A Survey on Crop Recommendation Using Machine Learning

M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran

Abstract--- Agriculture arranging assumes an imperative job in any nation. Agriculture segment gives different yields, for example, sustenance, crude material for industry, affordable lift and business. The Agriculture part contains huge information regarding factors influencing its info and yield. With advances in innovation different information mining systems are presented. These information mining methods can be utilized to dissect the multidimensional, time explicit information of horticulture area to create powerful learning from it which can be utilized to support the economy. Today, the term information mining [1][2] is an interdisciplinary procedure of breaking down, handling and assessing this present reality datasets and forecast based on the discoveries. Our case-based investigation gives observational proof that we can utilize diverse information mining arrangement calculations to group the dataset of horticultural districts based on soil properties. Moreover, we have explored the most performing calculation having amazing expectation exactness to suggest the best harvest for better yield. The proposed framework will coordinate the information got from archive, climate office and by applying machine learning calculation: Multiple Linear Regression, an expectation of most reasonable yields as indicated by current natural conditions is made. This furnishes an agriculturist with assortment of alternatives of harvests that can be developed. This exploration goes for examination of soil dataset utilizing information mining procedures. It centers around characterization of soil utilizing different calculations accessible. Another essential design is to foresee untested traits utilizing relapse procedure, and usage of computerized soil test grouping.

Keywords--- Data mining, classification, regression, soil testing, agriculture, Machine Learning.

I. INTRODUCTION

Agriculture part is a standout amongst the most critical division of numerous nations. Agriculture part gives nourishment, crude material and business. Agriculture segment faces numerous issues, for example, sporadic precipitation, floods, draft, environmental change and so on. To beat these issues mechanical arrangement is required which can encourage the ranchers. The profitability of cultivating isn't just rely upon regular assets yet it likewise relies upon info gave to the framework The Input gave to the framework are Type of Soil, Availability of Water, Type of Fertilizers, Weather conditions and Crops. These information sources can increment or pronouncements the profitability of any harvest. Customary Crop Selection isn't exact and does not have any examination subtleties. To beat such issues information mining and machine learning procedures can be utilized. In Agriculture division informational collection for long time term is accessible.

Information mining is use to examine the dataset and concentrate data from it. Machine Learning calculations can be utilized to anticipate the best possible crop with the goal that it will prompt less misfortune in data sources and builds the benefit.

Decent variety in India is extraordinary which speaks to assortments of physical highlights and social highlights. Practically all families in India are reliant on Agriculture and callings identified with horticulture. IoT is assuming significant job in Agriculture which is helping agriculturists from numerous issues and to concentrate on other related callings. Exactness Agriculture is a standout amongst the best creations. It is teaching ranchers from numerous points of view like anticipating infection in cutting edge with the goal that agriculturist can take activities and get kept from misfortune and prescribing crop reasonable for his field, climate data is given just as it likewise gives data about advertising and he can send out his items and looking after field. Sensors and actuators mechanize his errand, for example, water system and utilization of pesticides in proportionate. With every one of these systems he can boost benefit and can ceaselessly screen his field.

In India, exactness horticulture isn't greatly assessed. Presently multi day we found that consistently the earth is changing persistently which is destructive to the harvests and driving agriculturists towards obligation and suicide. Much of the time this way and with developing populace to augment yield ranchers are utilizing more pesticides and manures which are prompting the soil barrenness just as diminishing the holding limit of soil and expanding poisonous quality of soil. Cultivating land is utilized by developing industrialization, so again expanding rate of the soil contamination which influences the nature of plants. Different uses of exactness Agriculture are: expectation of infections, expectation of climate anticipating, arrangement of soil, checking crop, yield forecast, programmed water system framework, and so on.

Information mining is an imperative zone of present day inquire about world for handling, investigating and assessing extensive datasets; to distinguish affiliations, arrangements, and grouping, and so forth.; between various characteristics and anticipate the best outcomes with important examples [1] [2] [15]. Essentially, these strategies can be utilized in the field of Agriculture and can deliver uncommon critical advantages and expectations that can be utilized for business and logical purposes. Generally, Agriculture basic leadership depends on specialists' decisions [3] and these decisions may not have any

significant bearing to group the soil reasonableness and may lead the lower edit yield. The unequivocal dataset the board by the information mining systems and calculations contain the colossal expository potential for exact and legitimate outcomes and these can mechanize the order procedure, contingent upon the predefined parameters created by Agriculture look into focuses.

Choice tree [4], Naïve Bayes calculation [5][6][7], Rule-Based arrangement, Neural Networks[5][6], Support Vector Machine [8] and Genetic Algorithm [9] and so on., are very notable calculation for information characterization and further for learning revelation. In this exploration, we expected to comprehend the related area, broke down the conduct of various information mining order calculations on the soil dataset and assessing the most prescient and exact calculation. The dataset has been gathered from various soil reviews that were led at various horticultural zones situated in Kasur District, Punjab, Pakistan.

II.   RELATED WORK

Agricultural crop suggestion frameworks are accessible in the market which consider different parameters like climate at the time the yield is to be planted, soil type, geology of the locale, temperature and precipitation in the area, showcase costs of the harvest, trim length, and so on. Research has been completed in this field and the accompanying papers have been alluded with the end goal of research and study.

The framework in paper [1] recommended by creators S.Pudumalar and related co-creators utilizes a gathering system called Majority Voting Technique which consolidates the intensity of numerous models to accomplish more prominent forecast precision. The strategies utilized are Random Trees, KNN, CHAID and Naïve Bayes for gathering so that regardless of whether one strategy predicts erroneously, alternate models are probably going to make right forecasts and since the greater part casting a ballot system is utilized, the last expectation is right one. On the off chance that rules are the fundamental segments which are utilized in the expectation procedure.

The exactness got is 88% utilizing the gathering model. Paper [2] is an audit paper for concentrate different calculations and their exactness in the rural field proposed by Yogesh Gandge and Sandhya. It was seen that Multiple Linear Regression gave a precision of 90-95% for rice yield. Choice tree utilizing ID3 calculation was considered for soybean edit and the proposals were created. The third calculation was SVM which was utilized on every one of the harvests and the precision was great with computationally less prerequisites. Neural system was utilized on corn information to accomplish 95% of precision. Different calculations were additionally utilized which are KNN, C4.5, K-implies, J48, LAD Tree and Naïve Bayes. The end was that still enhancement is required for the calculations to accomplish better precision.

Being used of Data Mining in Crop Yield Prediction [3], paper [3], the dataset utilized was gathered from Kaggle.com The creator has broke down the information utilizing WEKA apparatus for calculations which are LWL, J48, LAD Tree and IBK. The exactness was estimated utilizing explicitness, affectability, precision, RMSE and mean outright blunder. For every classifier, perplexity network was utilized to get the effectively distinguished occurrences. The perception was that better precision can be acquired if pruning is utilized.

Paper [4] displayed by Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh proposed utilization of seven machine learning methods for example ANN, SVM, KNN, Decision Tree, Random Forest, GBDT and Regularized Gradient Forest for crop choice. The framework is intended to recover every one of the yields sowed and time of developing at a specific time. Yield rate of each harvest is acquired and the crops giving higher yields are chosen. The framework additionally proposes an arrangement of crops to be planted to get the higher yields.

Prof. Rakesh Shirsathand other co-creator in paper [5] proposed a framework which causes the clients to settle on choices with respect to the crop to be planted. The framework utilized is a membership based framework which would have customized data of each rancher enlisted. The framework incorporates a module which keeps up the data of the past harvests planted gathered from different sources and demonstrates a coordinating crop that can be planted. The entire procedure is finished with the assistance of fake impartial systems. Toward the end a criticism framework is given with the goal that the designer can make changes required if the rancher discovers some trouble while utilizing the framework.

Enormous Data Analysis Technology Application in Agricultural Intelligence Decision System paper creators Ji-chun Zhao and Jian-xin Guo in paper [6] considers information database as large information and derivations from the information is drawn. It considers different modules like clients, information build, area master, man-machine interface, deduction motor and learning base. The learning procurement framework gets information for the choice framework and sets up a powerful learning base to take care of the issue. The paper utilizes different Hadoop modules with the end goal of highlight extraction. It utilizes the unstructured information and procedures it utilizing NoSQL, Hive, Mahout and utilizations HDFS to store the information. The information was simply introduced for wheat edit and different crops were not considered.

RSF as referenced in paper [7] is a proposal framework for famers which considers an area location module, information investigation and capacity module, trim developing database, physiographic database. The comparative area discovery module distinguishes the areas which are like the client's areas and checks the comparable harvests that are planted in those areas. Likewise, utilizing comparability framework, the proposals for the client is produced. Area recognition module utilizes the Google API administrations to get the present area of the client to distinguish the comparative areas. Be that as it may, the framework does not inspire client criticism to enhance the procedure.

III. PREDICTION AND CLASSIFICATION

Foresee the element climate and soil condition. Prescient examination of systems utilized for dissecting the best yield generation dependent on the climate condition and soil condition dependent on grouped datasets. Climate Prediction utilizes the intensity of PCs to make a conjecture. Complex PC programs, otherwise called conjecture models, keep running on supercomputers and give expectations on numerous air factors, for example, temperature, weight, wind, and precipitation. Because of the improvement of innovation huge volumes of information is delivered, and this information is put away cropively as it is helpful for examining and distinguishing designs in the information. Built up a major information structure demonstrate which is exceptionally valuable in putting away the vast volumes of information to which the Machine Learning is connected for expectation purposes. At last forecast dataset from to grouped the harvest condition.

IV. CROP AND RECOMMENDATIONS

This framework prescribed to related yields suggestion helps ranchers related on harvests and climate estimating. Be that as it may, imperative in Agriculture is, all yields generation depends on the soils since soils are basic to horticulture improvement and harvests creation. On the off chance that soil isn't reasonable for specific harvest, ranchers can't get benefit creation. So suggest the yields with estimating of climate and related on soil will help to agriculturists for effectively distinguish reasonable crops.

V. CROP PREDICTION USING ENSEMBLING TECHNIQUE

Gathering is an information mining model otherwise called the Model Combiners that join the intensity of at least two models to accomplish better expectation, cropivity than any of its models could accomplish alone. In our framework, we utilize a standout amongst the most recognizable Ensembling strategy called Majority Voting system .In the casting a ballot method any number of base students can be utilized. There ought to be no less than two base students. The students are picked such that they are fit to one another yet being complimentary moreover. Higher the challenge higher is the opportunity of better expectation. Be that as it may, it is vital for the students to be complimentary in light of the fact that when one or couple of individuals make a blunder, the likelihood of the rest of the individuals amending this mistake would be high. Every student incorporates itself with a model. The model gets prepared utilizing the preparation informational index gave. At the point when another information must be characterized, each model predicts the class individually. At long last, the class which is anticipated by greater part of the students is casted a ballot to be the class name of the new example.

VI. A COMPARATIVE STUDY OF SOIL CLASSIFICATION

The order of soil was viewed as basic to contemplate in light of the fact that relying on the fruitfulness class of the soil the area learning specialists figures out which yields ought to be gone up against that specific soil and which

manures ought to be utilized for the equivalent. The accompanying segment portrays Naive Bayes, J48, JRip calculations quickly.

6.1 Support Vector Machine

Support vector machines (SVM) is set of directed learning procedures utilized for order, relapse and exception's revelation. it's an arrangement procedure. Here, we tend to plot each data thing as some degree in n-dimensional house (where n is assortment of choices you have) with the value of each component being the value of a chosen arrange. it's a grouping procedure. amid this algorithmic principle, we tend to plot each data thing as some degree in n-dimensional house (where n is assortment of alternatives you have) with the value of each component being the value of a chosen facilitate. A Support Vector Machine (SVM) is discriminative classifier effectively limited by an isolating hyperplane. In elective words, given marked instructing data (directed learning), the algorithmic guideline yields relate degree best hyperplane that classifies new models. Bolster vector basic machine (SVM) might be a lot of administered learning procedures utilized for arrangement, relapse and anomaly's revealing.

6.2 Naïve Bayes

It isn't single calculation, yet a family of algorithmic standards. All innocent Bayes financial classifiers embraces that the estimation of a specific element is free of the estimation of some other element, given the class variable. Innocent Thomas Bayes classifier could be a direct probabilistic classifier that works upheld applying hypothesis (from Bayesian insights) with vigorous gullible freedom suppositions. it's a grouping method upheld Bayes' hypothesis with partner degree supposition of autonomy between indicators.

In clear terms, a Naive Thomas Bayes categoryifier expect that the nearness of a particular element in an exceedingly class is unlike the nearness of the other element. for instance, an organic crop could likewise be thought of to be relate degree apple if it's red, round, and concerning a couple of creeps in width. Regardless of whether these highlights rely upon one another or upon the presence of alternate highlights, a credulous Bayes classifier would consider these properties to freely add to the likelihood that this organic crop is an apple. These Learners foresee the class mark for every one of the preparation informational index. The class name that is anticipated by most of the models is casted a ballot through the lion's share casting a ballot method and the class name of the preparation informational index is chosen. From the ensembled models the principles are created.

6.3 MULTI-LAYER PERCEPTRON (ARTIFICIAL NEURAL NETWORK)

Multi-Layer perceptron (MLP) could be a feed forward neural system with a few layers among information and finished result layer. Feed forward infers that learning period one way from contribution to yield layer.

MLPs square measure normally utilized for example characterization, acknowledgment, expectation and estimation. Relate ANN depends on a lot of associated units or hubs alluded to as fake neurons (comparable to natural neurons in partner creature cerebrum). each affiliation (neurotransmitter) between neurons will transmit a sign from one to an alternate. The accepting (postsynaptic) substantial cell will strategy the signal(s) at that point flag neurons associated with that. Counterfeit neural systems (ANNs) or connectionist frameworks square measure registering frameworks aroused by the natural neural systems that speak to creature cerebrums

6.4 J48 (C4.5)

J48 is an open source Java execution of the C4.5 calculation in the Weka information mining apparatus. C4.5 is a program that makes a choice tree dependent on a lot of marked info information. This choice tree would then be able to be tried against concealed marked test information to measure how well it sums up. This calculation was produced by Ross Quinlan. It is an expansion of Quinlan's prior ID3 calculation. C4.5 utilizes ID3 calculation that represents consistent trait esteem ranges, pruning of choice trees, rule determination, and so on. The choice trees produced by C4.5 can be utilized for characterization, and therefore, C4.5 is regularly alluded to as a factual classifier [6].

6.5 JRIP

This calculation executes a propositional rule student, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as a streamlined variant of IREP. In this paper, three grouping procedures (credulous Bayes, J48 (C4.5) and JRip) in information mining were assessed and thought about on premise of time, exactness, Error Rate, True Positive Rate and False Positive Rate. Ten times cross-approval was utilized in the analysis. Our examinations demonstrated that J48 (C4.5) show ended up being the best classifier for soil tests.

VII.   COMPARISION WITH EXISTING SYSTEM

SNO	AUTHOR	TITLE	MDETHODOLOGY USED	RESULT
1	S.Pudumalar, E.Ramanujam, 2016	Crop Recommendation System for Precision Agriculture	1. Random tree 2. CHAID 3. KNN 4. Naïve Bayes 5. WEKA tool	1. Pre-processing of data 2. Handling missing and out-of-range values 3. Feature extraction 4. Ensemble model to get higher accuracy 5. Rule generation
2	Yogesh Gandge, Sandhya, 2017	A Study on Various Data Mining Techniques for Crop Yield Prediction	1. Attribute selection 2. Multiple Linear Regression 3. Decision Tree using ID3 4. SVM 5. Neural Networks 6. C4.5 8. K-means and KNN	1. Selection of agricultural field 2. Selection of crop previously planted 3. Input from user 4. Preprocess 5. Attribute Selection
3	Miftahul Jannat Mokarrama, 2017	RSF: A Recommendation System for Farmers	1.Location Detection 2.Data analysis and storage 3.Similar location detection 4. Recommendation generation module.	1. Physiographic, thermal, crop growing period, crop cropion rate 2. Seasonal crop database 2. Similar location detection 3. Generating the set of crops 4. Similarity between the crops planted in a region
4	Prof. Rakesh Shirsath, 2017	Agriculture decision support system using data mining	1.Subscription based system 2. ANN 3. Android application 4. Personalized content	1. Android app with a login module 2. Previously planted crops known to system 3. User feedback mechanism 4. Maintenance of crop.
5	Ji-chun Zhao, Jian-xin Guo, 2018	Big Data Analysis Technology Application in Agricultural Intelligence Decision System	1.Inference engine 2.Domain expertise 3.Knowledge engineering 4.Knowledge acquisition module 5.Knowledge base for recommendation system	1. Large database of crops 2. Processed using Hadoop 3. Professional knowledge 4. Past experiences 5. Feature selection using HDFS 6. Future Scope: Using Hadoop with Artificial Neural Networks.
6	Shruti Mishra Priyanka Paygude, 2018	Use of Data Mining in Crop Yield Prediction	1. J48 2. LAD tree 3. LWL 4. IBK algorithm	1. WEKA tool 2. LAD tree showed the lowest accuracy 3. Errors can be minimized by pruning the tree 4. IBK was observed to achieve higher accuracy

(Row 1 top cell continued: 6. Classification algorithm on data 7. Crop is recommended)

VIII.   CONCLUSIONS

This paper introduces an overview that utilizing Data digging systems for climate expectation yields great outcomes and can be considered as an option in contrast to conventional metrological methodologies.

The investigation depicts the abilities of different calculations in foreseeing a few climate wonders, for example, temperature, rainstorms, precipitation and inferred that real systems In this paper, we have proposed an examination of the soil information utilizing distinctive calculations and forecast strategy. From the investigation in this paper, we presumed that there is as yet a need of research in the Agricultural field to improve precision. Utilizing group techniques is a decent method to guarantee better precision of the framework. Additionally, on the off chance that we need to think about just a single calculation for the proposal framework, we can utilize SVM because of its basic computational necessities.

In future work crops proposal makes difference agriculturists related on crops and climate determining. Yet, imperative in Agriculture is, all yields creation depends on the soils since soils are essential to horticulture advancement and harvests generation. On the off chance that soil isn't appropriate for specific harvest, ranchers can't get benefit generation. So prescribe the crops with determining of climate and related on soil will help to ranchers for effectively recognize reasonable harvests. In our proposed work we execute framework for agriculturists to suggest the harvests based climate forecast and reasonable soil. We anticipate the sort of crop which one is reasonable for that specific soil, climate condition, temperature, etc. So for, we are utilizing machine learning with the arrangement of dataset we are recognize the harvest for the relating soil. This proposed framework serves to ranchers to precisely recognize crops without stressing of future climate and coordinated soils. By this framework agriculturists can get more harvests generation and benefit.

REFERENCES

1. 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC), "Crop Recommendation System for Precision Agriculture", S.Pudumalar*, E.Ramanujam*, R.HarineRajashreeń, C.Kavyań, T.Kiruthikań, J.Nishań.
2. 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), "A Study on Various Data Mining Techniques for Crop Yield Prediction", Yogesh Gandge, Sandhya.
3. 2017 IEEE Region 10 Humanitarian Technology Conference, "RSF: A Recommendation System for Farmers", Miftahul Jannat Mokarrama; Mohammad Shamsul Arefin.
4. 2017 International Conference on I2C2, "Agriculture decision support system using data mining", Prof. Rakesh Shirsath; Neha Khadke; Divya More.
5. 2018 the 3rd IEEE International Conference on Cloud Computing and Big Data Analysis, "Big Data Analysis Technology Application in Agricultural Intelligence Decision System", Ji-chun Zhao; Jian-xin Guo. Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018), "Use of Data Mining in Crop Yield Prediction", Shruti Mishra, Priyanka Paygude, Snehal Chaudhary, SonaliIdate.
6. IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO) 2015, "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction", Aakunuri Manjula, Dr.G .Narsimha
7. 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", Rakesh Kumar , M.P. Singh , Prabhat Kumar and J.P. Singh
8. Wang, Yuhong, Jiangrong Tang, and Wenbin Cao. "Grey prediction model-based food security early warning prediction." Grey Systems: Theory and Application 2, no. 1 (2012): 13-23.
9. Liu Z., Meng L., Zhao W. and Yu F. Application of ANN in food safety early warning, The 2nd International Conference on Future Computer and Communication, Wuhan, Vol. 3, 2010, pp. 677-80.
10. AnshalSavla, Parul Dhawan, HimtanayaBhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture", Innovations in Information ,Embedded and Communication systems (ICIIECS), 2015
11. Aakunuri Manjula, Dr.G .Narsimha, "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction" , Conference on Intelligent Systems and Control (ISCO), 2015
12. Yash Sanghvi, Harsh Gupta, Harmish Doshi, DivyaKoli, AmoghAnshDivyaKoli, Umang Gupta , "Comparison of Self Organizing Maps and Sammon's Mapping on agricultural datasets for precision agriculture", International Conference on Innovations in Information ,Embedded and Communication systems, 2015
13. 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
14. A.T.M Shakil Ahamed, NavidTanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman , "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" , (SNPD) IEEE/ACIS International Conference., 2015
15. Yang, Liying. "Classifiers selection for ensemble learning based on accuracy and diversity." Procedia Engineering 15 (2011): 4266-4270.
16. Aymen E Khedr, Mona Kadry, Ghada Walid, "Proposed Framework for Implementing Data Mining Techniques to Enhance Decisions in Agriculture Sector Applied Case on Food Security Information Center Ministry of Agriculture, Egypt", International Conference on Communications, management, and Information technology (ICCMIT'), 2015.
17. Monali Paul, Santosh K. Vishwakarma, Ashok Verma , "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", International Conference on Computational Intelligence and Communication Networks, 2015.
18. Pudumalar, S., E. Ramanujam, R. HarineRajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." In Advanced Computing (ICoAC), 2016 Eighth International Conference on, pp. 32-36. IEEE, 2017.

19. Daryl H. Hepting, Timothy Maciag, Harvey Hill, "Web-Based Support of Crop Selection for Climate Adaptation", 45th Hawaii International Conference on System Sciences, 2012.
20. S. Dhamodaran, Allwyn raja, "Prediction of Weather Condition Using Probability Function", International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 10, Number 4 (2015) pp.10665-10670.