INTERNATIONAL JOURNAL OF RESEARCH AND DEVELOPMENT IN ENGINEERING SCIENCES
e-ISSN: 2582 - 4201 Volume: 3 | Issue: 1 | Year: 2021

A Survey on Prediction of Indian Crop Yield using Machine Learning

RAJANA PRIYANKA

Software Engineer,
Veniso Solutions Private Limited (Hyderabad), India

Abstract *– Now a days the current situation faced by farmers in Maharashtra, we have observed that there is an increase in suicide rate over the years. The reason behind this includes weather conditions, debt, family issues and frequent change in Indian government norms. Sometimes farmers are not aware about the crop which suits their soil quality, soil nutrients and soil composition. The work proposes to help farmers check the soil quality depending on the analysis done based on data mining approach. Thus the system focuses on checking the soil quality to predict the crop suitable for cultivation according to their soil type and maximize the crop yield with recommending appropriate fertilizer.*
Keywords: Farmers, API (Application Programming Interface), Machine learning, Crop Yield,

I. INTRODUCTION

Agriculture is the backbone of every economy. In a country like India, which has ever increasing demand of food due to rising population, advances in agriculture sector are required to meet the needs. From ancient period, agriculture is considered as the main and the foremost culture practiced in India. Ancient people cultivate the crops in their own land and so they have been accommodated to their needs. Therefore, the natural crops are cultivated and have been used by many creatures such as human beings, animals and birds. The greenish goods produced in the land which have been taken by the creature leads to a healthy and welfare life. Since the invention of new innovative technologies and techniques the agriculture field is slowly degrading [1].

Due to these, abundant invention people are been concentrated on cultivating artificial products that is hybrid products where there leads to an unhealthy life. Nowadays, modern people don't have awareness about the cultivation of the crops in a right time and at a right place. Because of these cultivating techniques the seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food [2]. By analyzing all these issues and problems like weather, temperature and several factors, there is no proper solution and technologies to overcome the situation faced by us. In India there are several ways to increase the economical growth in the field of agriculture. There are multiple ways to increase and improve the crop yield and the quality of the crops.

Data mining also useful for predicting the crop yield production. Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining software is an analytical tool that allows users to analyze data from many different dimensions or angles, categorize, and summarize the relationships identified [4]. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The patterns, associations, or relationships among all this data can provide information.

As per the statistics of 2019 around 200+ million farmers dwell in Maharashtra. With this myriad number of farmers and increasing suicide rates, we want to help farmers to understand the importance of prior crop prediction, to flourish their basic knowledge about soil quality, understanding location-wise weather constraints, in order to achieve high crop yield through our technology solution. Most of the existing system is hardware based which makes them expensive and difficult to maintain. Also they lack to give accurate results. Some systems suggest crop sequence depending on yield rate and market price [5].

The system proposed tries to overcome these drawbacks and predicts crops by analyzing structured data. The project being "Prediction of soil quality using data mining approach" certainly focuses on agricultural aspects. Being a totally software solution, it does not allow maintenance factor to be considered much. Also the accuracy level would be high as compared to hardware based solutions, because components like soil composition, soil type, pH value, weather conditions all come into picture during the prediction process. Information can be converted into knowledge about historical patterns and future trends [2].

II. LITERATURE SURVEY

Agriculture sector plays a major role in Indian economy, as 70 percent households in India depends purely on this field. Agriculture in India contributes to about 17% of Gross Value Added as of 2015-16.But there is a continuous decline in agriculture's contribution to Gross Value Added. Food is essential for life and we depend on agricultural outputs, so farmers play a very important role [6].

A Machine Learning (ML) deals with problems where the relation between input and output variables is not known or hard to obtain. The "learning" term here denotes the automatic acquisition of structural descriptions from examples of what is being described. Unlike traditional statistical methods, ML does not make assumptions about the correct structure of the data model, which describes the data [14].

This characteristic is very useful to model complex non-linear behaviors, such as a function for crop yield prediction. ML techniques most successfully applied to Crop Yield Prediction (CYP).Supervised learning algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables).

Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc [7].

The following comparison is shown below:

The study in used Multiple Linear Regression (MLR) technique for crop analysis. Decision tree algorithm and Classification is used to perform analysis of over 362 datasets and provide result. The training dataset here is classified into as organic, inorganic and real estate for predicting the type of soil. Results computed by this system are accurate as well as reliable [8].

The study in field data to a Back Propagation Network to evaluate the test data set. Back Propagation Network uses a hidden layer which helps in better performance in predicting soil properties. Back Propagation Network here, is employed to develop a self-trained function to predict soil properties with parameters.

Bosubabu Sambana et al, "Machine Learning Convergence for Weather based Crop Selection and Processing for Smart Agriculture", proposed new mechanism on crop yield and selection processing for agricultural activities. Agriculture plays a vital role in Indian economy. It contributes 20% of total India's GDP. In India, most of the crops are solely dependent upon weather conditions. Hence, more yields of crops can be achieved by analyzing agro climate data using machine learning techniques [3].

Data Mining is emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Data mining techniques are the better choice for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production.

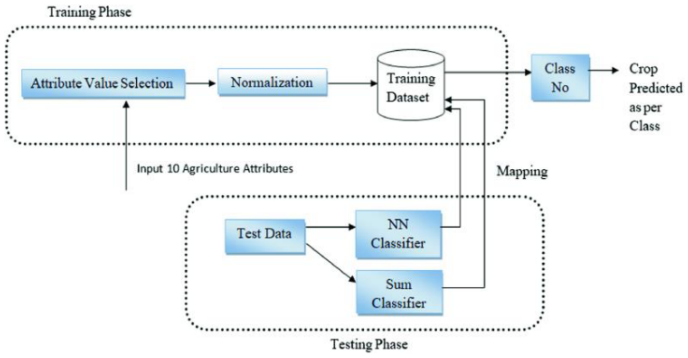

Figure. 1: Data pre-processing analysis

This research proposes and implements a system to predict crop yield from previous data. This is achieved by applying association rule mining on agriculture data. This research focuses on creation of a prediction model which may be used to future prediction of crop yield. This paper presents a brief analysis of crop yield prediction using data mining technique based on association rules for the selected region i.e. district of Tamil Nadu in India. The experimental results shows that the proposed work efficiently predict the crop yield production [3].

This gives more accuracy and performs better than the traditionally used methods, however, sometimes the system becomes slow and inconsistency is seen in the output. In two regressions supervised machine learning methods are used: Support Vector Machine (SVM) and Relevance Vector Machine (RVM) to show effectiveness in soil quality prediction. A smart wireless device for sensing soil moisture and meteorological data [9].

The wireless device gives an error rate of 15% and 95% accuracy. However, it has not been tested for real time data. The paper involves a check for Soil Fertility and Plant Nutrient by using back propagation algorithm. The results are accurate and enable improvement in soil properties. It performs better as compared to traditional methods. However, system is slowing inefficient and not stable. According to paper , three methods are used which includes Decision tree, Naive Bayes Classifier, and KNN Classifier which analyses soil and predicts crop yield, However rule based induction and SVM can be used for more accuracy as results are not accurate [1][2].

III. RELATED WORK

A case study of interpreting paddy distributions of three counties on Northern Taiwan during two crop seasons on year 2000 using multi-temporal imageries together with cadastre GIS by Bayesian posteriori probability classifier was studied. In order to integrating Bayesian conditional probability, priori probabilities of paddy's attributes were estimated from photogrammetric interpretation results provided by the Food Bureau, and the spectrum reflectance from different growth stages was used.

Bosubabu Sambana, "Smart Agricultural activities Monitoring and Control system using an Internet of Things", proves his complete research work on agricultural activities on Abstract—An Internet of things is a technology that encounters various remote monitoring and controlling of devices, which is connected [14]to it wirelessly and here the concepts of wireless sensors networks and Ethernet protocols are made of use. This research paper mainly concerns its application towards agricultural and Irrigational activities such soil moisture monitoring and water pressure monitoring and soil testing along with Protection against trespassing with motion detections and conservation of energy [4][5].

This paper implementation holds division of systems into multiple to one of which is the main system and the rest three are the subsystems [14]. The subsystem 1 consists of soil moisture monitoring control, and water pressure sensing along with pump set and vibration alert sensor for motion monitor in and around the related field. The second subsystem will be in use for protection against trespassing with detection of proximity of the object due to motion along the required field fence and will have voltage and current detection sensors for controlling, monitoring efficient energy usage [5].

The subsystem three will be in use for controlling that is taking all possible measures in accordance with the available current status of the incorporated alerted sensors wherein a user can control remotely access the sprinklers and protective measurable things like electrifying the fences and present activate status. The main system will take in the information from 1 and 2 the subsystem communicating through ZigBee communication and will be monitoring and transmitted to the network Router through ZigBee from where the required data will be displayed on to a web home page with necessary Ethernet protocols and current running data [5].

Due to the spatial heterogeneous of paddy's distribution, classifier parameters were established individually on each map-quadrangle. Temporal change of NDVI from different growth stages pass through rice's life cycle has been measured and [14] we find two-stage images make significant improvement on classification results. Results of the study help us to evaluate the accuracy of the classifier. Imagery classification results were compared with aerial photo's interpreting results for assessing accuracy [9].

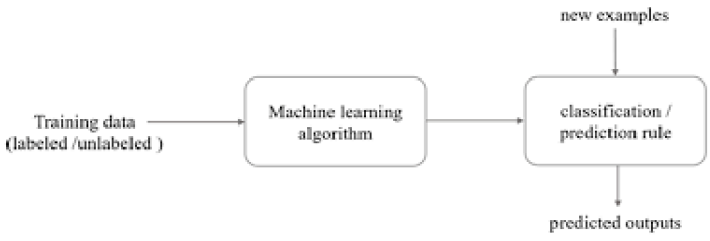

Figure. 2: Internal training Mechanism

Bayesin classifier has advantages including easy-to-adjusted and easy-to-computed rules and comparative stable results when limited SPOT satellite imageries

available. Bayesin method also provides results with probability that help the operator to assess the places having least confidence. These advantages allow us to suggest Bayesian method be used in paddy-area investigation in India. Studies conducted by agricultural researchers in Pakistan have shown that attempts of crop yield maximization through propesticide state policies have led to a dangerously high pesticide usage. These studies have reported a negative correlation between pesticide usage and crop yield. Hence excessive use of pesticides is harming the farmers with adverse financial, environmental and social impacts [10].

Study had shown that how data mining integrated agricultural data including pest scouting, pesticide usage and meteorological recordings is useful for optimization of pesticide usage. Unsupervised clustering of the data was performed first through Recursive Noise Removal (RNR). These clusters reveal interesting patterns of farmer practices along with pesticide usage dynamics and hence help identify the reasons for this pesticide abuse A mechanism of performing the mapping from nominal to numeric values(actually ranking) based on the transmittance as well as the statistical properties of the plants was proposed [11].

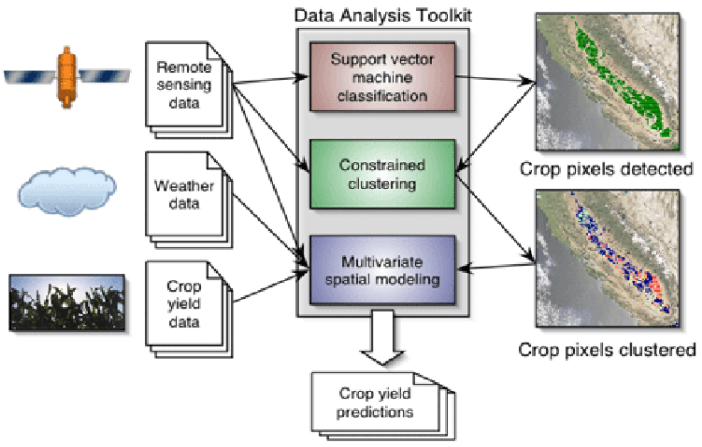

Figure. 3: Crop yield data analysis process

Spectral analysis (using chemical means) is a tedious and time consuming process, thus difficult to repeat, each and every time, for classification of (numerically) unclassified cotton varieties. A supporting statistical method was also proposed based on linear regression curve fitting using normalized nominal attributes. Subsequently a rank is assigned to the variety based on its R2 value and slope of the plot. This rank thus becomes the numeric equivalent of the nominal alphanumeric name of the variety being considered. Spectral analysis of 12 cotton varieties in the visible and near infra red regions was also performed. A 60% classification accuracy was achieved i.e., correspondence in order of ranking generated through regression, as compared with the spectral rank order.

A process model for analyzing data, and describes the support that Weka to Environment for Knowledge Analysis (WEKA) provides for this model. The domain model learned by the data mining algorithm can then be readily incorporated into a software application. This WEKA based analysis and application construction process was illustrated through a case study in the agricultural domain i.e., in mushroom grading. Effect of pesticides on humans can't be directly checked because of the poisonous nature of pesticides, therefore the usage of pesticides on cotton crop has been taken into consideration for the purpose [12].

The COF Clustering Tool cannot only be used for pesticide data, but also possesses the flexibility to deal with any numeric data. Spatial data mining methods to extract interesting and regular knowledge from large spatial databases of agriculture were studied aiming at discerning trends in agriculture production with reference to the availability of inputs. The predicted and real vs. Counter graph illustrated how closely the poly analyst prediction follows the actual value of the attribute over the range of the dataset.

Applying the data mining techniques to agriculture the target for different food grains can be achieved. Their study demonstrated the scope for application of spatial mining tools for a utility study and analysis. The specific application of Polyanalyst gave a clear scope for evaluation and comparison of predicted and real values. Influence of climatic factors on major kharif and rabi crops production in Bhopal District of Madhya Pradesh State was studied [13].

To improve the productivity of the item in this manner supporting both farmer and nation we have to use the development which evaluates the idea of gather and giving proposals. The Internet of things (IoT) is patching up the agribusiness drawing in the ranchers by the expansive grouping of procedures, for example, exactness and traditionalist development to go facing

difficulties in the field. IoT progression helps in get-together data on conditions like climate, temperature, and profitability of soil, reap web viewing draws in the territory of weed, level of water, bug affirmation, creature obstruction into the field, adjust improvement, development. IoT uses ranchers to get identified with his habitation from any place and at whatever point. Remote sensor structures are utilized for checking the homestead conditions and little scope controllers are utilized to control and robotize the property shapes [3][5].

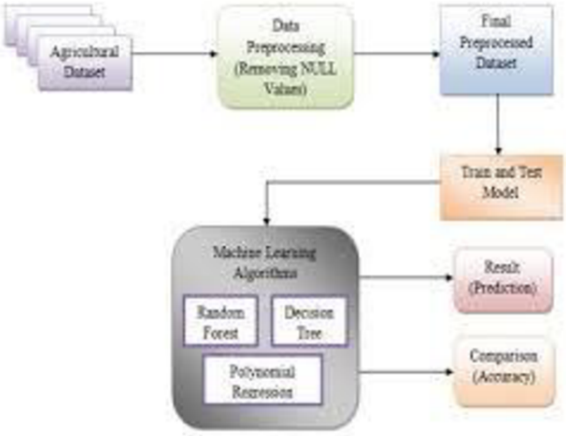

Figure. 4: Block diagram of working mechanism

The findings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by Relative humidity followed by rainfall and temperature. The decision tree analysis indicated that the productivity of paddy crop was mostly influenced by Rainfall followed by Relative humidity and Evaporation [10]. For Wheat crop the analysis indicated that the productivity is mostly influenced by Temperature followed by Relative humidity and Rainfall. The findings of decision tree were confirmed from Bayesian classification[14]. The decision tree in the study area fast to execute and much to be desired as representations of knowledge interpretations. The rules formed from the decision tree are helpful in identifying the conditions responsible for the high or low crop productivity [11].

Random Forest Classifier:

Random forest is a most popular and powerful supervised machine learning algorithm capable of performing both classification, regression tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction. Random decision forests correct for decision trees habit of over fitting to their training set [7].

The data sets considered are rainfall, perception, production, temperature to construct random forest, a collection of decision trees by considering two-third of the records in the datasets. These decision trees are applied on the remaining records for accurate classification [8]. The resultant training sets can be applied on the test data for correct prediction of crop yield based on the input attributes .RF algorithm was used to study the performance of this approach on the dataset [12]. The advantage of random forest algorithm is, over fitting is less of an issue with Random Forests, unlike decision tree machine learning algorithms. There is no need of pruning the random forest. Random Forest machine learning algorithms can be grown in parallel.

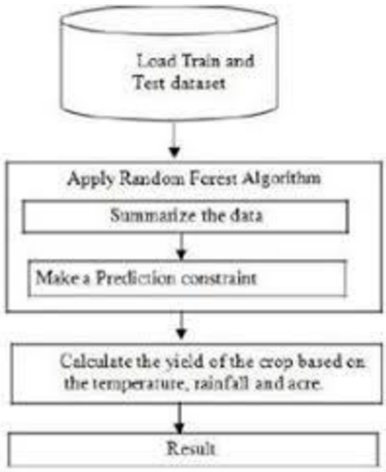

Figure. 5: Random Forest Classifier model

Application of Data Mining Techniques in Agriculture

Many techniques for learning rules and relationships automatically from diverse data sets were developed, to simplify the often tedious and error-prone process of acquiring knowledge from empirical data. While these techniques are plausible, theoretically well-founded, and perform well on more or less artificial test data sets, they depend on their ability to make sense of real-world data. This paper describes a project that is applying a range of machine learning strategies to problems in agriculture and horticulture.

They briefly surveyed some of the techniques emerging from machine learning research, describe a software workbench for experimenting with a variety of techniques on real-world data sets, and describe a case study of dairy herd management in which culling rules were inferred from a medium-sized database of herd information. They also described a range of machine learning strategies to problems in agriculture and horticulture. They briefly surveyed some of the techniques emerging from machine learning research, described a software workbench for experimenting with a variety of techniques on real-world data sets, and described a case study of dairy herd management in which culling rules were inferred from a medium-sized database of herd information.

IV. CONCLUSION

This paper review on a system which will help farmers to have an idea of crop yield estimates based on various weather parameters and area under cultivation using this farmer can make decisions on whether to grow that particular crop or go for alternate crop in case yield predictions are unfavorable.

References

[1] P.Priya et al., " Predicting Yield of the Crop using Machine Learning Algorithm", International Journal of Engineering Sciences & Research Technology ISSN: 2277-9655 , April 2018

[2] N G Yethiraj, International Journal of Business Intelligent Volume: 01 Issue: 02 December 2012,Pages No.40-42 ISSN: 2278-2400

[3] Bosubabu Sambana, " Machine Learning Convergence for Weather based Crop Selection and Processing for Smart Agriculture ", Journal of Xidian University, ISSN No:1001-2400, Volume-14 , Issue-11, 2020 , Page No: 351–364 ,

[4] Bosubabu Sambana, " Air Pollution Monitoring and Prediction System using the Internet of Things " , International Journal of Research and Development in Engineering Sciences , ISSN:2582-4201 , Volume: 2, Issue: 3 , Year: 2020

[5] Bosubabu Sambana, " Smart Agricultural activities Monitoring and Control system using an Internet of Things" , Journal of Xidian University, ISSN No:1001-2400 , Volume-14 Issue-7, July 2020,Page No: 1645 – 1654

[6] Bosubabu Sambana, "Signal Processing and Display of Linear Frequency Modulated Continuous Wave Radar on a Chip" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, Special Issue , NCETC-17, April 2017.

[7] Uppe Nanaji, Prof. S.Pallam Setty, "Investigation Misbehavior Configuration Nodes with Secure Neighborhood on Energy Consumption for DYMO Routing Protocol in MANETS", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 10, Issue 3, December 2020.

[8] Bosubabu Sambana, "Identifying Datasets by Pattern Recognition Techniques using Neural Networks" Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol. 6 No. 1, 2017, pp.34-39.

[9] Uppe Nanaji, Prof. S.Pallam Setty, "Investigate the Impact of Buffer to Provide Security for DYMO in MANETS", Journal of Science, Technology and Development, Volume 9, Issue 12 December 2020.

[10] Bosubabu Sambana, "Internet of Things: Applications and Future Trends", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 3, March 2017, ISSN (Online): 2320-9801,

[11] Bosubabu Sambana, "Signal Processing and Display of Linear Frequency Modulated Continuous Wave Radar on a Chip" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, Special Issue , NCETC-17, April 2017.

[12] Bosubabu Sambana , " An Internet of Things Approach to Monitoring and Control System of Agricultural Activities", March 2020 , DOI: 10.13140/RG.2.2.33522.63688

[13] Uppe Nanaji, Prof. S.Pallam Setty, "Investigating the Impact of Network Parameters on Wormhole Attack with ANODR Routing Protocol in MANET", Journal on Wireless Communication Networks, Volume 8, Issue 4, January-June 2020.

[14] Bosubabu Sambana, "Integrative Spectrum Sensing in Cognitive Radio using Wireless Network", Springer Nature Conference - International conference on Remote Sensing for Disaster Management (ICRSDM-2017) during 11-13 October 2017 at Andhra University-Visakhapatnam. First Online: 24 June 2018, Publisher Name: Springer, Cham. Print ISBN: 978-3-319-77275-2, Online ISBN: 978-3-319-77276-9

Rajana Priyanka received the B.Tech in Electronics and Communication Engineering from Avanthi Institute of Engineering and Technology, India in 2017.

She is currently working as a Software Engineer in Veniso Solutions Private Limited, (Hyderabad) India.