IMPLEMENT A SYSTEM FOR CROP SELECTION AND YIELD PREDICTION USING RANDOM FOREST ALGORITHM

M. Devaki[1], **K. B. Jayanthi**[2], **M. Priya**[3]

Abstract

Agriculture is the bone that plays important part in the frugality of India. India is an agrarian country and its frugality largely grounded upon crop product. It'll allow policy makers and growers to take effective marketing and storehouse way to prognosticate crop yields before in their crop. This design will allow growers to capture the yield of their crops before civilization in the field of husbandry and therefore help them make the necessary opinions. Perpetration of such a system with a use and the machine learning algorithm can also be distributed. The results attained are granted access to the planter. And yet there are colorful styles or protocols for similar veritably data analytics in crop yield vaticination, and we're suitable to prognosticate agrarian productivity with guidance of all those algorithms. It utilizes a Random Forest Algorithm. By probing similar problems and issues similar as rainfall, temperature, moisture, downfall, moisture, its overcome using Random Forest technique, acceptable results and inventions to resolve the situation. In countries like India, indeed in the agrarian sector, as there are numerous types of adding profitable growth. In addition, the processing is useful for vaticinating the product of crop yields.

KeyWords: Climate, crop yield, Indian Agriculture, Machine Learning Techniques, Random Forest Algorithm.

DOI Number: 10.14704/nq.2022.20.8.NQ44702 **NeuroQuantology 2022; 20(8): 6781-6787**

[1] Assistant Professor, K.S.Rangasamy college of technology,Tiruchengode, Tamil nadu, India
[2] Dean/SES, K.S.Rangasamy college of technology,Tiruchengode, Tamil nadu, India
[3] Post Graduate Student, K.S.Rangasamy college of technology,Tiruchengode, Tamil nadu, India

Introduction

For numerous pattern bracket problems, a advanced number of features used don't inescapably restate into a advanced bracket delicacy. In some cases the performance of algorithms devoted to speed and prophetic delicacy of the data characterization can indeed drop. Thus, point selection can serve as apre-processing tool of enormous significance previous to working the bracket problems. The purpose of the point selection is to reduce the maximum number of inapplicable features while maintaining an respectable bracket delicacy. A good point selection system can reduce the cost of point dimension, and increase classifier effectiveness and bracket delicacy. Point selection is of considerable significance in pattern bracket, data analysis, multimedia information reclamation, medical data processing, machine literacy, and data mining operations. PSO is used to apply a point selection, and SVMs with the one-versus- rest system were used as observer for the PSO fitness function for five multiclass problems taken from the literature. The results reveal that our system illustrated a better delicacy than the bracket styles they were compared to.

DATA MINING

Data mining(the analysis step of the" Knowledge Discovery in Databases" process, or KDD), a field at the crossroad of computer wisdom and statistics, is the process that attempts to discover patterns in large data sets. It utilizes styles at the crossroad of artificial intelligence, machine literacy, statistics, and database systems The overall thing of the data mining process is to prize information from a data set and transfigure it into an accessible structure for farther use Away from the raw analysis step, it involves database and data operation aspects, data preprocessing, model and conclusion considerations, interestingness criteria, complexity considerations,post-processing of discovered structures, visualization, and online updating. Generally, data mining(occasionally called data or knowledge discovery) is the process of assaying data from different perspectives and recapitulating it into useful information- information that can be used to increase profit, cuts costs, or both. Data mining software is one of a number of logical tools for assaying data. It allows druggies to dissect data from numerous different confines or angles, classify it, and epitomize the connections linked. Technically, data mining is the process of chancing correlations or patterns among dozens of field in large relational databases.

Data Mining Techniques

There are several major data mining ways have been developed and used in data mining systems lately including association, bracket, clustering, vaticination and successional patterns.

Association

Association is one of the best known data mining fashion. In association, a pattern is discovered grounded on a relationship of a particular item on other particulars in the same sale. For illustration, the association fashion is used in request handbasket analysis to identify what products that guests constantly buy together. Grounded on this data businesses can have corresponding marketing crusade to vend further products to make further profit.

Classification

Classification is a classic data mining fashion grounded on machine literacy. Principally bracket is used to classify each item in a set of data into one of predefined set of classes or groups. Classification system makes use of fine ways similar as decision trees, direct programming, neural network and statistics. In bracket, make the software that can learn how to classify the data particulars into groups. For illustration, can apply Classification in operation that " given all once records of workers who left the company, prognosticate which current workers are presumably to leave in the future. " In this case, divide the hand's record into two groups that are " leave " and " stay ".

Clustering

Clustering is a data mining fashion that makes meaningful or useful cluster of objects that have analogous characteristic using automatic fashion. Different from bracket, clustering fashion also defines the classes and put objects in them, while in bracket objects are assigned into predefined classes. To make the conception clearer, can take library as an illustration. In a library, books have

a wide range of motifs available. The challenge is how to keep those books in a way that compendiums can take several books in a specific content without hassle.

LITERATURE REVIEW

[1] Gabriel M. Alves, Paulo E. Cruvinel, In perfection husbandry an increase of data and information has been observed and new approaches to ameliorate knowledge are now needed. Thus, studies on Big Data are being conducted to find innovative results as a means to dissect large data sets. In this work, we present a Big Data terrain for agrarian soil analysis from reckoned tomography(CT) images. Our structure is planned in three layers source; Bigdata terrain, and operations. We use Hadoop frame in the alternate subcaste to reuse CT images and bandy how the 3D reconstruction is performed. Another operation in the structure is the statistical analysis of soil samples. The Big Data terrain is developed as a soil analysis system to gain an understand of the problems related to agrarian lands.

[2] Tyrone T. Lin, Chung-Shiao. Hsieh This paper substantially explores when the agrarian assiduity faces grain crop price oscillations and natural climate changes, it'll take which position of price of grain crops and what probability of climate changes for developing a dynamic grain crop gyration model. In the former paper, the authors introduce the mixed strategy of game proposition to construct a 2- player game. In consideration of the pursuit of the maximization of their own interests, the decision- timber of dynamic grain crop gyration is the main focus of the former paper, and it'll be extended to a multiple stable dynamic grain crop gyration strategy cycle. And now the authors develop a stationary Markov process as the base for a final decision. Markov chain is a system constantly used in decision- timber and is a model simple to be bandied.

[3] Niketa Gandhi, Leisa J. Armstrong, OwaizPetkar, Food product in India is largely dependent on cereal crops including rice, wheat and colorful beats. The sustainability and productivity of rice growing areas is dependent on suitable climatic condition. Variability in seasonal climate conditions can have mischievous effect, with incidents of failure reducing product. Developing better ways to prognosticate crop productivity in different climatic conditions can help planter and other stakeholders in better decision making in terms of agronomy and crop choice. Machine literacy ways can be used to ameliorate vaticination of crop yield under different climatic scripts. This paper presents the review on use of similar machine learning fashion for Indian rice cropping areas. This paper discusses the experimental results attained by applying SMO classifier using the WEKA tool on the dataset of 27 sections of Maharashtra state, India. The dataset considered for the rice crop yield vaticination was sourced from intimately available Indian Government records. The parameters considered for the study were rush, minimal temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, product and yield for the Kharif season(June to November) for the times 1998 to 2002. For the present study the mean absolute error(MAE), root mean squared error(RMSE), relative absolute error(RAE) and root relative squared error(RRSE) were calculate. The experimental results showed that the performance of other ways on the same dataset was much better compared to SMO.

[4] Vidya V. PolProf. S.M.PatilThe term Big Data, refers to vastly substantial data whose volume, variability, and haste make it veritably laborious to manage, process or anatomized. To dissect this vastly substantial kind of data Hadoop will be employed. Still, Processing is veritably time-consuming. To resolve this dilemma & to diminishment replication time one result is to executing the job incompletely, where an approximate, early result becomes available to the use, afore completion of job. Proposed system gives a more nascent Chart Reduce armature that warrants data to be divided for easier & early processing. This isn't time consuming and amends system application for batch jobs as well. Proposed system presents a more nascent interpretation of the Hadoop Map Reduce frame that fortifies on- Process aggregation, which warrants & avails druggies to get early results of

a job as it's calculating. It'll estimate this fashion exercising authentic world datasets and operations and endeavor to amend the systems performance in terms of perfection and time. Also the combiner introduced in this system is original reducer. Combiner will get execute later chart function & before reducer. Rather of processing complete train on- process aggregation divides the train into number of blocks which helps to gives the result in places. Dividing the train into number of data sets helps to give result as early as possible by giving intermediate result to the stoner. The ideal of the proposed fashion is to amend the performance of Hadoop Map Reduce for effective & easy Immensely Big Data Processing time.

EXISTING SYSTEM

A good point selection system can reduce the cost of point dimension, and increase classifier effectiveness and bracket delicacy. Point selection is of considerable significance in pattern bracket, data analysis, multimedia information reclamation, medical data processing, machine literacy, and data mining operations. PSO is used to apply a point selection, and KNN with the one-versus- rest system were used as observers for the PSO fitness function for five multiclass problems taken from the literature. The results reveal that our system illustrated a better delicacy than the bracket styles they were compared to.

Disadvantages

- It doesn't classify the unlabeled data.
- It take further time for training.

PROPOSED SYSTEM

Random timber is a principally supervised literacy algorithm that's used for both groups as well as retrogression. Random timber algorithm creates decision trees on different data samples and also prognosticate the data from each subset and also by advancing gives better the result for the system. Random Forest used the bagging system to trained the data. Principally, the bagging system is a admixture of studying different models and increase the final result of the system. For getting high delicacy we used the Random Forest algorithm which gives delicacy which predicate by model and factual outgrowth of predication in the dataset. In the arbitrary timber which beaters the decision tree from a sample of data and trees gives the vaticination from each family and selects the stylish result by voting which gives better delicacy for the model. It gives optimum results for the system.

Random Forest works in two- phase first is to produce the arbitrary timber by combining N decision tree, and second is to make prognostications for each tree created in the first phase.

The Working process can be explained in the below way and illustration

Step- 1 Select arbitrary K data point from the training set.

Step- 2 Figure the decision trees associated with the named data points(Subsets).

Step- 3 Choose the number N for decision trees that you want to make.

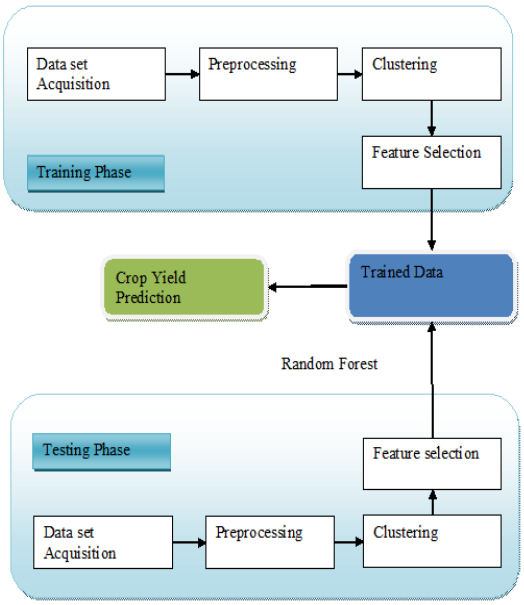
Step- 4 Reprise Step 1 & 2

Advantages

- It takes lower training time as compared to other algorithm.
- It predicts affair with high delicacy, indeed for the large dataset it runs efficiently.
- It can also maintain delicacy when a large proportion of data is missing.

DESIGN AND DEVELOPMENT PHASE

The proposed System is intended to help the farmers and researchers to understand the crop field and to choose a suitable crop for it. Various parameters are considered from soil to atmosphere for predicting the suitable crop. Soil parameters such as type, ph level, iron, copper, manganese, sulphur, organic carbon, potassium, phosphate, nitrogen.



MODULE DETAILS

The Product will perform the following functions

- Dataset Acquisition
- Preprocessing
- Clustering
- Feature Selection
- Classification

6.1 Dataset Acquisition

- In this module is used to upload the rainfall details.
- It contains the' Time',' Downfall',' Area of Sowing',' Yield',' Diseases'(Nitrogen, Phosphorous and Potassium) and' Product'.

6.2 Pre-processing

pre-processing is an important step in the data mining process. The expression" scrap in, scrap out" is particularly applicable to data mining and machine systems. Data- gathering styles are frequently approximately controlled, performing in out- of- range values, insolvable data combinations, missing values, etc. Assaying data that has not been precisely screened for similar problems can produce deceiving results.

6.3 Clustering

Clustering is a fashion in data mining to find intriguing patterns in a given dataset. The k-means algorithm is an evolutionary algorithm that gains its name from its system of operation. The algorithm clusters information into k groups, where k is considered as an input parameter. It also assigns each information's to clusters grounded upon the observation's propinquity to the mean of the cluster. The cluster's mean is also more reckoned and the process begins again. The k- means algorithm is one of the simplest clustering ways and it's generally used in medical data and related fields. K- Means algorithm is a divisive, unordered system of defining clusters.

6.4 Feature Selection

feature selection is the process of opting a subset of applicable, useful features for use in erecting an logical model. Point selection helps constrict the field of data to just the most precious inputs, reducing noise and perfecting training performance

6.5 Classification

In this module, apply Classification algorithm to classify the data, eventually prognosticate the yield product using Random Forest Classification. Random timbers are the aggregation of tree predictors in such a way that each tree depends on the values of a arbitrary subset tried singly and with the same distribution for all trees in the timber. Random Forest used the bagging system to trained the data which increases the delicacy of the result. For getting high delicacy we used the Random Forest algorithm which gives delicacy which predicate by model and factual outgrowth of predication in the dataset.

Pseudocode of the Proposed System

1. In this principally we first aimlessly named the k to point out of the total m point in the model.

2. Using the stylish split point choose the k point and calculate the knotd.

3. So we used the split system, resolve the bumps into the son knot.

4. Reprise 1 to three way until l number of bumps has been reached

5. Figure timber by repeating way 1 to 4 for n number times to make n number of trees.

To perform vaticination using the trained arbitrary timber algorithm uses the below

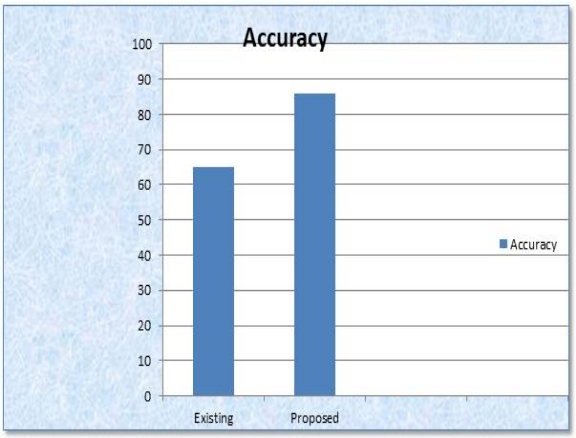
pseudocode

1. In this, we took the test features and every arbitrary decision tree to rest the affair and anticipated the outgrowth which is stored

. And also we calculated the vote which is given by every decision tree for each rested outgrowth.

3. Eventually, we considered high suggested rested outgrowth which gives the final vaticination from the arbitrary timber algorithm.

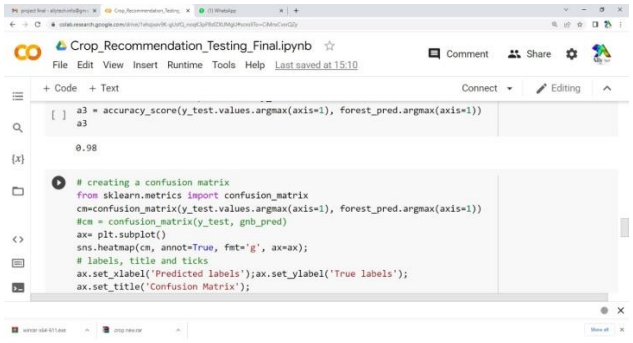
6.6Result and Discussion

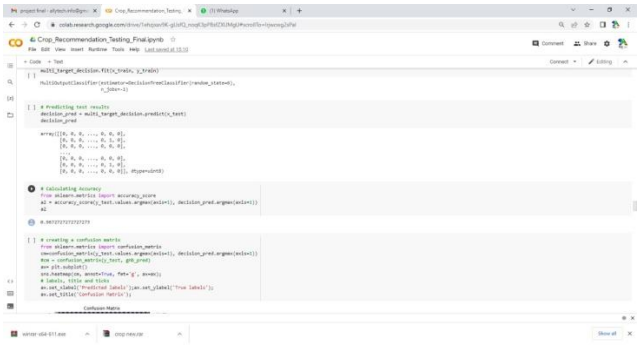


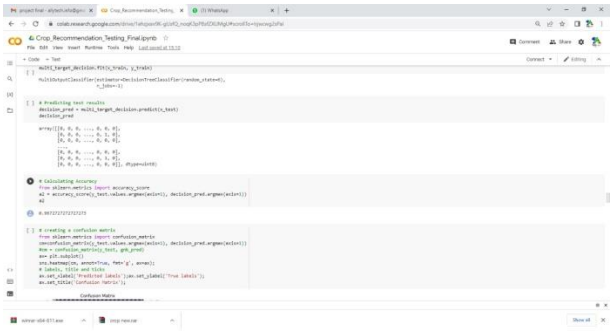
PERFORMANCE EVALUATION

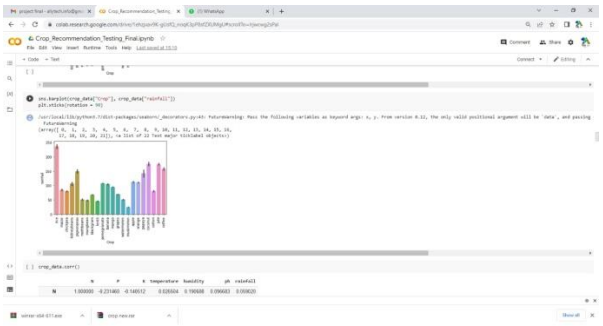
In this above graph is represent the comparison of the existing and proposed system classification algorithms. In KNNAlgorithm is provide the accuracy is higher than the other classification method.

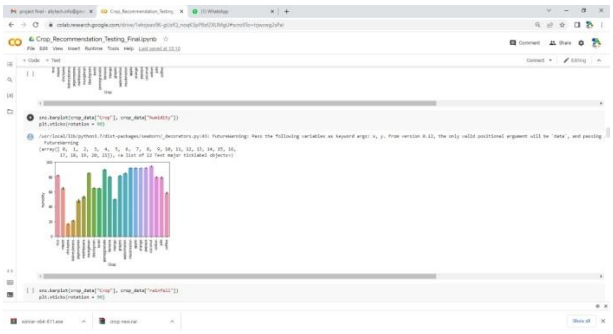
SCREEN SHOT

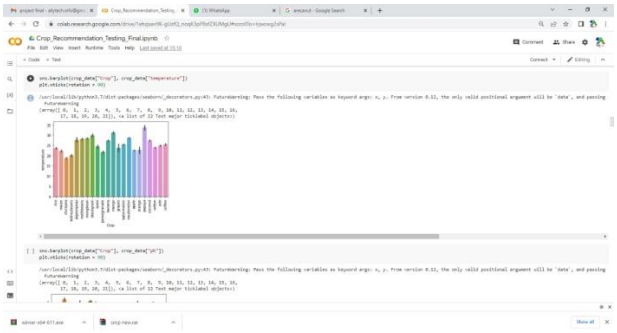


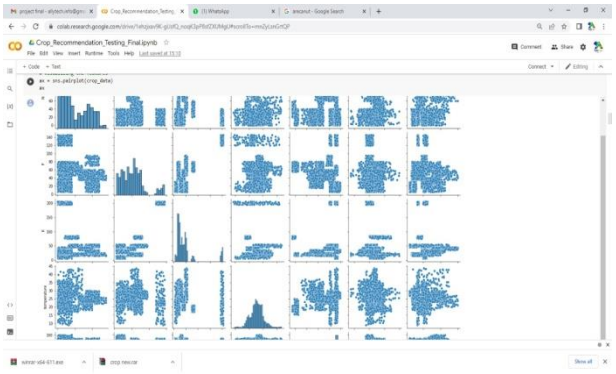












CONCLUSION

This system focuses on the vaticination of crop and computation of its yield with the help of machine literacy ways. Several machine learning methodologies used for the computation of delicacy. Random Forest classifier was used for the crop vaticination for chosen quarter. Enforced a system to crop vaticination from the collection of once data. The proposed fashion helps growers in decision timber of which crop to cultivate in the field. This work is employed to search out the gain knowledge about the crop that can be stationed to make an effective and useful harvesting. The accurate vaticination of different specified crops across different sections will help growers of india. This improves our Indian frugality by maximizing the yield rate of crop production.

References

[1] Gabriel M. Alves, Paulo E. Cruvinel, "Big Data environment for agricultural soil analysis from CT digital images", published in Semantic Computing (ICSC), IEEE Tenth International Conference,Feb 2016.

[2] Pallavi V. Jirapure, Prof. Prarthana A. Deshkar "Qualitative data analysis using Regression method for Agricultural data", published in Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), World Conference, 29 Feb/March 2016.

[3] Mohammed Zakariah, "Classification of large datasets using Random Forest Algorithm in various applications: Survey" published in International Journal of Engineering and Innovative Technology (IJEIT), Volume 4, Issue 3, September 2014.

[4] Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A., and Jain, A. K., "Dimensionality Reduction Using Genetic Algorithms," IEEE Trans. Evolutionary Computation, vol. 4, no. 2, pp. 164-171, July 2000.

[5] Narendra, P.M. and Fukunage, K., "A Branch and Bound Algorithm fo Feature Subset Selection," IEEE Trans. Computers, vol.6, no. 9, pp. 917-922, Sept. 1977.

[6] Pudil, P., Novovicova, J., and Kittler, J., "Floating Search Methods in Feature Selection," Pattern Recognition Letters, vol.15, pp. 1119-1125, 1994.

[7] Roberto B., "Using mutual information for selecting features in supervised neural net learning," IEEE Transactions on Neural Networks, 5(4):537-550, 1994.

[8] Zhang, H. and Sun, G.. Feature selection using tabu search method. Pattern Recognition, 35: 701-711, 2002.

[9] Tyrone T. Lin, Chung-Shiao. Hsieh, "A Decision Analysis for theDynamic Crop Rotation Model with Markov Process's Concept" published in Industrial Engineering and Engineering Management (IEEM), IEEE International Conference, 10-13 Dec. 2013.

[10] Snehal S.Dahikar1, Dr.SandeepV.Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach" published in International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, Vol. 2, Issue 1, January 2014.

[11] Wu Fan, Chen Chong, GuoXiaoling, Yu Hua, Wang Juyun, "Predictionof crop yield using big data", Published in: Computational Intelligence and Design (ISCID), 2015 8th International Symposium, 12-13 Dec. 2015.

[12] Snehal S. Dahikar, Sandeep V. Rode, PramodDeshmukh, "An Artificial Neural Network Approach for Agricultural Crop Yield Prediction Based on Various Parameters", published in International Journal of Advanced Research in Electronics and Communication Engineering