A Survey on Agriculture Crop Recommender System

Saiyyad Mohmmad Ali Muzffar Ali.

***Abstract*— Agriculture planning plays an important role in any country. Agriculture sector provides various outputs such as food, raw material for industry, economical boost and employment. The agriculture sector contains vast data with respect to factors affecting its input and output. With advances in technology various data mining techniques are introduced. These data mining techniques can be used to analyze the multidimensional, time specific data of agriculture sector to produce effective knowledge from it which can be used to boost the economy. In this paper we survey various approaches used by different researcher to deal with analysis of agriculture sector to provide effective way to increase production and effective use of resources which in turn increases the economy in this sector.**

***Index Terms*—Agriculture Sector, Data Mining, Multidimensional, Time Specific Data.**

I. INTRODUCTION

Agriculture sector is one of the most important sector of many countries. Agriculture sector provides food, raw material and employment. Agriculture sector faces many problems such as irregular rainfall, floods, draught, climate change etc. To overcome these problems technological solution is needed which can help the farmers. The productivity of farming is not only depend on natural resources but it also depends on input provided to the system The Input provided to the system are Type of Soil, Availability of Water, Type of Fertilizers, Weather conditions and Crops. These inputs can increase or decrees the productivity of any crop. Traditional Crop Selection is not precise and does not have any analysis details. To overcome such problems data mining and machine learning techniques can be used. In agriculture sector data set for long time duration is available. Data mining is use to analyze the dataset and extract information from it. Machine Learning algorithms can be used to predict the proper crop so that it will lead to less loss in inputs and increases the profit.

II. LITERATURE SURVEY

In [1], Fertilizer type and quantity for the farm based on soli type, soil nutrient, water availability, weather conditions. Sensors are used to collect the data across the field. Wireless sensor network is used to store the data in database. The data get analyzed and used in the predictive model to recommend the crop.

In [2], The model uses SMS or email to deliver advisory services to the farmers. It uses Static information, semi dynamic information, and dynamic information as input to track growth of each crop. The model provides three advisories which are informational general, information specific and action specific. It uses data of Yield at the time of harvest for accumulating historical data.

In [3], Crop related information is given to farmers in this model. The information related to Varity differ with respect to location. Farming situation has also some impact on such information. Crop performance depends on soil type, Geographic location seasonal conditions and nutrition. The model uses data mining techniques to explain crop performance variability. The model can be used to understand impact of location and weather condition on crops. It also be useful for farmers to identify high performing verities according to the location and conditions. The model defines correlation between the traits and varieties

In [4], The dependency between the moment of prediction and the accuracy of the forecast is studied. The linear model is used for yield forecasting. Then, the model is extended with non-linear components to improve the accuracy of the forecasts. The Crop yield of a particular culture is predicted using the model. The model uses function having parameter as values of vegetation indices during growth and ripening period of the crop. Forecast of the crop yield is depend on historical track record of the indices. The linear model assumes that the soil and climate have small variation. If the amount of statistical data is not sufficient the number of adjustable parameter got reduced. The Factors which cause difference in crop yield are soil fertility, climatic difference, amount of solar radiation.

In [5], Feed management is been discussed in this model. Clustering of feed resources is done using data mining techniques for efficient management of feed resources. Different clustering techniques are discussed for classification of resources into different group on the basis of accuracy of the technique used. To evaluate and to suggest a best technique for clustering feed resources on two different data sets containing 236 and 106 feed resources three approaches are used which are K-means, spectral k-means and auto spectral .The major constituents of the feed resources are crude protein, crude fiber ash, fat etc. The

Manuscript received Feb, 2018.
***Saiyyad Mohmmad Ali Muzffar Ali**, Department of Computer Engineering, R. C. Patel Institute of Technology, Shirpur, Shirpur-425405, India.*

output of the clustering technique were validated on the basis of precision, recall and F-measure. During the evaluation process K-means was the best among the three techniques used for clustering of feed resources. K-means is not being effected by sample size and distribution of the feed resource in data sets. The clustering of feed resource gives opportunity to the farmers to choose feed resources on the basis of locality, price and effectiveness. The correct choice leads to balanced diet and profit due to proper feed management.

In [6], The author proposes a model to discuss the currant usage of land and agriculture land vanished in past seven years. Data mining tools are used to analyze the data of land to determine land vanished in last seven years, usage of land and amount of rice production in last seven years. The model used to analyze classification of land across six districts as Coimbatore, Erode, Dindigul, Salem&Namakkal, Dharmapuri in Tamilnadu state. In India. K-means clustering algorithm is used to make cluster. Clusters are analyzed to understand usage of land and its impact on production in last seven years.

In [7] The author proposes the model evaluate the applications of agriculture loans using different data mining techniques. It uses comparative study of decision tree, neural network, Logistics regression on the basis of their misclassification rate. The model uses data of loans of grain producer for a specific period. The data is divided into different categories according to different factors. The result of the model shows importance of different models for borrowers of single loan with respect to borrower with more than one loan.

In [8], Crop yield production is discussed. The crop production is depends on geographical, biological, climatic factors. Data mining techniques can be used to extract knowledge from this data so that crop production can be estimated. Accuracy of crop production is depend on the historical data about crop yield. Scheduling of production is depends on crop size. K-means method is used to classify soil and plants. SVM is used to analyze the atmospheric changes. The study analyze the data collected from Kolhapur district from Maharashtra State and effect of different data mining technique to analyze the data.

In [9] an early warning prediction for food security risk is developed. The quality index model is constructed. The quality index model is used to measure level of food quality. The improved grey prediction model is also introduced in the study. The grey prediction model is used to predict quality index for food. The food security risk is analyze by comparing the prediction of the model and expert opinion.

In [10], the data mining technique to predict food quality using a back-propagation neural network is developed. If the predicted data is near the threshold value error may occur. To overcome such problem data near threshold values are used during training process.

In [11], The system uses various data mining algorithms. It performs comparative analysis for the yield prediction of soybean crop. Seed, yield, production and area harvested are the parameters during the analysis phase of the data. Support Vector Machine, Neural Network, Random Forest, REPTree Bayes and Bagging. The Bagging Algorithms is the bset algorithm among the other algorithms for crop yield prediction of soybean crop for the given dataset.

In [12], the model uses the concept of precision agriculture to effectively use inputs and produce maximum outputs. Data mining techniques has effective role in precision agriculture. The productivity of the crops can be assessed on the basis Temperature Condition Index (TCI), Vegetation Condition Index (VCI) and Normalized Difference Vegetation Index (NDVI).The System developed for eXtensible Crop Yield Prediction Framework (XCYPF). The framework provides facility to select crops, dataset, dependent and independent variable. The framework uses temperature data, weather condition to predict crop yield for sugarcane and rice crop.

In [13], the author developed a model to introduce the technique to reduce high dimensional data to smaller size. Self-organizing maps and multi-dimensional scaling techniques are used reduce the data. In SOM set of data items is projected onto a regular two-dimensional grid. The sammon's mapping is used to match original distances of data items with their projections. The study of the model for agriculture data reveals that SOM is suitable when dataset is large and sammon's mapping is useful when datset is small
In [14], The System addresses selection of crop issue. Crop selection is based on production rate, government policies and market price. If more than one crop are available then the system provides the best solution for such situations. The System provides crop selection method according to which crop can be chosen and which leads to high economic growth. Crop production rate depends upon region, weather condition, soil type, soil composition and harvesting method. Input for the system is sowing time, plantation days and predicted yield rate and it produces output as a sequence of crops whose production per day are maximum over season
In [15], data mining techniques are used to predict Annual yield of major crops and recommender system is used to recommend the crops for specific regions. The dataset contains attributes as rainfall, temperature variation, humidity, irrigated area, sunshine, soil salinity, soil ph value. Clustering of the district is made on the basis of similar attributes. K-means clustering algorithm is used to make the cluster. Linear regression k-NN, Neural net are used obtain the crop yield prediction results. X-means operator is used to provide K-means clustering. Four cluster are formed on the basis of weather attribute, soil salinity and soil-ph, irrigated area, individual crop yields. Crop recommendations are based on crop species per hectare area and its net worth.
In [16], the system developed to address the problem of selection of classifier for ensemble learning. The system proposed the method to select set of classifier from pool of classifiers. Classifier diversity is considered during the selection process. Classifier correlation is calculated on the basis of Q statistics diversity measures which are based on correlation between errors. Classifier selection algorithm works on the principle of selection of accuracy. The algorithm trains set of classifiers then examine their accuracy and removes the week classifier so that remaining classifier get combined and ensemble learning can be used to evaluate

the performance. Classifier can also be selected using Selection by Accuracy and Diversity. In this one of the Classifier with most accuracy is selected. Diversity is calculated between selected and not selected classifier. Strong diversity classifier is selected into ensemble and process get repeated till the requirement

In [17], the model is developed to overcome the problem of food security in Egypt. The model calculate the need of Egyptian people for different crops. The model uses multilayer perceptron in the ANN as a function for predicting the values.WEKA tool is used to build the prediction model. The model shows the difference between available and required amount of food. The model deals with the production of wheat, beans, rice. The model can be extended for more number of food items and it can effectively provide food insecurity details according to the data given to it.

In [18], In the system soil dataset is classified on the basis of data mining techniques. The predicted category will indicate the yielding of the crop.Naïve Bayes and k-Nearest Neighbor algorithms are used for crop yield prediction. Soil is categories into high, low, medium category. Categorization of soil leads to crop yield prediction. Rapid Miner tool is used to implement data mining technique. The model can be enhanced by using different categorization technique such as SVM, PCA.

In [19], the model is developed to recommend the crop for respective soil type. The model uses the concept of precision agriculture. On the basis of soil type, crop yield precision agriculture suggest the proper crop to the farmers. The System uses majority voting technique. Random tree, CHAID, K-Nearest Neighbor and Naive Bayes are used to provide the recommendations.

III. CONCLUSION

We discuss various approaches presented by different researchers for agriculture data analysis. The basic data mining approaches such as clustering, classification is done by algorithms such K-means, SVM and PCA etc. The Agriculture sector contains dataset which is multidimensional. To deal with it more algorithms are need to address this sector effectively.

REFERENCES

[1] Daryl H. Hepting, Timothy Maciag, Harvey Hill, "Web-Based Support of Crop Selection for Climate Adaptation", 45th Hawaii International Conference on System Sciences, 2012.

[2] Satish Babu, "A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India,2013.

[3] D. Diepeveen and L. Armstrong, "Identifying key crop performance traits using data mining" World Conference on Agriculture, Information and IT, 2008.

[4] K. Gorokhovskyi, V. Ignatiev, and A. Murynin, "Efficiency of crop yield forecasting depending on the moment of prediction based on large remote sensing data set," in Proceedings of the International Conference on Data Mining, Las Vegas Nevada, USA (CSREA Press USA, 2013, pp. 36–41.

[5] Rani, V. Uma. "Efficient management of feed resources using data mining techniques." PhD diss., Christ University, 2010.

[6] Megala, S., and M. Hemalatha. "A novel data mining approach to determine the vanished agricultural land in Tamilnadu." International Journal of Computer Applications 23, no. 3 ,1-6, 2011.

[7] Salame, Emile. "Applying data mining techniques to evaluate applications for agricultural loans." 2011.

[8] Raorane, A. A., and R. V. Kulkarni. "Data Mining: An effective tool for yield estimation in the agricultural sector." *International Journal of Emerging Trends of Technology in Computer Science* 1, no. 2 (2012): 75-79.

[9] 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.

[10] 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.

[11] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, 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

[12] Aakunuri Manjula, Dr.G .Narsimha, "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction" , Conference on Intelligent Systems and Control (ISCO), 2015

[13] Yash Sanghvi, Harsh Gupta, Harmish Doshi, Divya Koli, Amogh Ansh Divya Koli, 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

[14] 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

[15] A.T.M Shakil Ahamed, Navid Tanzeem 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

[16] Yang, Liying. "Classifiers selection for ensemble learning based on accuracy and diversity." *Procedia Engineering* 15 (2011): 4266-4270.

[17] 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.

[18] 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.

[19] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, 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.

First Author Saiyyad Mohmmad Ali Muzffar Ali has completed his ME in Computer Engineering. He is currently working at R. C. Patel Institute of technology, Shirpur.