled is as of now moderate, in that capacity crop contemplates are consistently endeavored in remote and dispersed zones, and such data are ordinarily accumulated physically. Besides, the nature of physically assembled item execution data is low, since it doesn't consider earlier conditions that have not been seen by the human administrators but instead is principal to filter through accumulated data that will incite invalid ends. Rising Internet of Things (IoT) progressions, for instance, IoT gadgets (eg., remote sensor systems, Organize associated climate stations, cameras, and PDAs) can be utilized to group immense measure of ecological and yield execution information, going from time arrangement information from the sensors, to spatial data from camera, to human perceptions gathered and recorded by means of portable PDA applications. Such information analyzed then be able to be investigated to sift through invalid information and process customized edit proposals for a particular farm

II. LITERATURE SURVEY

An instructive foundation needs an inexact earlier learning of selected understudies to anticipate their execution in future scholastics. This encourages them to distinguish promising understudies and furthermore gives them a chance to focus on and improve the individuals who might likely get lower grades. As an answer, we have built up a framework which can foresee the execution of understudies from their past exhibitions utilizing ideas of information mining strategies under Classification. We have dissected the informational index containing data about understudies, for example, gender, marks scored in the board examinations of classes X and XII, checks and rank in selection tests and results in first year of the past group of understudies. By applying the ID3 (Iterative Dichotomiser3) and C4.5 classification algorithms [1] on this data, we have anticipated the general and individual execution of crisply conceded understudies in future examinations.

There are several applications of Data Mining procedures in the field of farming. The specialist's implemented K-Means algorithm to forecast the contamination in the climate, the KNN is applied for recreating day by day precipitations and other climate factors and distinctive possible changes of the climate situations are analyzed using SVM. [3]

Human capital is of a high stress for associations organization where their most intrigue is in procuring the exceedingly qualified personnel which are required to perform significantly too. Recently, there has been a creating eagerness for the data mining an area, where the objective is the disclosure of discovering that is correct and of high preferred standpoint for customers. In this paper, data mining methodologies were utilized to develop a gathering model to predict the execution of specialists. To amass the request model the CRISP-DM data mining framework was grasped. Decision tree was the essential data mining tool used to manufacture the classification model, where a couple of classification rules were created.

Increasing crop intensity leads to increase in yield. For some of crops the same climatic condition will suits to get high yield so in that case growing multiple crops which utilizes same climatic condition results in high yield and high crop intensity. In few cases high rainfall and good climatic condition does not suits to grow few crops, sometimes even though the cropping intensity of few crops may be high but the yield result by them will be less. Due to this factor yield for particular crops is defined in terms of Kg per hectare. But it is important to consider all factors that affect crop growth like land, time, labor should be used to define yield of various crops.

Overall disease study reports that Stomach dangerous development is the third driving explanation behind malignancy demise in the two sexual orientations. The

destruction rate can be restricted in early desire for the proportionate. Different Data Mining computations are used to mine the significant information from the broad course of action of data. It is in like manner accepting a vital activity in various disease estimates with the guide of different figurings. This paper thinks the expectation of the savage infection Stomach Cancer using two extraordinary information mining calculations named C4.5 calculation and Naive Bayes calculation [5]. It likewise examinations the execution and the activity of both the calculations in forecast.

Yuan Guo [4] proposed a remote sensor based computerized water system framework to streamline water use for rural reason. The system comprises of distributed WSN of temperature and soil moisture sensors mounted in the product field. The agriculture insightful system was depended on IoT which is presented for organic melons and organic products creation and quality. A considerable lot of the advances were used as a part of the system. Such as sensors, RFID and etc.

III. METHODOLOGY

Crop management and agricultural production can be improved and estimated by predicting crop yield. Various machine learning algorithms and traditional approaches are becoming techniques and tools for predicting yield by agricultural scientists. For prediction all parameters must be considered because every parameter plays an essential part in development of the crop and predicting yield.

After collecting all needed information for prediction the data is pre-processed and analyzed using data analysis tools and features from this preprocessed data is collected and extracted using extraction module. Classification of crops is done to classify the data into training and testing sets. The prediction rules will be applied to training data set for predicting the maximum yield for various parameters. Data were collected from numerous sources namely: Data.gov.in, Directorate of economics and statistics, indiawaterportal.in, raitamitra.nic.in, Apmc Karnataka., data.world, kaggle.

A. Analysis:

It is checking variation of information because the distinctive sensor information comes at various time. This information has to be analyzed using Time Series analysis method additionally to be checked for errors, excess information or undesirable garbage information. It also applies Outlier Removal technique with the goal that information will be error free.

B. Pre-processing:

During the upload part all the incomplete records regarding crop data will automatically be removed and also others errors like spelling correction will automatically be handled. Some of the missing calculations can also be deleted with in this stage. All these pre-processing takes place during the import operation.

C. Naive Bayes Classifier:

Naive Bayes is a basic strategy for building classifiers: models that relegate class labels to issue occurrences, spoke to as vectors of highlight esteems, where the class labels are

drawn from some finite set. There is definitely not a single algorithm for training such classifiers, however a group of algorithm dependent on a common principle: all naive bayes classifiers assume that the estimation of a specific element is independent of the estimation of some other feature, given the class variable. In this module the classification of district is done which can be high yield, medium yield and low yield district based on the data sets which are available.

D. C5.0 Algorithm:

The Count No of instances of 1 and then Count No of instances of 2 is taken. After that Calculate the Overall Information Gain,

$$Information\ Gain = -\frac{P}{P+n} \log\left(\frac{P}{P+n}\right) - \frac{n}{P+n} \log\left(\frac{n}{P+n}\right)$$

Where,

P = count of class labels of 1

n = count of class labels of 2

Find the different attributes of each of the classes and then for each of the attribute information gain is computed. Compute the gain for each of the attribute and then entropy of the attributes is computed. The decision tree is generated with the root node being the parent. Find the classes for each of the data and then predict the class.

E. Support Vector Machine:

A Support Vector Machine (SVM) is a discriminative classifier formally described by a disconnecting hyperplane. As it were, given named preparing information (managed learning), the calculation yields a perfect hyperplane which sorts new points of reference. In two dimensional space this hyperplane is a line isolating a plane in two areas where in each class lay in either side.

The learning of the hyper plane in linear SVM is finished by changing the issue using some straight variable based math. This is the place the bit assumes job. For linear kernel the condition for forecast for another data using the dot product between the data and each support vector.

IV. RESULTS

Year	Month	Bajra	Bull	Cow	Ground Nut	He Buffalo	Jowar	Milch	Ram	Sajje	She Buffalo	She Goat	Sheep	Son Flower	Alan De Go	Soil PH	Rainfall	Goat	Cross Name
2012	1	242	65	151	164	11	4	233	181	1649	182348	1027	190	1	1	5.7	0	492	3
2012	1	242	75	111	164	21	5	233	181	1649	182348	1027	206	2	20	1.7	1	292	3
2012	1	242	85	121	164	31	6	233	181	1649	182348	1027	206	3	30	1.7	2	192	3
2012	1	242	95	131	164	41	7	233	181	1649	182348	1027	490	4	40	1.7	3	292	3
2012	1	242	105	141	164	51	8	233	181	1649	182348	1027	590	5	50	1.7	4	392	1
2012	1	242	115	151	164	61	9	233	181	1649	182348	1027	690	6	60	1.7	5	492	1
2012	1	242	125	161	164	71	10	233	181	1649	182348	1027	790	7	70	1.7	6	592	2
2012	1	242	135	171	164	81	11	233	181	1649	182348	1027	890	8	80	1.7	1	692	2
2012	1	242	145	181	164	91	12	233	181	1649	182348	1027	990	9	90	1.7	2	792	2
2012	1	242	155	191	164	101	13	233	181	1649	182348	1027	1090	10	100	1.7	3	892	3
2012	1	242	165	201	164	111	14	233	181	1649	182348	1027	1190	11	110	1.7	4	992	3
2012	1	242	175	211	164	121	15	233	181	1649	182348	1027	1290	12	120	1.7	5	1092	3
2012	1	242	185	221	164	131	16	233	181	1649	182348	1027	1390	13	130	1.7	6	1192	1
2012	1	242	195	231	164	141	17	233	181	1649	182348	1027	1490	14	140	1.7	7	1292	2
2012	1	242	205	241	164	151	18	233	181	1649	182348	1027	1590	15	150	1.7	8	1392	2
2012	1	242	215	251	164	161	19	233	181	1649	182348	1027	1690	16	160	1.7	9	1492	2
2012	1	242	225	261	164	171	20	233	181	1649	182348	1027	1790	17	170	1.7	10	1592	2
2012	1	242	235	271	164	181	21	233	181	1649	182348	1027	1890	18	180	1.7	11	1692	3
2012	1	242	245	281	164	191	22	233	181	1649	182348	1027	1990	19	190	1.7	12	1792	3
2012	1	242	255	291	164	201	23	233	181	1649	182348	1027	2090	20	200	1.7	13	1892	1
2012	1	242	265	301	164	211	24	233	181	1649	182348	1027	2190	21	210	1.7	14	1992	1

Fig. 1: Viewing Dataset

Figure 1 shows the dataset view after giving the dataset as an input to algorithm and in this 75% is training set, 25% is test set.

Fig. 2: UI for data input

Figure 2 shows the user interface provided for user to input all the attribute values to predict the crop yield and gives output to the user.

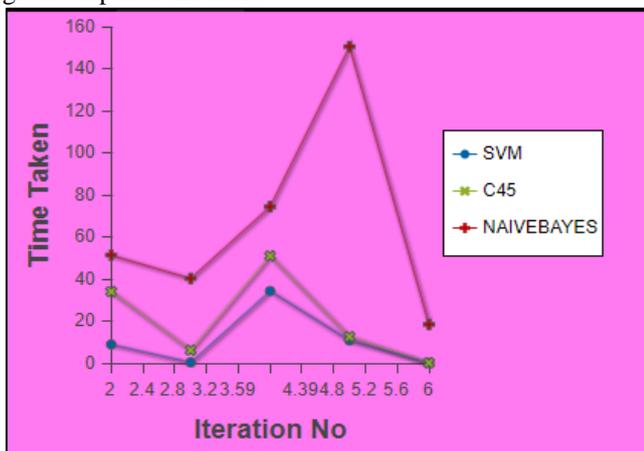


Fig. 3: Time comparison graph

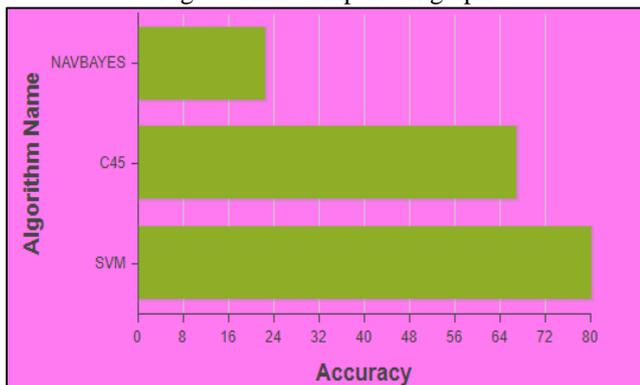


Fig. 4: Accuracy comparison graph

Figure 3 & 4 shows the metrics comparison graphs after completion of several iterations of algorithms.

V. CONCLUSION

Agriculture can be well organized by predicting yield. This yield is predicted using C 5.0, SVM, Naive Bayes algorithms in this project. By this the user can know the yield of the particular crop depending on the specific area. The project uses classification algorithms which helps to deal with the large amount of datasets in agricultural domain.

ACKNOWLEDGMENT

This project work is supported as part of major project for M.Tech, Software Engineering Post Graduate Program in the RV College of Engineering®, Bengaluru-560059.

REFERENCES

- [1] D Ramesh , B Vishnu Vardhan. "Data Mining Techniques and Applications to Agricultural Yield Data". International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue 9, September 2018,pp.3477-3480.
- [2] Mohammad Motiur Rahman, Naheena Haq and Rashedur M Rahman "Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield", 20th. IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, 2018.
- [3] LIU Dan, Cao Xin, Huang Chongwei, Ji Liang Liang, "Intelligent agent greenhouse environment monitoring system based on IOT technology", 2017 International Conference on Intelligent Transportation, Big Data & Smart City.
- [4] Joseph Haule, Kisangiri Michael, "Deployment of wireless sensor networks (WSN) in automated irrigation management and scheduling systems: a review", Science, Computing and Telecommunications(PACT), 2017, Pan African Conference.
- [5] Weimin Qiu, Linxi Dong, Haixia Yan, Fei Wang, "Design of Intelligent Greenhouse Environment Monitoring System Based on ZigBee and embedded technology", 2016 IEEE International conference.
- [6] Yuan Guo, "The Application of a ZigBee Based Wireless Sensor Network in the LED Street Lamp Control System", 2016, College of Automation & Electronic Engineering, Qingdao University of Scientific & Technology, Qingdao, China embedded technology, Consumer Electronics - China, 2014 IEEE International Conference.
- [7] D.K. Sreekantha, Kavya A.M.. "Agricultural crop monitoring using IOT - a study", 2017 11th International Conference on Intelligent Systems and Control (ISCO), 2017.
- [8] Piara Singh, N. P. Singh, K. J. Boote, S. Nedumarian, K. Srinivas, "Management options to increase groundnut productivity under climate change", Anantapur, Mahboobnagar and Junagadh, 2016.
- [9] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique" International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),2015.
- [10] Ramesh A Medar and Vijay S Rajpurohit " A survey on Data Mining Techniques for Crop Yield Prediction" International Journal of Advance Research in Computer Science and Management Studies Volume(2) 2014.
- [11] Ratchaphum Jaikla, Sansanee Auephanwiriyaikul and Attachai Jintrawet, "Rice Yield Prediction using a Support Vector Regression method", Proceedings of ECTI-CON 2008.