ISSN 2319 – 1953
International Journal of Scientific Research in Computer Science Applications and Management Studies

Crop Yield Prediction Using Machine Learning and Data Mining Algorithms for Modern Agriculture

[1]Mythra .N, [2]Dr.Velayudham.A, [3]Dr.Shamila E.S, [4]Pavithra.M
[1]PG Scholar, [2]Professor & Head, [3]Associate Professor,[4]Assistant professor Department of Computer Science and Engineering, Jansons Institute of Technology, Coimbatore, Tamil Nadu, India
mythra154@gmail.com, velayudham.jit@gmail.com, shamila.e.s@jit.ac.in,smilepavi17@gmail.com

Abstract: **Now –a-days the most important emerging field in the real world is agriculture Farming is the main occupation of India. About 70% of primary and secondary business is based on farming and it is also the backbone of our Indian economy. Agriculture plays an important role in our country. In farming, season cultivation of the crop is based on the soil. Another important work in agriculture is selecting soil is based on the weather condition and also when to harvest the crop for the best cultivation. The crop selection is an important task based on the weather conditions. Data mining techniques is very popular in the Area of agriculture. Data mining is the process of finding the new templates from the large data sets there are various data mining techniques are used and evaluated in agriculture for estimating the future year's crop production. Utilization of information and communications technology enables automation of extracting significant data in an effort to obtain knowledge and trends crop is an essential key factor of agriculture. So data mining techniques are useful for new framers. In our work we can select the crop based on following parameters with the help of artificial neural network. These are temperature, rainfall, Humidity, weather condition and Area. A prerequisite of intelligent system has brought artificial neural network (ANN) to become a new technology which provides assorted solution for the complex problems in agriculture researches.**

Keywords: **Data Mining, Agriculture, framer, Crop Selection, soil selection, methodology, high accuracy, prediction capabilities, artificial neural network (ANN)**

I. INTRODUCTION

India is one among the oldest countries which is still practicing agriculture. But in recent times the trends in agriculture has drastically evolved due to globalization. Various factors have affected the health of agriculture in India. Many new technologies have been evolved to regain the health. One such technique is precision agriculture. Precision agriculture is budding in India .Precision agriculture is the technology of "site-specific" farming. It has provided us with the advantage of efficient input, output and better decisions regarding farming. Although precision agriculture has delivered better improvements it is still facing certain issues.

DATA MINING:

Data mining is the term used to describe the process of extracting value from a database. A data-warehouse is a location where information is stored. The type of data stored depends largely on the type of industry and the company. Many companies store every piece of data they have collected, while others are more ruthless in what they deem to be "important". Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. A number of advances in technology and business processes have contributed to a growing interest in data mining in both the public and private sectors. Some of these changes include the growth of computer networks, which can be used to connect databases; the development of enhanced search-related techniques such as neural networks and advanced algorithms; the spread of the client/server computing model, allowing users to access centralized data resources from the desktop; and an increased ability to combine data from disparate sources into a single search source. There exist many systems which propose the inputs for a particular farming land. Systems propose crops, fertilizers and even farming techniques. Recommendation of crops is one major domain in precision agriculture. Recommendation of crops is dependent on various parameters. Precision agriculture aims in identifying these parameters in a site-specific manner in order to resolve issues regarding crop selection. The "site-specific" technique has improved the results yet there is a need to supervise the results of such systems. Not all precision agriculture systems provide accurate results. But in agriculture it is important that the recommendations made are accurate and precise because in case of errors it may lead to heavy material and capital loss. Many research works is being carried out, in order to attain an accurate and efficient model for crop prediction. Ensemble is one such technique that is included in such research works. Among these various machine learning techniques that are being used in this field; this paper proposes a system that uses the voting method to build an efficient and accurate model. The crop is varied depends upon the following environmental parameters such as rainfall, temperature, Humidity.

II.RELATED WORK

USE OF DATA MINING IN CROP YIELD PREDICTION

Agriculture is the most important sector that influences the economy of India. It contributes to 18% of India's Gross Domestic Product (GDP) and gives employment to 50% of the population of India. People of

India are practicing Agriculture for years but the results are never satisfying due to various factors that affect the crop yield. To fulfil the needs of around 1.2 billion people, it is very important to have a good yield of crops. Due to factors like soil type, precipitation, seed quality, lack of technical facilities etc. the crop yield is directly influenced. Hence, new technologies are necessary for satisfying the growing need and farmers must work smartly by opting new technologies rather than going for trivial methods. This paper focuses on implementing crop yield prediction system by using Data Mining techniques by doing analysis on agriculture dataset. Different classifiers are used namely J48, LWL, LAD Tree and IBK for prediction and then the performance of each is compared using WEKA tool. For evaluating performance Accuracy is used as one of the factors. The classifiers are further compared with the values of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Relative Absolute Error (RAE). Lesser the value of error, more accurate the algorithm will work. The result is based on comparison among the classifiers.

A STUDY ON VARIOUS DATA MINING TECHNIQUES FOR CROP YIELD PREDICTION

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality ,pH ,EC,N,P,K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining. Through data mining we extract the knowledge from the huge size of data. This paper presents the study about the various data mining techniques used for predicting the crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. This paper summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation.

ANALYSIS OF SOIL BEHAVIOUR AND PREDICTION OF CROP YIELD USING DATA MINING APPROACH

Yield prediction is very popular among farmers these days, which particularly contributes to the proper selection of crops for sowing. This makes the problem of predicting the yielding of crops an interesting challenge. Earlier yield prediction was performed by considering the farmer's experience on a particular field and crop. This work presents a system, which uses data mining techniques in order to predict the category of the analyzed soil datasets. The category, thus predicted will indicate the yielding of crops. The problem of predicting the crop yield is formalized as a classification rule, where Naive Bayes and K-Nearest Neighbor methods are used.

PREDICTING CROP DISEASES USING DATA MINING APPROACHES: CLASSIFICATION

Agriculture research is rapidly growing, due to advancement of technologies and upcoming challenges. It has been proven to be leading role in improving the overall growth rate of any country. Especially in Pakistan, there is a dire need to do extensive research for better productivity in agriculture. To improve the growth rate of agriculture, researchers of this domain used different data mining techniques to solve agriculture related problems. Data mining approaches such as classification helps to predict the crops diseases, production and loss. It supports farmer while taking right decisions. This paper focuses on prediction of loss due to grass grub insect. We analyze the damages by using well-known classifiers such as Decision Tree, Random Forest, Neural Networks, Naïve Bayes, Support Vector Machines and K-Nearest Neighbor and design Ensemble Models of above mentioned classifiers which gave better results as compared to classifiers. Neural Networks and Random Forest produced slightly better results than other classifiers. Ensemble model improve the results of weak classifiers and proven as fruitful technique for our agriculture related problem. To improve the results further, hybrid of evolutionary algorithms and data mining techniques will be used which is our future research direction.

CROP RECOMMENDATION SYSTEM FOR PRECISION AGRICULTURE

Data mining is the practice of examining and deriving purposeful information from the data. Data mining finds its application in various fields like finance, retail, medicine, agriculture etc. Data mining in agriculture is used for analyzing the various biotic and abiotic factors. Agriculture in India plays a predominant role in economy and employment. The common problem existing among the Indian farmers are they don't choose the right crop based on their soil requirements. Due to this they face a serious setback in productivity. This problem of the farmers has been addressed through precision agriculture. Precision agriculture 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 based on their site-specific parameters. This reduces the wrong choice on a crop and increase in productivity. In this paper, this problem is solved by proposing a recommendation system through an ensemble model with majority voting technique using Random tree, CHAID, K-Nearest Neighbor and Naive Bayes as learners to recommend a crop for the site specific parameters with high accuracy and efficiency.

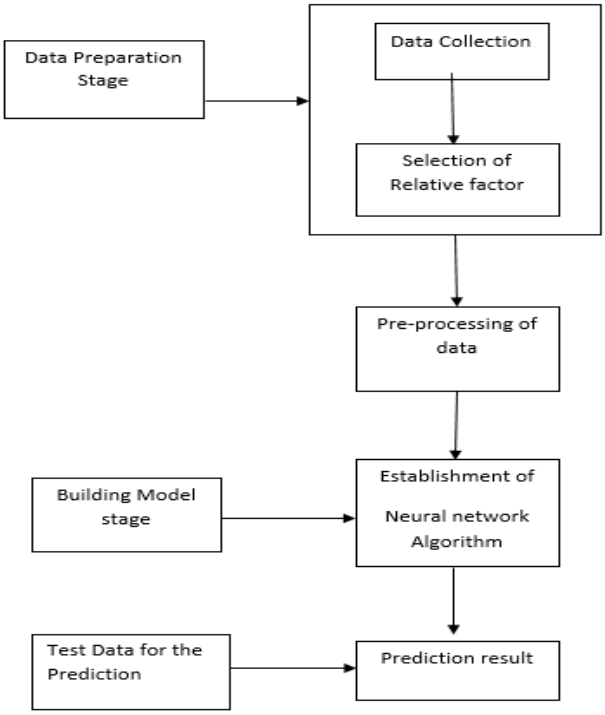
USE OF DATA MINING TECHNIQUE FOR PREDICTION OF TEA YIELD IN THE FACE OF CLIMATE CHANGE OF ASSAM, INDIA

Data mining is an emerging field of research in Information Technology as well as in agriculture. The present study focus on the applications of data mining

techniques in tea plantations in the face of climatic change to help the farmer in taking decision for farming and achieving the expected economic return. This paper presents an analysis using datA mining techniques for estimating the future yield prediction in tea cultivation with climatic change trends observed in last 30 years (1977-2006). The patterns of crop production in response to the climatic (rainfall, temperature, relative humidity, evaporation and sunshine) effect across the four tea growing regions (South Bank, North Bank, Upper Assam and Cachar) of Assam were developed using Multiple Linear Regression (MLR) technique. The tea production estimation equations developed for the regions were validated for the future yield prediction (2007, 2009 and 2010) and were found to be significant. Thus it is suggested that the planters/farmers could use the technique to predict the future crop productivity and consequently adopt alternative adaptive measures to maximize yield if the predictions fall below expectations and commercial viability.

METHODOLOGY

In our work, we predict the better crop yield for the respected season and soil based on the temperature, rainfall, humidity of the various Areas. We Use Neural Network Classifier for Prediction of Crop yield.



ARCHITECTURE DIAGRAM

Data collection:

Collect the data on various environmental conditions and different zones. The type of crop is varied corresponding to the temperature, rainfall, and Humidity and Soil type on the period of time. We collect the data and create a table.

Pre-processing:

In the table may be occur invalid data. In the pre-processing stage remove the invalid numbers. And also convert the string values into numerical values for the further implementation. Pre-processing is a process of convert the dataset into reasonable format for the prediction of Crop yield.

Neural Network Classifier:

Classifier: In order to classify a set of data into different classes or categories, the relationship between the data and the classes into which they are classified must be well understood to achieve this by computer, the computer must be trained

- Training is key to the success of classification
- Classification techniques were originally developed
- Out of research in Pattern Recognition field Computer classification of remotely sensed images involves the process of the computer program learning the relationship between the data and the information classes

Artificial Neural Network:

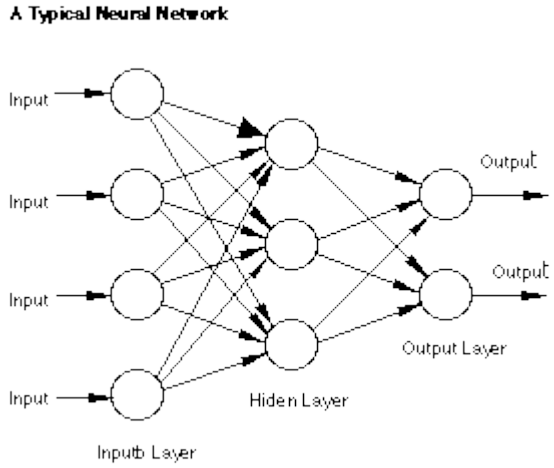
Advantages of using ANN

1. ANNs have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are non-linear as well as complex.
2. ANNs can generalize—after learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well, thus making the model generalize and predict on unseen data.
3. Unlike many other prediction techniques, ANN does not impose any restrictions on the input variables (like how they should be distributed). Additionally, many studies have shown that ANNs can better model heteroskedasticity i.e. data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data. This is something very useful in financial time series forecasting (e.g. stock prices) where data volatility is very high

We use the Machine learning method Artificial Neural Network for classifying the normal crop and the weed parts in the farms using extracted features. 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. They 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.



A neuron in an artificial neural network is

1. A set of input values (xi) and associated weights (wi).
2. A function (g) that sums the weights and maps the results to an output (y).

Neurons are organized into layers: input, hidden and output. The input layer is composed not of full neurons, but rather consists simply of the record's values that are inputs to the next layer of neurons. The next layer is the hidden layer. Several hidden layers can exist in one neural network. The final layer is the output layer, where there is one node for each class. A single sweep forward through the network results in the assignment of a value to each output node, and the record is assigned to the class node with the highest value.

Training an Artificial Neural Network

In the training phase, the correct class for each record is known (termed supervised training), and the output nodes can be assigned correct values -- 1 for the node corresponding to the correct class, and 0 for the others. (In practice, better results have been found using values of 0.9 and 0.1, respectively.) It is thus possible to compare the network's calculated values for the output nodes to these correct values, and calculate an error term for each node (the Delta rule). These error terms are then used to adjust the weights in the hidden layers so that, hopefully, during the next iteration the output values will be closer to the correct values.

The Iterative Learning Process

Once a network has been structured for a particular application, that network is ready to be trained. To start this process, the initial weights (described in the next section) are chosen randomly. Then the training (learning) begins.

The network processes the records in the Training Set one at a time, using the weights and functions in the hidden layers, then compares the resulting outputs against the desired outputs. Errors are then propagated back through the system, causing the system to adjust the weights for application to the next record. This process occurs repeatedly as the weights are tweaked. During the training of a network, the same set of data is processed many times as the connection weights are continually refined.

Cross validation method:

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

Train & Test data using ANN

The word network in the term 'artificial neural network' refers to the inter–connections between the neurons in the different layers of each system. An example system has three layers. The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations.

III. RESULT AND DISCUSSION

The dataset is opened in R tool and the result is obtained by performing various experiments for crop yield prediction using data mining techniques

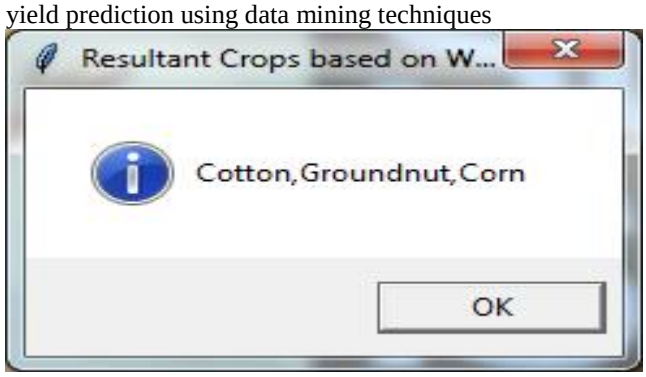


Figure 1: crop based on weather condition

Figure 2: rainfall rate based on the humidity

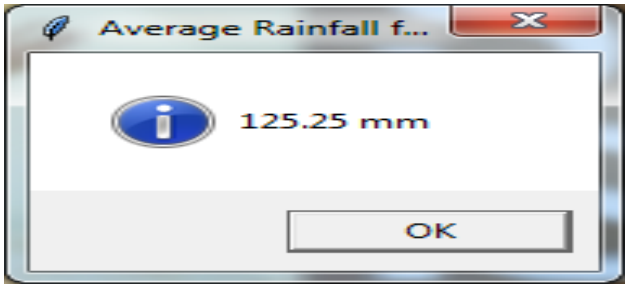


Figure 3: average rainfall rate

Figure 4: Final crop obtained for area

IV. CONCLUSION

Agriculture plays the major role in developing countries like India. Using some new technology in framing can change the scenario of decision making and framers to yield crop production in better way. Data mining in the agriculture discipline can be perceived to be novel undertaking a practice that is targeted at improving the welfare of the people .every framer is highly interested with the level of yields to expected in given month or season earlier on prediction of crop yield was performed through analysis of the past experience by the framer .data mining procedures can be employed to perform prediction on possible future yields Data mining, through better management and data analysis in data mining with amount of available information it is easier to make the decision .analysis of our parameters give the impact to the agriculture , bye the use of data mining techniques acquired knowledge can be used to make successful decision which will advance the success of the agriculture organization on the market .data mining technology provides user oriented access to new and hidden patterns in data from which knowledge is generated which can help with decision making in agriculture organization use of data mining techniques in agriculture field creates condition for mankind adequate decision and with that achieving competing advantage .increased drought and floods are likely to increased production variability Data mining plays important role in decision making We proposed a Data mining algorithm to predict the crop yield based on different attributes like rainfall and temperature, weather condition, Areas and Soil type. We use Regression tree and Artificial Neural Network for Classification and Prediction of Crop yield. And also we concentrate on Prediction of crop yield based on weather condition and temperature, rainfall respective to particular areas.

REFERENCE

[1]. Ramesh A, Vijay. S. Rajpurohit (2014), 'A survey on data mining techniques for crop yield prediction', International Journal of Advance Research in Volume 2, No 9, page 25.
[2]. Zekarais Dirida, Berhanu Borena (2015), 'Application of data mining techniques for crop productivity predication', International Journal of Computer Science and Technology in Volume 1, No 5, page 15.
[3]. Dakshayini Patil, Dr.M.S.Shirdhonkar (2017), 'Rice crop yield prediction using data mining techniques', International Journal of Advanced Research in Computer Science and Software Engineering in Volume 7, No 5, page 6'.
[4]. DR.A.Senthil Kumar,P.Arun (2017), 'A survey on agriculture for crop yield prediction using data mining techniques', IOSR Journal of Computer Engineering in Volume 6,No 3,page 11.
[5]. Yogesh Gandge,sandhya (2013), 'A study on various data mining techniques for crop yield prediction', International Conference on Computer and Optimization Techniques in Volume 7,No 3,page 15.
[6]. Shute Mishra, priyanka paygude, Snehal Chaudhary, Sonali idate (2018), 'Use of Data Mining in Crop Yield Prediction', Proceedings of the Second International Conference on Inventive Systems and Control in Volume 4, No 4, page 14.
[7]. Monali Paul, Santosh K. Vishwakarma, Ashok Verma (2015), 'Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach', International Conference on Computational Intelligence and Communication Networks in Volume 5,No 4,page10
[8]. R. Sujatha, Dr. P.Isakki (2016), ' A Study on Crop Yield Forecasting Using Classification Techniques', International Journal of Research in Engineering and Technology in Volume 3 ,No 2,page 14 .
[9]. Rajeswari and k.arunesh (2016), 'Analysing Soil Data using Data Mining classification techniques', Indian Journal of science and Technology, Volume 9, No 2, page 10.
[10]. K. L. Ponce-Guevara1, J. A. Palacios-Echeverr´ıa1, E. Maya-Olalla1, H. M. Dom´ınguez-Limaico1,L. E. Su´arez-Zambrano1, P. D. Rosero-Montalvo1,4, D. H. Peluffo-Ord´o˜nez1,2, and J. C. Alvarado-P´erez2 (2017), "Green Farm-DM: A tool for Analysing Vegetable Crops Data from a Greenhouse using Data Mining Techniques(First trial)"in Volume 4,No 4,page 7.
[11]. D Ramesh, B Vishnu Vardhan (2015), 'Analysis of Crop Yield Prediction using Data Mining Techniques', International Journal of Research in Engineering and Technology, Volume 4, No 1, page 14.
[12]. Aakunuri Manjula and Dr.G.Narsimha (2016), 'Crop Yield Prediction with Aid of Optimal Neural Network in Spatial Data Mining: New Approaches', International Journal of Information & Computation Technology, Volume 6, No 1, page 25-33.
[13]. A. K. Tripathy, J. Adinarayana, D. Sudharsan, K. Vijayalakshmi , S. N. Merchant, U. B.Desai (2013), 'Data Mining and Wireless Sensor Network for Groundnut Pest/Disease Interaction and predictions –A Preliminary Study', International Journal of Computer Information Systems and Industrial Management Applications, Volume 5,No 4,page 427-436.
[14]. S. Hari Ganesh, Mrs. Jayasudha (2015), 'Data Mining Technique to Predict the Accuracy of the soil Fertility', International Journal Of Computer Science and Mobile Computing, Volume 4, No 7, page 330-333.
[15]. B. Milovi and V. Radojevi (2015), 'Application of Data Mining in Agriculture', Bulgarian Journal of Agricultural Science, Volume 21, No 1, page 26-34, May 2015.