CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS

Guru Murthy B

Dept. of Computer Science and Engineering,

Coimbatore Institute of Engineering and Technology,

Tamil Nadu, India.

Karuppasamy M

Dept. of Computer Science and Engineering,

Coimbatore Institute of Engineering and Technology,

Tamil Nadu, India.

Shree Sakthi Keerthna S

Dept. of Computer Science and Engineering,

Coimbatore Institute of Engineering and Technology,

Tamil Nadu, India.

Sunmathi M

Dept. of Computer Science and Engineering,

Coimbatore Institute of Engineering and Technology,

Tamil Nadu, India.

ABSTRACT:Agriculture plays a vital role in the growth of the country's economy. But, over the past few years there is a huge development in various sectors due to swiftchanges in the technology. One such place where there is yet large reach to developis agriculture. The main purpose of its downside is not having a well-formeddesign about farming and proper instruction to the farmers. Agriculture whenmerged with technology can yield exceptional results. Farmers should be conscious of thenature of the field in order to produce best results. Of all the other stages inagriculture, production also plays a pivotal role. To attain various crop yieldforecast, various machine learning algorithms such as Decision tree, Artificialneural networks (ANN), Support vector machine (SVM), Multiple linear regressionshave been used. This paper focuses on a incisive work of various papers thatdeals with different techniques used to assess the crop yield. It aims at predictingthe yield of crop (Wheat) by observing and analyzing the datasets of foregoingyears of the crop. In addition, paper contours diverse existing techniques to inspect crop yield. It also establishes the disparity of various algorithms along with its merits and de merits.

Keywords—Artificial neural network, classification, prediction.

I. INTRODUCTION

Agriculture is thought to be an essential occupation of our nation. Production also plays a major role in it. At times the yield of crops is average. As a result, the urge of food production is increasing. Researchers are trying to make extra attempts for more production. In the past, farmers used to follow naked eye observation. This method of naked eye examining is difficult when large hectares of productions are considered. Hence in order to obtain great results in predicting the crop yield, some of the machine learning algorithms can be applied. It is the branch of computer science which deals with machine intelligence where an intelligent agent as a system, boosts its change of success by taking action. It is the study of concepts which empowers the computer to do the things that make people seem brilliant. It is the

study through knowledge, experience concerning some tasks and performance. It is a subgroup of artificial intelligence that has enabled the system to impulsively learn and improve. Planning, learning, communication, reasoning, knowledge and perception are some of the central principles that have the ability to move and handle objects. It is the science and engineering of collating intelligent machines.

Modern technological advancements in the field of yield forecast may also aid farmers in cost prediction based upon the production. Mainly there are two varieties in yield prediction., classification and prediction phase. In this paper we are presenting an outline on classification and predictions techniques. Predictions can be acquired by analysing the large sets of pre-existent databases in order to generate a new understanding. It is also considered as one of the proven methods of settling the issues. Thus, resulting in automatic and rough predictions.

The different types of machine learning are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

SUPERVISED LEARNING

Supervised learning is the machine learning task of inferring a function from supervised training data. Training data includes a group of examples with paired input subjects and desired output.Needs the training algorithm to generalize from the training data to unseen situations during a reasonable way.

The system learns from the data provided which contains the features and the output as well. After the completion of learning, new data is provided without outputs, and therefore the system generates the output using the knowledge it gained from the info on which it trained. This is how supervised learning model works.

Supervised learning is oftendivided into classification and Regression problems.

- Classification

To predict the outcome of a given sample where the output variable is in the form of categories(output in the form of a or b)

- Regression

To predict the outcome of a given sample where the output variable is in the form of real values (predicted data is numerical in nature) Some of the supervised machine learning algorithms are as follows:

- Decision Trees

It is a kind of supervised learning algorithm that's mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets. This is done supported most vital attributes/ independent variables to form as distinct groups as possible

- Random Forest

Random Forest may be a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so referred to as "Forest"). To classify a replacement object supported on attributes, each tree gives a

classification and that we say the tree "votes" for that class. The forest chooses the classification having the foremost votes (over all the trees within the forest).

- Support Vector machine

In this work the supervised machine learning algorithm SVM is used for classification of microaneurysms. Support Vector Machine may be a supervised machine learning algorithm which may be used for both classification or regression challenges. However, it is mostly used in classification problems. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine may be a frontier which best segregates the two classes (hyper-plane/ line).

UNSUPERVISED LEARNING:

Unsupervised learning is where you simply have input file (X) and no corresponding output variables. The goal for unsupervised learning is to model the underlying structure or distribution within the data so as to find out more about the info. Unsupervised learning problems are often further grouped into clustering and association problems.

- Clustering

A clustering problem is where you would like to get the inherent groupings within the data, like grouping customers by purchasing behavior.

- Association

An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Some popular examples of unsupervised learning algorithms are:

- k-means for clustering problems

It is a kind of unsupervised algorithm which solves the clustering problem. Its procedure follows an easyand straight forward thanks to classify a given data set through a particular number of clusters (assume k clusters). Data points inside a cluster are homogeneous and heterogeneous to see groups.

- Apriori algorithm for association rule learning problems

The Apriori Algorithm is used in a transactional database to mine frequent item sets and then generate association rules. It is popularly utilized in market basket analysis, where one checks for combinations of products that regulalry co-occur within the database.

REINFORCEMENTLEARNNING:

Reinforcement Learning is a kind of Machine Learning, and thereby also a branch of Artificial Intelligence. It allows machines and software agents to automatically determine the perfect behavior within a selected context, so as to maximise its performance

II. RELATED WORK

a) Arun Kumar & et al [2]. "Efficient Crop Yield Prediction Using Machine Learning Algorithms". In this work, the crop yield classification has been performed to group on the basis of yield productivity using Artificial neural networks. And the range of productivity will be defined. Regression will be performed to get the actual crop yield and estimated cost.

b) SML Venkata Narasimhamurthy& et al [1]. "Rice Crop Yield Forecasting Using Random Forest Algorithm". In this paper, this algorithm is used to predict the crop yield (Rice), by considering the climatic conditions as the parameters. This also

aids farmers as well as the contributors in a better decision making. In research authors have taken the datasets of rainfall, temperature, rice production(tonnes), perception. Highest accuracy of 85.89% is obtained by using this method.

c) P. Priya*[1]& et al [4]. "Predicting the Crop Yield Using Machine Learning Algorithm". In this paper, Random forest algorithm has been used to check the yield of the crop as per the hectares, before farming on to the field. Data set such as rainfall, perception, temperature, production is taken into account. Decision tree is used for classification purposes. Here, accuracy can be predicted by relating the resultant class value with the test data set.

d) KawsarAkhand& et al [5]. "Yield Prediction in Bangladesh Using Satellite Remote Sensing Data and Artificial Neural Network". In this paper, feed-forward ANN with a back-propagation learning algorithm is used to train dataset and an AVHRR sensor-based vegetation health product (VHP) developed by NOAA was used to build a model for predicting wheat yield.

III. PROPOSED SYSTEM

In the proposed system, techniques such as pre-processing of data, classification and predictions have been carried out in aspect of obtaining finest results

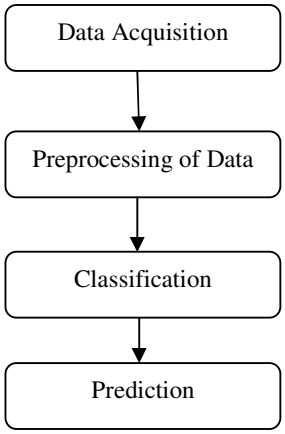


Fig-2:Flow Diagram of proposed system

IV. METHODOLOGY

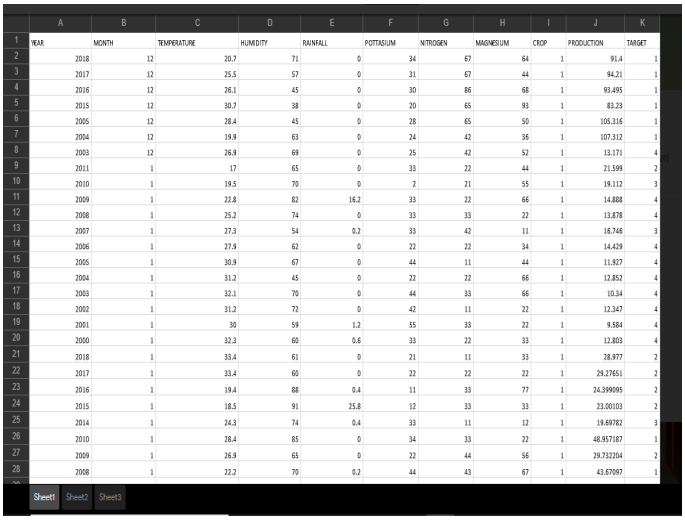
The proposed methodology contains of data acquisition, preprocessing of data, classification and prediction.

DATA ACQUISITION

It is a process of collecting data of yield production and giving it as input. It needs two things to work, data and models. The acquired data should have enough features (aspect of data that can aid for a prediction).

Datasets Used:

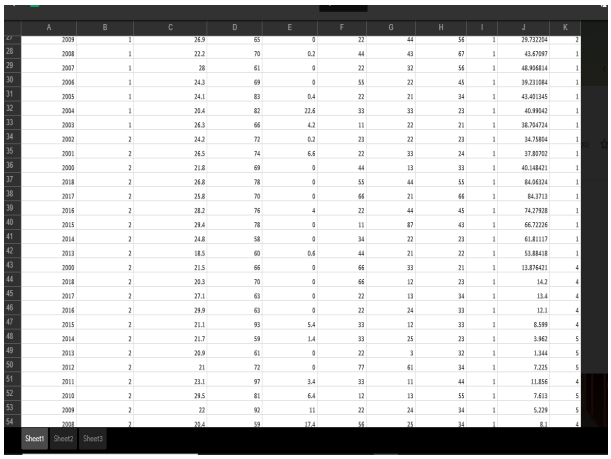
Only a limited number of important factors which have the highest impact on agricultural yield were selected here. The parameters selected are year, month, temperature, humidity, rainfall, production, target. The rainfall directly affects the crop production. Here is the total amount of precipitation of crops of each month of every year.Year Wise temperature details are taken into consideration as one of the inputs.



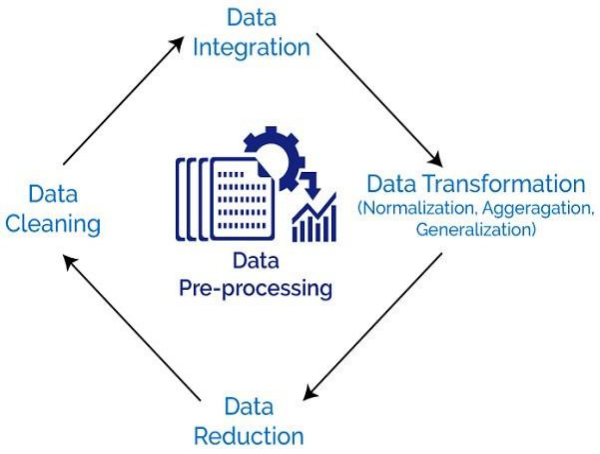
	A	B	C	D	E	F	G	H	I	J	K
1	YEAR	MONTH	TEMPERATURE	HUMIDITY	RAINFALL	POTTASIUM	NITROGEN	MAGNESIUM	CROP	PRODUCTION	TARGET
2	2018	12	20.7	71	0	34	67	64	1	91.4	1
3	2017	12	25.5	57	0	31	67	44	1	94.21	1
4	2016	12	26.1	45	0	30	86	68	1	93.495	1
5	2015	12	30.7	38	0	20	65	93	1	83.23	1
6	2005	12	28.4	45	0	28	65	50	1	105.316	1
7	2004	12	19.9	63	0	24	42	36	1	107.312	1
8	2003	12	26.9	69	0	25	42	52	1	13.171	4
9	2011	1	17	65	0	33	22	44	1	21.599	2
10	2010	1	19.5	70	0	2	21	55	1	19.112	3
11	2009	1	22.8	82	16.2	33	22	66	1	14.888	4
12	2008	1	25.2	74	0	33	33	22	1	13.878	4
13	2007	1	27.3	54	0.2	33	42	11	1	16.746	3
14	2006	1	27.9	62	0	22	22	34	1	14.429	4
15	2005	1	30.9	67	0	44	11	44	1	11.927	4
16	2004	1	31.2	45	0	22	22	66	1	12.852	4
17	2003	1	32.1	70	0	44	33	66	1	10.34	4
18	2002	1	31.2	72	0	42	11	22	1	12.347	4
19	2001	1	30	59	1.2	55	33	22	1	9.584	4
20	2000	1	32.3	60	0.6	33	22	33	1	12.803	4
21	2018	1	33.4	61	0	21	11	33	1	28.977	2
22	2017	1	33.4	60	0	22	22	22	1	29.27651	2
23	2016	1	19.4	88	0.4	11	33	77	1	24.399095	2
24	2015	1	18.5	91	25.8	12	33	33	1	23.00103	2
25	2014	1	24.3	74	0.4	33	11	12	1	19.69782	3
26	2010	1	28.4	85	0	34	33	22	1	48.957187	1
27	2009	1	26.9	65	0	22	44	56	1	29.732204	2
28	2008	1	22.2	70	0.2	44	43	67	1	43.67097	1

Sheet1 Sheet2 Sheet3

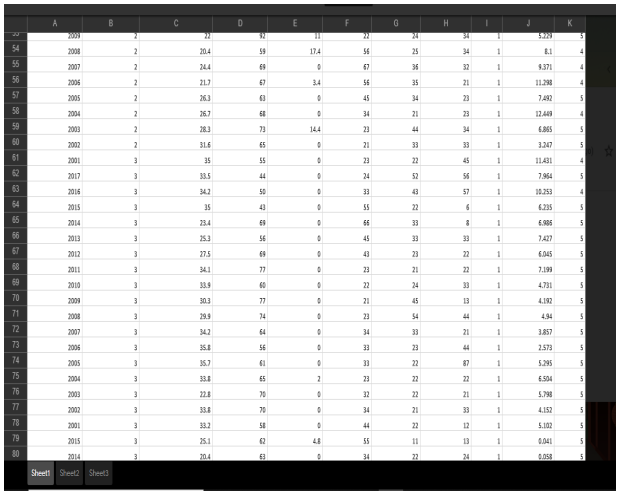
SHEET 1



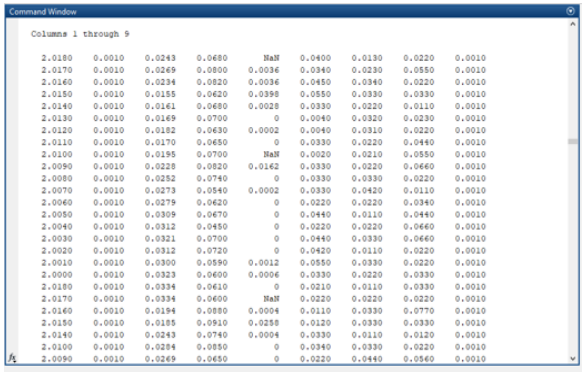
SHEET 2



DATA PRE-PROCESSING



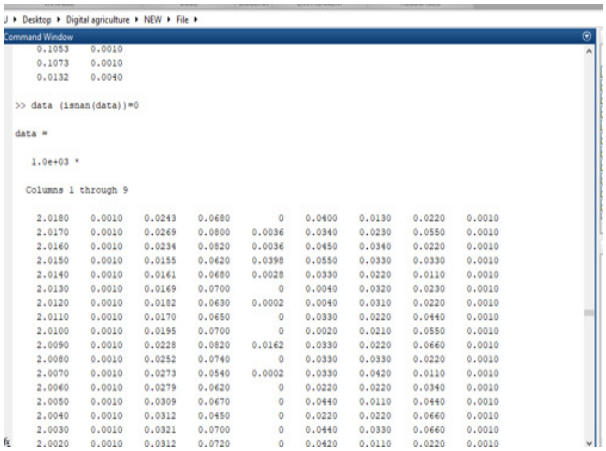
SHEET 3



BEFORE PRE-PROCESSING

PRE-PROCESSING OF DATA

Data pre-processing is a technique that is used to modify the raw data into a clean data set (i.e. reasonable format). In short, whenever the data is clustered from various sources, it is obtained in raw format which is not probable for the analysis. The aim of pre-processing is to improve the generalizability of the model. It handles many irrelevant and missing data. In this, pre-processing is carried out by the in-built functions in MAT LAB.



AFTER PRE-PROCESSING

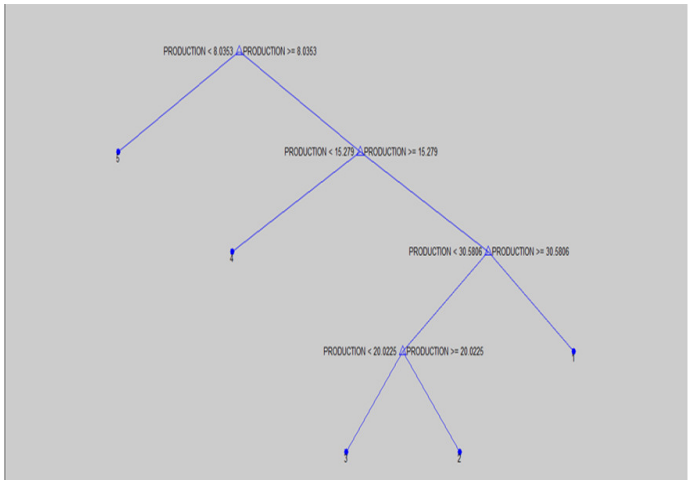
CLASSIFICATION

It is a process of determining, in which a set of categories (sub divisions) a new consideration belongs, on account of a training set of data containing observations whose class associates are established. A flowchart-like structure in which each internal node in represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). Classification rules are represented by the path from root to leaf. A decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated. Decision tree consists of three nodes,

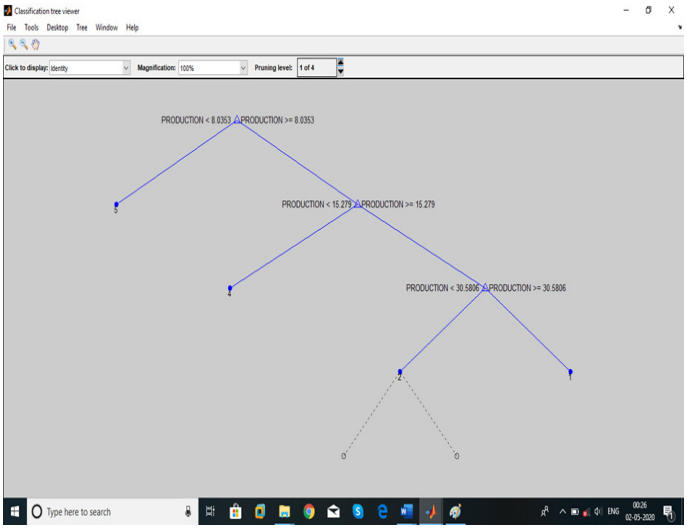
- Decision nodes – typically represented by squares
- Chance nodes – typically represented by circles
- End nodes – typically represented by triangles

Decision trees are also used in operations research and operations management. If, in practice, decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. It also used as a descriptive means for calculating conditional probabilities.

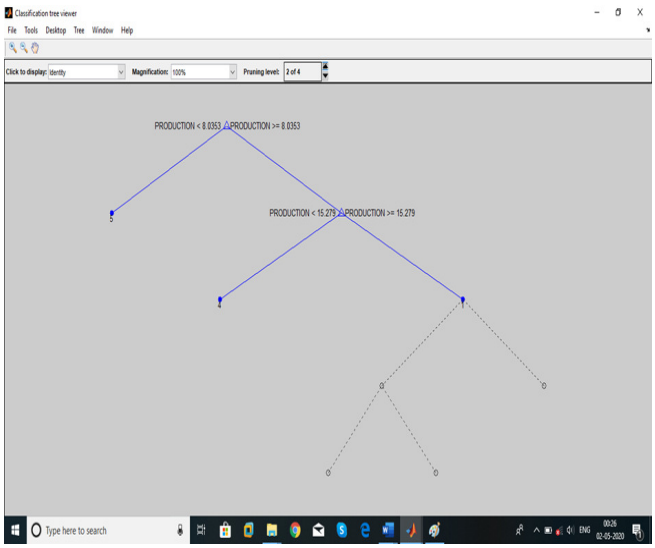
Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods. In this method, data from excel sheet (rows and columns) are given as input for the classification purpose, so that those inputs have been processed by MAT LAB and decision tree is being generated. Pruning is also carried out in order to reduce the complexity. As the level of pruning gets increased the complexity is reduced, which resultsincreasing the performance level.



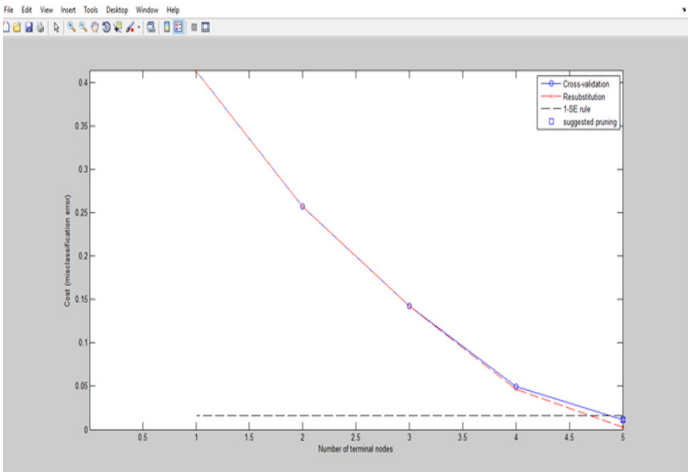
DECISION TREE



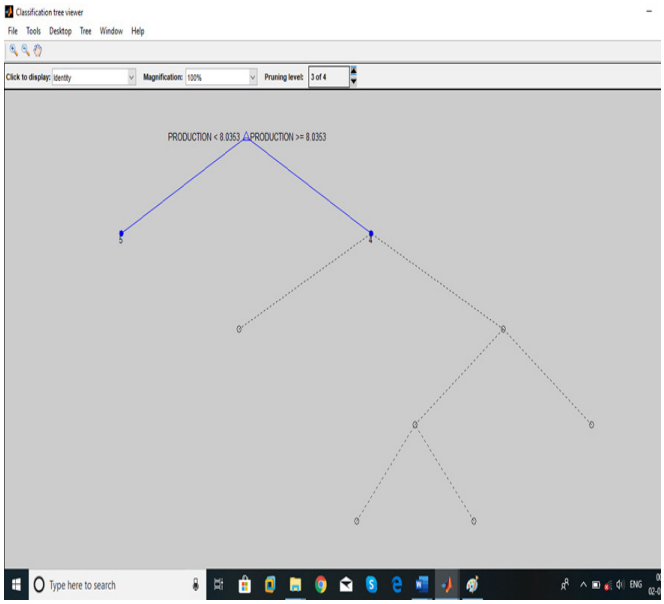
DECISION TREE – LEVEL 1 PRUNING



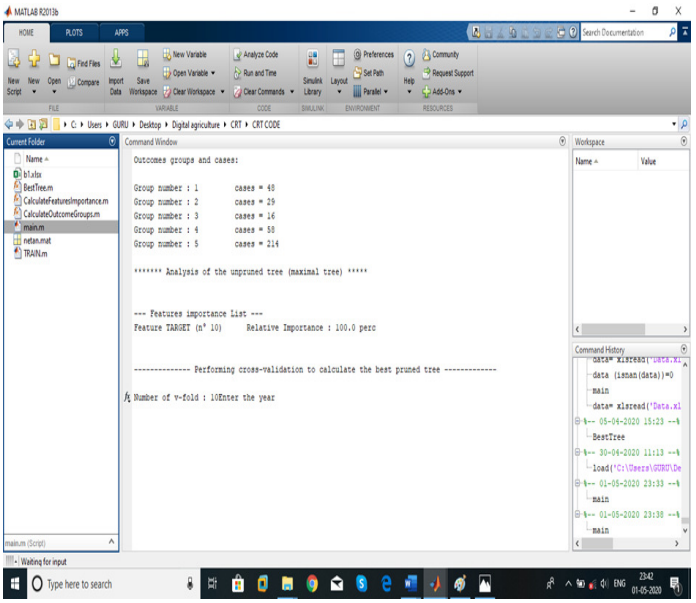
DECISION TREE – LEVEL 2 PRUNING



DECISION TREE GRAPH



DECISION TREE – LEVEL 3 PRUNING



OUTPUT OF CLASSIFICATION

PREDICTION

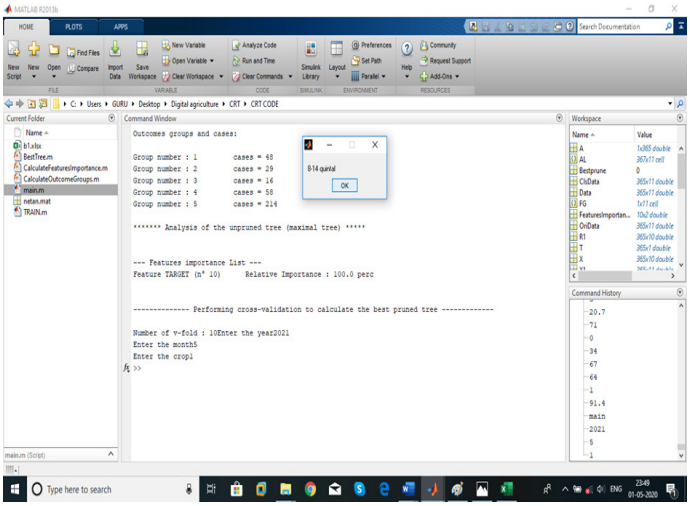
It is the process of referring to an algorithm's output, after it has been skilled on a factual dataset and applied to advanced data when forecasting the tendency of a particular outcome.

PREDICTION USING ARTIFICIAL NEURAL NETWORK- Machine learning method Artificial Neural Network is used for prediction of software defect. A neural network has several inputs, hidden, and output nodes. Each node applies a function some data (could be SoftMax, linear, logistic), and returns an output. Every node in the proceeding layer takes a weighted average of the outputs of the previous layer, until an output is reached. The reasoning is that multiple nodes can collectively gain insight about solving a problem (like classification) that an individual node cannot. The cost function differs for this type of model -- the weights between nodes adjust to minimize error.

Artificial neural networks are relatively crude electronic networks of neurons based on the neural structure of the brain. ANN process records one at a time, and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm for further iterations.

ANN is a complex adaptive system which can change its internal structure based on the information pass through it. It is achieved by adjusting the weight of connection. Each connection has a weight associated with it. A weight is a number that control the signal between two neurons. Weights are adjusted to improve the result. Popular methods of learning are given as:

SUPERVISED LEARNING: This strategy involves a trainer which is smarter than the network.



OUTPUT OF ANN

V. CONCLUSION

Agricultural examines would aid the agricultural bodies in order to uphold farmers for making meaningful and profitable decisions. In this paper we have presented observation on predicting the yield of the crop (classification and prediction) by using different machine learning algorithms. In addition, we have discussed juxtaposition of different algorithms used for crop yield prediction along with its benefits and drawbacks.

The results show that we can attain an accurate crop yield (wheat) prediction using the Artificial neural network algorithm. This makes the farmer to take the right decision for right crop such that the agricultural sector will be developed by agriculture.

There is scope for future development of this project. These project weather condition have a direct effect on crop yield. Artificial neural network has been demonstrated to be powerful tools for the prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameter of soil, PH, nitrogen, phosphate, potassium,

depth, temperature, rainfall, for that purpose we are used artificial neural network (ANN). The project aims of the prediction the yield of **wheat** crop by considering and analyzing survey of previous year of the crop. This project used for that increase in agriculture production by 70%.

VI.REFERENCES

[1] J. Ramirez-Villegas and A. Challinor, "Assessing relevant climate data for agricultural applications". Agricultural & Forest Meteorology, 2012, vol. 161(3), pp. 26–45.

[2] Mrs. K.R. Sri Preethaa, S.Nishanthini, D.Santhiya, K.Vani Shree ,"Crop Yield Prediction", International Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III, Issue III, March- 2016.

[3] Arun Kumar, Naveen Kumar, Vishal Vats**,** "Efficient Crop Yield Prediction Using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET)- e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 05 Issue: 06 | June-2018.

[4] Aditya Shastry, E. Bhanushree, H. A Sanjayand, "Regression Technique to Predict the Yield of The Crop". International Journal of computing12 (2):96-102 2017, ISSN:1816-9503

[5] S. Djodiltachoumy, E. Manjula, "Model for Predicting the Crop Yield", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

[6] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L., "Crop Yield Prediction using Machine Learning" (IRJET), Issue 2Feb-2018. Volume: 05

[7] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya NasrinRayta, AmitabhaChakrabarty, "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques", 978-1-5386-3831-6/17/$31.00©2017 IEEE.

[8] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Maximize Crop Yield Rate using Machine Learning Technique by Crop Selection Method", 2015 (ICSTM)

[9] Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta "Crop Prediction System using Machine Learning", Dec 2017 International Journal of Advance Engineering and Research Development

[10] Nain A. S., Dadhwal, V. K. and T. P. Singh, 2002 Crop Simulation Model with Minimal Data Set and Real Time Wheat Yield Assessment Using Technology Trend. Current Science. 82(10): 1255-1258.

[11] S. Bejo, S. Mustaffha and W. Ismail, "Predicting the crop yield by the Application of Artificial Neural Network", Journal of Food Science and Engineering, vol. 4, pp.1-9, 2014.

[12] W. Guo and H. Xue, "Crop Yield Forecasting Using Artificial Neural Networks", A comparison between spatial and temporal models, Mathematical Problems in Engineering, pp.1-7, 2014.

[13] S. Tripathi, V.V. Srinivas and R.S. Nanjundiah, "Downscaling of precipitation for climate change scenarios" A support vector machine approach, vol. 330, no. 3, pp.621-640, 2006. Journal of Hydrology.

[14] N.Gandhi, L.J. Armstrong and O. Petkar, "Rice Crop Yield Prediction in India using Artificial Neural Network", International Conference on 2016 IEEE TechnologicalInnovations in ICT for Agriculture and Rural Development (TIAR), Chennai, India scheduled on 15th and 16th July 2016.

[15] V. Anandhi and R.M. Chezian, "Support Vector Regression in Forecasting", International Journal of Advanced Research in Computer and Communication Engineering, October 2013. Vol. 2, no.10.

[16] K. Tanaka and T. Kiura, "Crop yield prediction systems for rainfed areas and mountainous areas in Thailand", Proceedings of the 9th Conference of the Asian Federation for Information Technology in Agriculture "ICT's for future Economic and Sustainable Agricultural Systems", 2014

[17] Nain A. S., Dadhwal,V. K. and T. P. Singh, 2002 Realtime wheat yield assessment using technology trend and crop simulation model with minimal data set.Current Science. 82(10): 1255-1258.

[18] Mandal, M.S.N., Pandit, D.B., Sufian, M.A., Mannan, M.A. and Hossain, A.K., "Problems and Prospects of wheat cultivation in Bangladesh", Int.J. BioRes., Vol.7(3), pp.45-52, 2009.

[19] Rahman, M.A., Chikushi, J., Yoshida, S. and Karim, A.J., "Growth and yield components of wheat genotypes exposed to high temperature stress under control environment", Bangladesh J. Agril. Res., Vol.34(3), pp.361-372, 2009.

[20] Curtis, B.C., Rajaram, S. and Macpherson, H.G., "Bread Wheat Improvement and Production", FAO Plant production and protection series, No.30, pp.12-15, 2002.

[21] Cakir, Y., Kirci, M. and Gunes, E. O., "Yield prediction of wheat in south-east region of Turkey by using artificial neural networks", IEEE Agrogeoinformatics, 3rd International Conference on 11-14 August, 2014, Beijing, China, ISBN: 978-1-4799-4157-5.