International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Machine Learning Technique for Crop Recommendation in Agriculture Sector

Nitin N. Patil, Mohmmad Ali M. Saiyyad

Abstract: ***Numerous efforts have been demonstrated through various innovations to lead towards the betterment of agriculture sector till now. Effective and innovative use of Science and Technology can help to improve crop quality and production, yield prediction and crop disease analysis. Agriculture sector provides various productions such as food, raw material for industry and has a significant impact on economy and employment of a country. The agriculture sector contains huge data with respect to factors affecting its input and output. With advances in technology, various data mining techniques have been introduced accordingly. Also different analytical models like Decision Tree, Random Forest, Support Vector Machine and Bayesian Neural Network are available for required analysis. These data mining techniques can be used to analyze the multidimensional, time specific data of agriculture sector to produce effective knowledge in support of the efforts to boost the economy. These methods can be further used to analyze soil, climate, moisture, humidity, temperature. In this paper, we used Naive Bayesian Classification technique to identify the class of crop and used the food grain dataset to analyze the technique over different attribute. Further we used Machine Learning approach for the accurate crop recommendation. Crop recommendation will help for effective decision making to the end user to take proper decision according to the output of the system.***

Keywords: Agriculture Sector, Data Mining, Multidimensional, Time Specific Data, Machine Learning.

I. INTRODUCTION

Agriculture sector provides food, raw material and employment and many important products which increases the revenue of any country. Agriculture sector faces many issues such as floods, draught, plant diseases etc. To overcome these problems technological solutions are needed which can help the farmers. The productivity of farming is not only dependent on natural resources but it also on input provided to it. It includes type of soil, availability of water, type of fertilizers, weather conditions and type of crops. These inputs have a great influence on increase or decrease the productivity of any crop. Traditional crop selection is not precise and not based on any analytical details. To overcome such problems data mining and machine learning techniques can be used appropriately. In agriculture sector, seasonal dataset for various crops are available. Data mining is used to analyze the dataset and extract information from it. Machine Learning allows to procure good extent to examine the massive data which can be used to predict the proper crop so that it will lead to less loss and consequently increase in the profit.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Dr. Nitin N. Patil, Department of Computer Engineering, R. C. Patel Institute of Technology, Shirpur, India (e-mail:er.nitinpatil@gmail.com).

Mr. Mohammad Ali M. Saiyyad, Department of Computer Engineering, R. C. Patel Institute of Technology, Shirpur, India (e-mail: sayyad1188@gmail.com).

II. LITERATURE SURVEY

The model introduced by D. Diepeveen and L. Armstrong describes the crop yield and crop related information. Information regarding the crop is basically based on the linked location of the crop. Information obtained from the model has an impact of farming conditions. Crop growth depends on soil type, geographic location, seasonal conditions and nutrition. The said 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 [1].

Satish Babu suggested farmers about the crops using SMS and email service in his proposed model. Static, dynamic and semi dynamic information are used to track the growth of the plants [2].

Daryl H. Hepting et al. has used sensor for collecting the crop data. Fertilizers are important according to the soil type, soil nutrient, water availability, weather conditions. Network across the field can be used to store the data onto the cloud. Various analytical techniques provide the predictive functionality for recommendation system [3].

A liner model is used to provide the prediction of the crop by K. Gorokhovskyi et al. also nonlinear components are used to improve the accuracy. The model provides information regarding growing period of the crop. Soil and Climate are the linear parameter in the system. If the amount of statistical data is not sufficient the number of adjustable parameter get reduced [4]. Feed management is the important aspect in agriculture field. Feed management is been discussed with their model by Rani and V. Uma. Clustering of feed resources is done using data mining techniques for efficient management of feed resources. Clustering techniques have been used to make classification of resources [5]. Megala S. and M. Hemalatha [6], proposed a model to understand the usage of agriculture land and land vanished in last several years.

Data mining techniques are utilized to analyze and understand the data of 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. The 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 [6].

Data mining technique has been utilized by Salame and Emile to evaluate the application of loan. Various data mining techniques have been used for comparative study of decision tree, neural network, logistics regression on the basis of their misclassification rate. Here the model provides the information regarding loan application according to various parameters. The model utilizes the time series data of loan. 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 [7].

Raorane, A. A has developed model using data mining techniques crop production analyisis. The crop production depends on biological, climatic factors and availability of resources. Data mining techniques can be used to extract knowledge from this data so that crop production can be estimated. Accuracy of crop production is dependent on the historical data about crop yield. Scheduling of production depends on crop size. SVM is used to analyze the atmospheric changes.

A model which can give an early warning regarding food is proposed by Wang et al. As food is the most important part in every human being's life, a constructive model is proposed which can predict the food security warning. The quality index model is constructed. Food quality is predicted in this model. The food security risk is analyzed by comparing the prediction of the model and expert opinion [9].

Another model is proposed by Liu Z et al. using the data mining technique to predict food quality using a back-propagation neural network. 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 [10].

The system developed by Anshal Savla et al. uses various data mining algorithms. Soyabean crop is analyzed for comparison purpose. Various parameters are considered for analysis purposes. Support Vector Machine, Neural Network, Random Forest, REPTree Bayes and Bagging. The Bagging Algorithms is the best algorithm among the other available algorithms for crop yield prediction of soybean crop for the given dataset [11].

The proposed model uses the concept of precision agriculture to effectively use inputs and produce maximum outputs. Data mining techniques have effective role in precision agriculture. Various index like VCI, TCI, NDVI are used to calculate the productivity. The System proposed 'eXtensible Crop Yield Prediction Framework' (XCYPF). The framework provides facility to select crops, dataset, dependent and independent variables. The framework uses temperature data, weather condition to predict crop yield for sugarcane and rice crops [12].

Yash Sanghvi et al. proposed a model to introduce the technique to reduce high dimensional data to smaller size. Self-organizing map (SOM) with multidimensional scaling is used to scale down large data into smaller parts. In SOM set of data items is projected onto a regular two-dimensional grid. The salmon' 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 salmon's mapping is useful when dataset is smaller [13].

The system demonstrated by Rakesh Kumar et al. addresses selection of crop issue. Crops are selected according to market trends, use and various policy of government. If more than one crops are available then the system provides the best solution for such situations. This model provide a unique way to identify and select crop according to various selection criteria. Crop growth has attributes like area, climate, soil nutrients, 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 the season [14].

A.T.M Shakil Ahamed et al. used data mining techniques to predict annual yield of major crops. Also the recommender system was 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 to obtain the crop yield prediction results. X-means operator provided the K-means clustering. Four clusters 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 it's net worth [15].

The system proposed by L Yang 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 [16].

A model is developed to recommend the crop for respective soil type. The model uses the concept of precision agriculture. Crop recommendation to the farmers is based on various attributes. The System uses majority voting technique. Random tree, CHAID, K-Nearest Neighbor and Naive Bayes techniques are used to provide the recommendations [17].

III. RECOMMENDED PARAMETER

In Agriculture sector, the crop yield is depend upon various factors like temperature, soil, moisture, rainfall, humidity. We will understand the impact of above factors with respect to crop yield [4-17].

- **Temperature**

Temperature is the important aspect in the crop recommendation system. Every crop has maximum and minimum temperature variance. Low temperature may lead to damage to the crop such as chilling injury, freezing injury and suffocation. High temperature may lead to problem like reduction in nutrient absorption, photosynthesis disruption and plant death.

- **Soil**

Soil is the important part in the crop recommendation process. The soil has nutrients such as Nitrogen, Phosphors, Potassium, Magnesium, and Sulfur. The different types of soil according to the nutrients are Clay, Sandy, Silty, Peaty, Chalky, and Loamy.

- **Moisture**

Moisture in soil is the main component for proper growth of plant. For recommending the crop moisture also plays the important role in the system.

- **Rainfall**

Rainfall is the main source for water for many crops. Crops are taken according to the rainfall. Crops which requires more water will be taken in the area where sufficient rainfall occur. Rainfall also leads to availability of various water resources due to which various different crops can be taken

- **Humidity**

Humidity is important to make photosynthesis possible. If the Humidity is low plant growth is compromised, if the humidity is high the plant quality may get affected.

IV. METHODOLOGY

In this approach, we used machine learning for the intended recommendation. In this modified algorithm, we use Naive Bayesian classification for crop and fertilizer recommendation. The training data with the class label is given to the model. After training, the testing data is used to test the working of model.

- Naive Bayesian Classification

Naive Bayes classifier works on Bayes theorem. Naive Bayes Classifier is used to assign label to the class where the class labels are drawn from some finite Set. It works on the principle that the features are independent from each other.

The Class label are identified on the basis of the input dataset [20].

- **The Algorithm**

Following are the steps for the modified algorithm:

Step 1: Input the area of crop filed details.
- It contains the land area and area under cultivation.

Step 2: Read the dataset for the training purpose with class labels.
- Dataset contain the parameter for recommendation with class labels.

Step 3: Preprocess the data to remove inconsistency and outlier detection.
- Unknown variables and outliers are detected and corrected to improve the accuracy.

Step 4: Apply data cleaning with suitable Technique.
- Normalization and other

Step 5: Apply the Naive Bayesian Classification technique to obtain the classes for the cleaned data.
- Output of the classifier is the label for the input data

Step 6: Generate the model to obtain the appropriate fertilizer Recommendation.
- The Model is generated with class labels which is useful for decision making.

Step 7: Obtain required Crop Recommendation from the developed model.

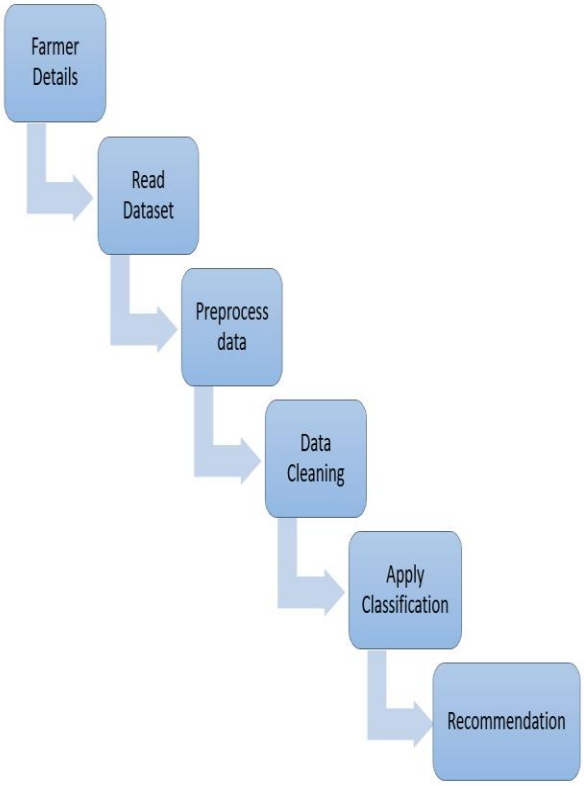

Fig 1. Modified Algorithm for Recommendation System [18]

The figure 1 illustrates various steps of modified algorithm for recommendation system. It starts with details of farmer and goes through reading, preprocessing, cleaning of data. Then classification is applied which is useful for the final stage of recommendation.

The figure 2 depicts the working flow of the described methodology [18].

The Table – I shows appropriate and accurate crop recommendation by considering various factors of soil, environment and fertilizers.

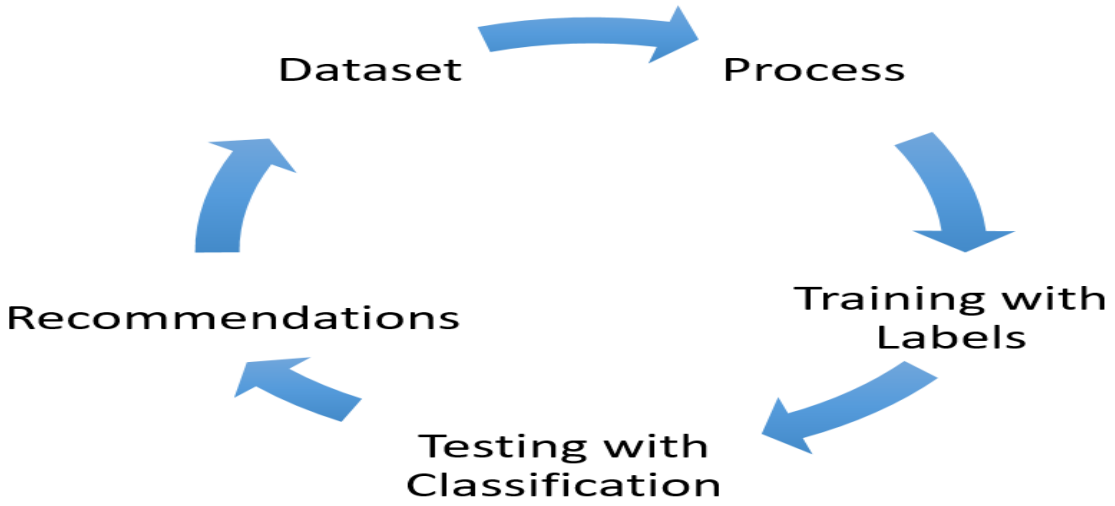

Fig 2. Methodology Working Flow

Table-I. Experimental Result for Fertilizers and Crop Recommendation Syste

Sr. No.	Crop	Season	Soil Type	Temperature	Moisture	Rainfall	Humidity	Fertilizer	Crop
1	Wheat	Winter	Sandy	Low	Low	Low	High	Yes	Yes
2	Bajara	Winter	Loam	High	Low	Medium	High	Yes	No
3	Maize	Summer	Loam	Medium	Medium	Low	Low	Yes	Yes
4	Ground Nut	Summer	Sandy	Low	High	High	Medium	Yes	Yes
5	Cotton	Rainy	Clay	High	Low	High	Low	Yes	Yes
6	Rice	Rainy	Loam	Low	High	High	High	No	Yes
7	Pulse	Winter	Clay	High	Medium	Low	Medium	Yes	Yes
8	Carrot	Rainy	Clay	Medium	Medium	High	Medium	No	Yes
9	Wheat	Winter	Clay	Medium	Low	Medium	Medium	Yes	Yes
10	Bajara	Summer	Clay	Medium	Medium	Medium	High	No	Yes
11	Maize	Winter	Clay	High	Low	Medium	High	No	Yes
12	Ground Nut	Winter	Clay	Medium	High	Low	High	No	No
13	Cotton	Summer	Sandy	Medium	Low	Medium	High	Yes	Yes
14	Rice	Winter	Sandy	Medium	Medium	High	Low	Yes	Yes
15	Pulse	Rainy	Loam	Medium	High	Medium	High	Yes	Yes
16	Carrot	Winter	Sandy	High	Medium	Low	Low	No	No
17	Wheat	Winter	Loam	High	Low	Medium	High	No	No
18	Bajara	Winter	Sandy	High	Medium	Medium	Low	Yes	Yes
19	Ground Nut	Rainy	Loam	Medium	Medium	Medium	Medium	Yes	Yes
20	Pulses	Summer	Clay	High	High	Low	High	No	Yes

V. CONCLUSION

In this paper, we described various approaches presented by different researchers for agriculture data analysis. In our implementation, we used the Naive Bayesian classification technique to recommend the crops and fertilizers. The proposed technique uses the input data more accurately using five weather and soil related parameters to obtain reliable

crop recommendations. In this approach, the results are promising and useful for crop and fertilizer recommendation which will help the farmers according to crop fields. In future work, a collaborative framework can be developed using Internet of Things based techniques to make the quick, more accurate and reliable predictions. Real time dataset using Internet of Things and machine learning technique is the next task in future work.

REFERENCES

1. D. Diepeveen and L. Armstrong, "Identifying Key Crop Performance Traits using Data Mining", World Conference on Agriculture, Information and IT, 2008.
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. Daryl H. Hepting, Timothy Maciag, Harvey Hill, "Web-Based Support of Crop Selection for Climate Adaptation", 45th Hawaii International Conference on System Sciences, 2012.
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 (2011): 1-6.
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. Liying Yang, "Classifiers Selection for Ensemble Learning Based on Accuracy and Diversity", Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS], 2011.
17. 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.
18. Han, Jiawei, Jian Pei, and Micheline Kamber. "*Data mining: Concepts and Techniques",* Elsevier, 2011.

AUTHORS PROFILE

Prof. Dr. Nitin N. Patil has completed his Ph.D. in Computer Engineering from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra, India in 2018.
He is currently working as the Head and Associate Professor in the Computer Engineering Department at R. C. Patel Institute of Technology, Shirpur (Maharashtra), India. He has more than 16 years of teaching experience. His area of interest are Information hiding, Information Security, Data Mining, Machine Learning and Internet of Things.
Dr. Nitin N. Patil is a life member of Indian Society for Technical Education and also a life member of International Association of Computer Science and Information Technology. He has published many papers in International/National conferences and Journals. He has also actively participated in various seminars and workshops.

M. M . Saiyyad, has completed his M.E. in Computer Engineering from North Maharashtra University, Jalgaon, Maharashtra, India in 2014.
He is currently working as Assistant Professor in the Computer Engineering Department at R. C. Patel Institute of Technology, Shirpur (Maharashtra), India. His area of interest are Image Processing, Data Mining and Machine Learning. He has published papers in International conferences and Journals. He has also in various seminars and workshops.