International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

Supervised Knowledge Analytics Model for Optimal Crop Production

Pratik Dubey[1], Shashikant Panday[2]
Department of CSE & IT, VNS Group of Institutes Faculty of Engineering,Bhopal (M.P), India

Abstract : India holds numerous variations in ecological and environmental conditions. It is rich in diversity. More than half of its population depends on agriculture for their livelihood. Due to large variations in environmental factors the difference in crop pattern can be seen easily.Due to variation in environmental factors, varieties of crops are there in India. There are more than 10 parameters associated with every crop. The crop related data generated throughout the year is quite large. It includes parameters such as soil type, N, P, K, soil PH, etc. Similarly, irrigation relate data is also necessary such as, rainfall, temperature, etc. Crop production can be increased in poor productivity area and better production can be seen in good productivity area.The Supervised Knowledge Analysis model is capable to extract knowledge from the raw data. The Supervised Knowledge Analysis model contains steps such as Problem objective, Data Gathering, Pre-processing, Model estimation or model design collaborated with other required technologies and interpretation and visualization. The last step includes the evaluation, validation of results. Final phase also includes the proper representation of results in order to better understand the recommendation.

Supervised Knowledge Analytics Model for optimal crop production generates recommendation for current harvested crop as well as for future crop planning.

Keywords: Data Analytics,Data Mining Basic Concepts,Data Mining Techniques for Crop Prediction

I. INTRODUCTION

In India, a major part of the population depends on Agriculture. Approximately 70 percent of the total rural area of the country uses agriculture as their primary source of livelihood. Agricultural Lands of India reflects various behaviour. Diversity in the geographical and ecological environment plays a vital role in overall crop production throughout the country. Increasing crop Production is always being a crucial task. Every crop has its certain features that hold some geographical, environmental and nutritional requirements. In Agriculture, approximately 4 to 6 crop-related data generated every year, there is a huge amount of data generated quarterly as well as annually. The data generated includes annual crop production, Overall Yield, Fertilizer information, diseases, soil type, etc. For Agriculture, some Agri - meteorological data such as temperature, rainfall, humidity, etc. can also be used for prediction and rule generation purposes. There are various methods and technologies available for various tasks in the agricultural field. Data mining is one of the most applied practices for achieving the specified intended objective. Data mining facilitates the exploration of intention - specific knowledge from the large raw heterogeneous data[29].

II. LITERATURE REVIEW

I. AnshalSavlaet, al in [10] performed a survey of a classification algorithm for formulating yield prediction accuracy in agriculture. The author explained the various classification techniques such as SVM, Random forest, Rep.Tree, Bagging, Bayes Algorithm & ANN.

II. A.T.M ShakilAhamed et. al. in [2] proposed a method to predict the annual yield of major crops in different districts of Bangladesh. The author predicts the annual crop production of Rice (AMON, AUS, BORO) & wheat crops. The author took data for 11 districts of Bangladesh.

III. Niketa Gandhi et. al in [8] performs a comprehensive review of the application of several classification and clustering techniques such as ANN, Bayesian network, SVM and association rule mining. These techniques explained with examples to reveal how these have been applied to a variety of agricultural data sets.

IV. Lei Shi et. al. in [28] have proposed a model for classification of agriculture data by using support vector machine classification algorithm. And the author also compares this algorithm with naive Bayes and neural network.

V. P. Surya et. al. in [11] have proposed a model for the Northwestern zone of Tamilnadu state to predict the area which has the highest crop yield production rate. This model predicted the production rate on the basis of harvested area and production using predictive analytic techniques.

III. DESCRIPTION

Data Analytics is a procedural phrase that focus on the numerous types of data analysis. Basically, It is the technique behind analyzing the raw data in order to generate a conclusion about the knowledge. The main objective of the analysis vary depending to the requirement. The requirement of information that needs to be extracted from the raw data makes data analytics purpose specific

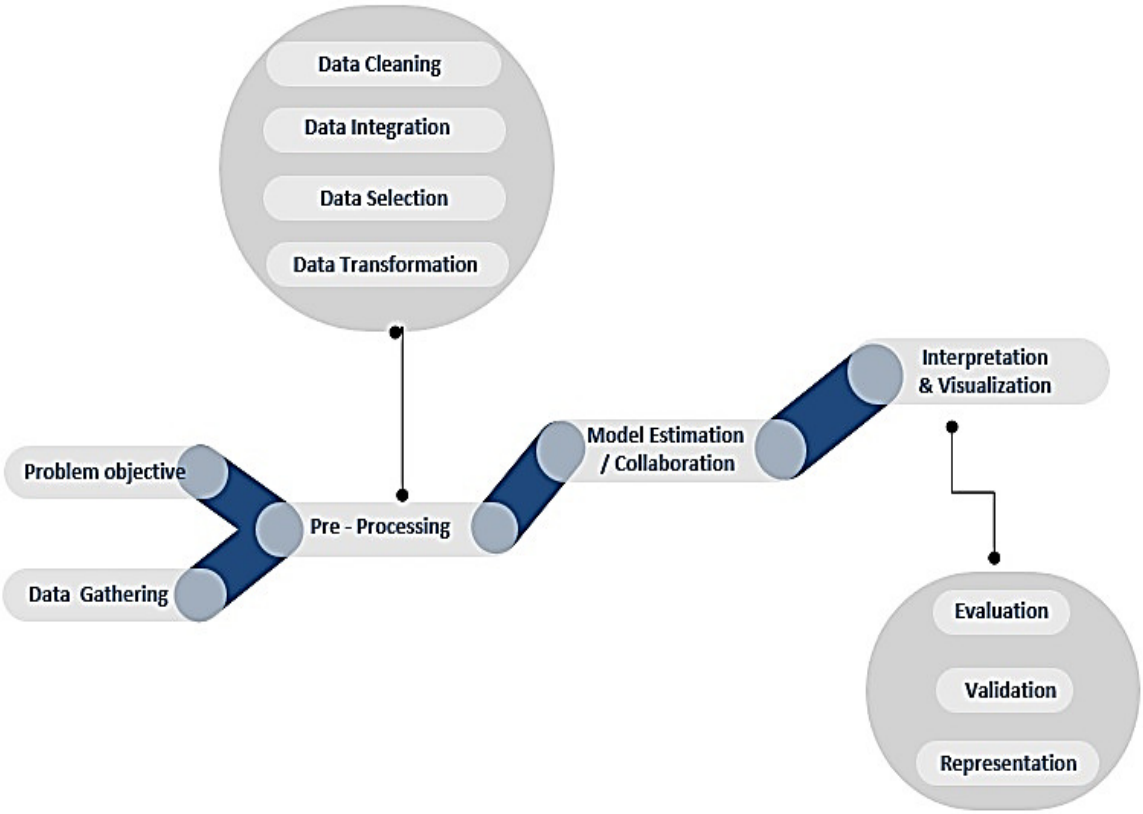

Figure 1.1: Generalized Analytical Model[29]

1. First step is data accumulation. Data can either be taken from single source or from more than one sources according to the aim behind the analysis.
2. Next, data should have properly cleaned. Missing values and redundant values can also be handled properly.
3. When data is nearly ideal to process, a model is designed for efficient analysis. The model can also be integrated along with various techniques and approaches purely determined by the objective.
4. Finally, results should be in readable form. The proper visualization is required to understand the results.

IV. RESEARCH METHODOLOGY

There is significant amount of data generation has been seen in last decade in the field of ariculture. Governments are also focusing on uploading the data digitally for processing. Digital data requires much more preprocessing. Agricultural data also contains missing and redundant values. Before moving for actual analysis the data is gone through 4 subphases of preprocessing such as Data cleaning, Integration, selection of suitable values and transformation in the processable form. The Supervised Knowledge Analysis model is capable to extract knowledge from the raw data. The Supervised Knowledge Analysis model contains steps such as Problem objective, Data Gathering, Preprocessing, Model estimation or model design collaborated with other required technologies and interpretation and visualization. The last step includes the evaluation, validation of results. Final phase also includes the proper representation of results in order to better understand the recommendation.

Preprocessing

As predictive modeling is focused on how much wheat crop is a product on the basis of area harvested, rainfall and soil type. It is a procedure associated with the data mining technique in which unanalyzed data is transformed into a usable and understandable format before actual processing. In the real world, data is often insufficient, incompatible, non-existent in certain bearing and is likely to certain many delusion. Data pre-processing is a method to overcome these issues.

1 Identify Problem Domain

In the field of Agriculture, the probability of optimizing crop production is always varies. There are various factors that affect the overall crop production. Various environmental and other parameters that need to be considered throughout the crop harvesting process.Various research institutes experimenting with the parameters for every individual crop. There is a need that those results can be utilized for proper crop planning and management

2 Data Selection (Department of Agriculture, M.P)

Table 3.1: Schema of Crop Dataset

S.no	Attributes	Type
1.	District	String
2.	Crop Name	String
3.	Soil Type	String
4.	Soil PH	Numeric
5.	Area Harvested	String
6.	Crop Production	String
7.	Crop Yield	Numeric

8.	Soil Nutrient (N)	Numeric
9.	Soil Nutrient (P)	Numeric
10.	Soil Nutrient (K)	Numeric
11.	Rainfall	Numeric
12.	Temperature	Numeric

3 Data preparation

Initially, the data was containing many erroneous values, so these are the following steps which come under data preparation:

4. Parameters Details

i. ***Area***: Area harvested.
Unit → Hectare.
ii. ***Production***: Crop Production Value.
Unit → Tonnes.
iii. ***Soil Ph***: Unit for measuring the Acidity or Basicity of soil.
iv. ***N, P, and K***: These are the Nitrogen, Phosphorous and Potash content of the soil.

3.2 Parameter Estimation

The data gathered from multiple sources is not in a required form, hence few estimations are required to evaluate the data values according to the Model requirement.

i. Rainfall Data Evaluation:

$$RD_i = \frac{1}{N} \{ \sum_{j=1}^{N} RGj \} \quad \text{.....(1)}$$

Where,
RD_i = *Rainfall in* i^{th} *District*
RGj = j^{th} *Rain Gauge Station in a* i^{th} *district.*
N = *Total no of Rain Gauge Stations*

ii. Crop Yield Estimation :

$$Crop\ Yield = Total\ Production\ \frac{Value * 1000}{Total\ Area\ Harvested} \quad \text{..... .....(2)}$$

iii. Fertilizer Per Nutrient :

Fertilizer Amount = N * i: P * j: K * k .....(3)
Per Hectare

Where,
i = Amount of Nitrogen (N)
j = Amount of Phosphorus (P2O5)
k = Amount of Potash (K2O)

Model and Illustration using Sample Dataset

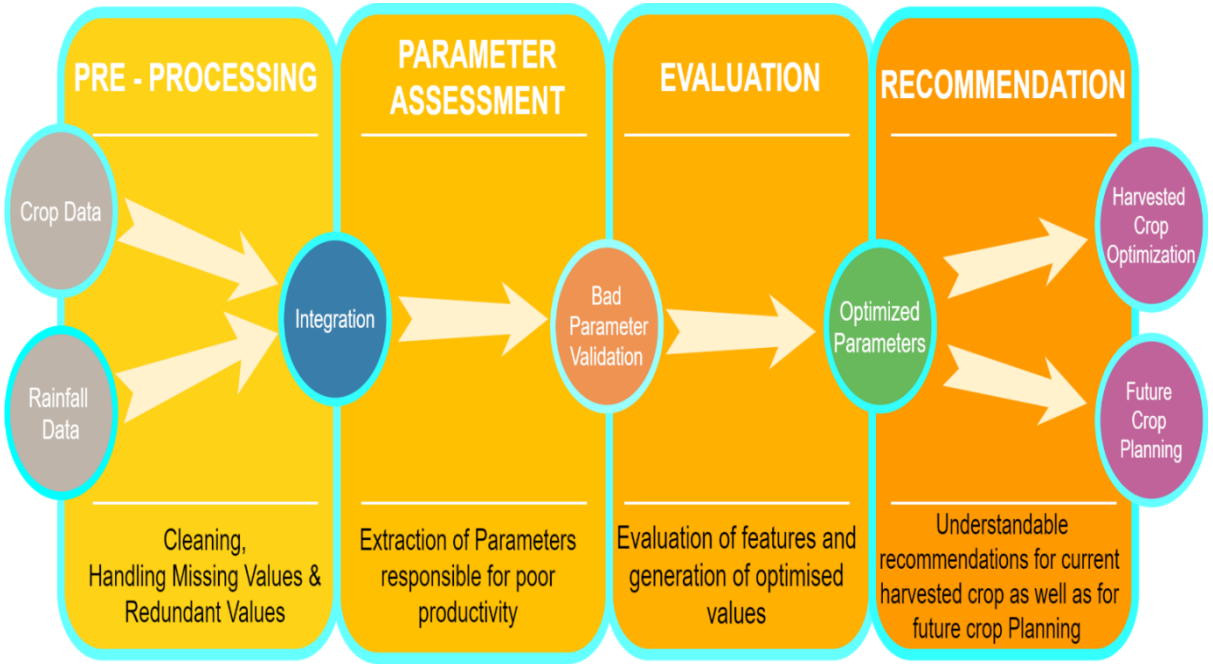

Figure 3.2: Supervised Knowledge Analytics Model

Input: Crop Dataset

Output: Recommendations for Crop Management

Step 1: Data Preprocessing

Step 2: Low production data parameters assessment.

Step 3: Evaluation of parameters value according to the Knowledge Base and validation of data values.

Step 4: Recommendation values for the current harvested crop as well as for future crop planning, better crop production, and management.

V. FUTURE WORK

At present, there are many problems in the agricultural field that need to be solved in detail. To validate the above analytics model, the data can be tested using machine learning techniques and it can be useful for better crop yield. which is useful in crop promotion. Algorithms of Data Science can be done in Efficiency Machine learning Approaches can also be incorporated along with model and knowledgebase For Real Time Efficiency.

VI. CONCLUSION

In agriculture field, there are numerous factors that affects the overall production of a crop. These factors include environmental as well as meteorological factors. In Agriculture, a huge amount of crop-related data is generating throughout the year. This data includes various parameters such as soil nutrient value, rainfall, soil type, etc. Crop production is always considered as a sensitive task for which high production has always been a priority. But at the rural level, farmers are not aware of certain factors that must need to be addressed in order to achieve high productivity of a crop.

Using Supervised Knowledge Analytics the data generated throughout the year can be analyzed for the parameters responsible for the poor productivity of a particular crop. In order to do so, the model assesses each and every record for that year and compare it with the knowledge base. The knowledge base is designed by taking experimentally and scientifically proven values of various parameters for every respective crop.

The analytical model evaluates the recommended values for the defected parameters. These parameters can be used for improving the production of the current harvested crop. The recommendations can also be used for future crop planning according to the area and other factors. The recommendations are authentic and trusted because the validation itself is done by experimentally proven values of various genuine research institutes.

REFERENCES

I. S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," *2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, 2018, pp. 796-802.

II. A. T. M. S. Ahamed, NavidTanzeem Mahmood, Nazmul Hossain, Mohammad TanzirKabir, 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," *IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Takamatsu, 2015, pp. 1-6.

III. B. Imane, B. Abdelmajid, T. A. Mohammed, T. A. Mohammed and T. A. Youssef, "Data mining approach based on clustering and association rules applicable to different fields," *International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, Kenitra, 2018, pp. 1-5.

IV. D.Rajesh, "Application of Spatial Data Mining for Agriculture", International Journal of Computer Applications (0975 – 8887) Volume 15– No.2, February 2011.

V. Miss. Snehal S. Dahikar, Dr. Sandeep V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", IJIREEICE, Vol. 2, Issue 1, January 2014.

VI. Manish Sahane, BalajiAglave, Razaullah Khan and SanjaynSirsat, "An Overview of Data Mining Techniques Applied to Agriculture Soil Data", International Journal of Agriculture Innovations and Research, Volume 3, No. 2, September 2014.

VII. M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," *International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, 2015, pp. 766-771.

VIII. N. Gandhi and L. Armstrong, "Applying data mining techniques to predict yield of rice in humid subtropical climatic zone of India," *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 1901-1906.

IX. N. Hemageetha, "A survey on application of data mining techniques to analyze the soil for agricultural purpose," *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 3112-3117.

X. A. Savla, N. Israni, P. Dhawan, A. Mandholia, H. Bhadada and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2015, pp. 1-7.

XI. P. Surya, Dr. I. Laurence Aroquiaraj, "Crop Yield Prediction in Agriculture Using Data Mining Predictive Analytic Techniques", IJRAR December 2018.

XII. K. Pavya, Dr. B. Srinivasan, "Feature Selection Techniques in Data Mining: A Study", IJSDR, June 2017.

XIII. Snehal S. Dahikar, "An Artificial Neural Network Approach for Agriculture Crop Prediction Based on Various Parameters", IJARECS, Volume 4, Issue 1, January 2015.

XIV. Gang Liu, Xuehong Yang and Minzan Li, "An Artificial Neural Network Model for Crop Responding to Soil Parameters", Springer 2005.

XV. P. Priya, U. Muthaiah and M. Balamurugan, "Predicting Yield of The Crop Using Machine Learning Algorithm", IJESRT, April 2018.
XVI. Georg Ruß, "Data Mining of Agricultural Yield Data: A Comparison of Regression Models", Springer 2009.
XVII. SunitaBeniwal, Jitender Arora, "Classification and Feature Selection Techniques in Data Mining", IJERT, August 2012.
XVIII. Ms. Sonali. B. Maind and Ms. Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network", IJRITCC, January 2014.
XIX. K.Bharatha Krishna and S.S.Suganva, "Application of Data Mining in Agriculture", IJRCAR, August 2017.
XX. Dai, Qing-yun& Zhang, Chun-ping & Wu, Hao. Research of Decision Tree Classification Algorithm in Data Mining. International Journal of Database Theory and Application, 2016.
XXI. Bhumika Gupta, Aditya Rawat, Akshay Jain, Arpit Arora and Naresh Dhami. Analysis of Various Decision Tree Algorithms for Classification in Data Mining. International Journal of Computer Applications 163(8):15-19, April 2017.
XXII. Priyanka Gaur, "Neural Networks in Data Mining", International Journal of Electronics and Computer Science Engineering, ISSN- 2277-1956.
XXIII. Dr.Yashpal Singh, Alok Singh Chauhan, "Neural Networks in Data mining", Journal of Theoretical and Applied Information Technology.
XXIV. Nayak, Janmenjoy&Naik, Bighnaraj&Behera, Dr. H. A comprehensive survey on support vector machine in data mining tasks: Applications & challenges, 2015.
XXV. Yujun Yang, Jianping Li and Yimei Yang, "The research of the fast SVM classifier method," *12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2015.
XXVI. Bharati, M &, Ramageri., DATA MINING TECHNIQUES AND APPLICATIONS. Indian Journal of Computer Science and Engineering, 2010.
XXVII. Han, J., &Kamber, M., Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2001.
XXVIII. Lei Shi, QiquoDuan, Xinming Ma and Mei Weng, "The Research of Support Vector Machine in Africultural Data Classification", IFIP, 2012.
XXIX. Sharma, Gaurav. (2019). Data Analytics Model for Optimal Crop Management. International Journal for Research in Applied Science and Engineering Technology. 7. 690-693. 10.22214/ijraset.2019.6119.

