INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

AI ALGORITHMS USED IN THE ESTIMATION AND ANTICIPATION OF CROP PRODUCTION: A SURVEY

[1]Mrs. P. ANBARASI MCA., MPhil., **[2]Dr. T. PRABHU** MCA., MPhil., ME., PhD

[1]Assistant professor and head, department of computer science, sri muthukumaran arts and science college, chikkarayapuram, near mangadu, chennai-69.

[2]Associate Professor Cum Deputy HOD, Department of Computer Applications, Dr. MGR Educational & Research Institute, maduravoyal, chennai-600095.

[1]Mail to:anbarasip29@gmail.com, [2]Mail to:drprabhut22@gmail.com

Abstract: - This article is a survey study of various techniques and the models used in the forecasting of the plant produce in the agricultural sector. Agro based sector is a huge domain with the large economic status of any country. This paper is aimed to focus on the spatial data mining algorithm based on the deep learning models to anticipate the crop growth through the estimation of various factors including climatic changes and soil attributes. Data mining is a novel technique in the artificial intelligence model which is used to design a pattern based on the big data set and inadequate data sets. Hence these techniques are used to forecast the yield as the data obtained in the agriculture domain are always non-linear and unstable.

Keywords: Data mining algorithm, KNN model, Deep learning algorithms, Spatial Data mining framework, crop prediction

1. INTRODUCTION

India being the second largest producer of the agro based industries; agriculture plays an essential role in the economic growth of the country. This crop production business is based on many attributes like climatic and economic factors like temperature, moisture, weather, rainfall, manpower, fertilizers, pests and insects, pesticides, weeds and many more. Hence the scientist and researchers based on the agro sector focus more on the development of effective methodology or algorithm to anticipate and enhance the production. Most of the studies and research focus on the development of an Al triggered biological mechanism to estimate and develop the yield. By

investigating the soil and climatic condition at a particular area, prediction and estimation of the yield based on the analyzed factors are done. This helps the farmer to select appropriate crops suitable to the soil and prevailing weather conditions.

Data analysis is a series of processes of investigating, clarifying and designing data to devise innovative algorithms and methodologies. In other words, this process involves continuous steps of data evaluation, abstraction and prediction of a novel pattern from a large data set. Thus, the big data is processed into useful information. With the advancement in the neural networks and data processing algorithms , Data analysis is applied in the agro based industries to forecast the weather changes and the corresponding productivity of the crops India being second largest country in terms of population, with the grow h in the population demand and supply for the plant-based products also increases exponentially[5] . Using the advanced techniques many hybrid plant based products are produced in the market [6]. These exploitations make the land infertile and cause harmful effects to the people and ecosystem as well [7]. To overcome all these issues and challenges Researchers are now focusing on anticipating the physical and climatic attributes of the soil and the geographical location to estimate the production of the crops in a natural way thereby producing healthy products[8].

Generally, two different models were used for the prediction of the crop product namely simulation methodology and regression methodology [9]. Simulation methodology is purely an analytical framework which distinguishes and symbolizes the attributes which are deeply related to the data set involved in the previous researches and studies. This framework involves the simulation of different dataset under different conditions to design and scrutinize the framework. But the dataset related to agriculture is always inadequate and fragmentary. Due to these limitations, the regression model is mostly preferred over the simulation model for the anticipation of the crop produce. Also, the recent studies are based on the statistical, multivariate regression algorithms.

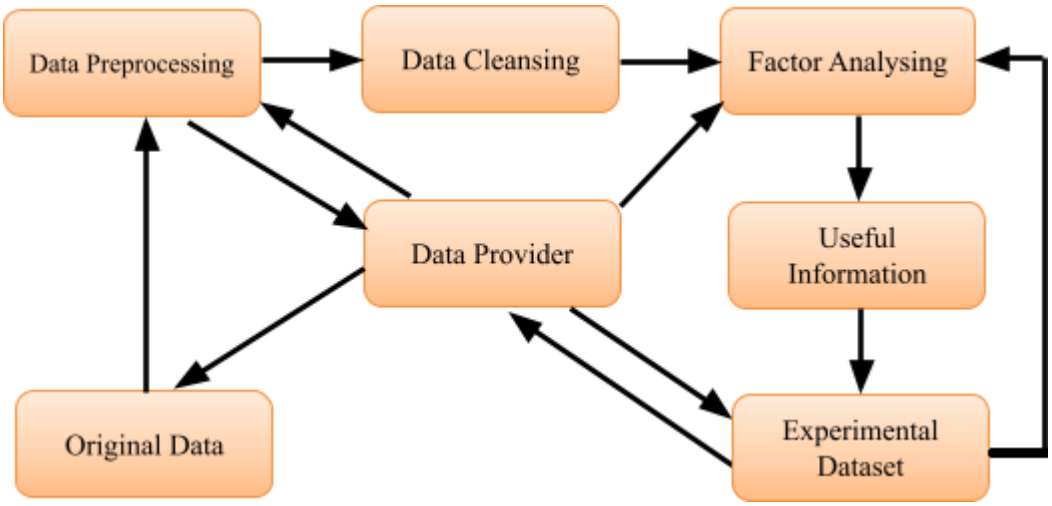


Fig.1.Data Analysis process

Data mining techniques are considered as the most important tools for the data set analysis related to the agro based industries. These techniques involve the computation and identification of the novel patterns to analyze the big data set. The analysis procedures involve different techniques like Al, deep learning algorithms, mathematical regression models, machine learning algorithms, etc. as shown in the Fig.1. Farming being the primary source of finds produce and secondary source of employment in India, in order to improvise the farming and produce more yield many new models depending upon the updated technologies were used. Due to the lack of the knowledge about crop selection according to the geographical area and the corresponding weather conditions people face difficulties in farming and suffer from loss. Hence this article is framed to be a survey of the previous researches and studies to anticipate and forecast the weather conditions and predict the corresponding crop yields.

2. LITERATURE REVIEW

Study of the previous works and research plays an important role in any research. This helps us to identify the challenges faced by the researchers and the corresponding solutions proposed by the scholars to fix those issues. Also helps us to identify the scope for the future research and development. This section focuses on the previous

studies carried out in Agro based industries for the prediction of the climatic conditions and the corresponding yields.

Ponce et al [11] implied the developed technique for his prediction algorithm which is a combination based on decision tree and data entropy modelling. The results were simulated to obtain the graphical curves. Later Hybrid algorithms incorporating SVR algorithms into data entropy methods were developed by Lee at al [12]. Other Al techniques like regression models and fuzzy entropies for the reliable classification were suggested by Manjunatha et al

Roberto et al suggested a procedural technique where the interested attributes are estimated statistically to develop a stable and generalized numerical algorithm. But yet the algorithm is not generalized as the database access was given to a single user

Narsi Reddy et al tried to estimate the crop produced through Lilli for method with the implied qualitative normalization. But the null hypothesis result of this method failed to normalize the crop production.

Lansigan et al] used the attributes of the prevailing weather in the geographical area under study to anticipate the changes in those attributes. The sequenced data set was generated from the specially designed weather generator. Multiple regression analysis proved that the changes in the weather attributes affect the yield respectively.

Supervised and unsupervised algorithms were involved in the forecasting of the product. Clustering technology was considered as an unsupervised model and was applied for many different applications in the agro based industries. There are different types of clustering techniques like partitioned, hierarchical, Grid and density based model, etc.

In the partitioned clustering model, means and the medians are used to assign the cluster number to each cluster and is reassigned on qualitative iteration. Whereas in the hierarchical model the assignment was done based on the decision tree algorithms and the lowest level cluster is identified using the data points [17].

Density based unsupervised learning algorithms consider the neighbouring clusters and the Euclidean distance between them is calculated as the density and it has a threshold point. Each neighbour cluster is defined to have minimum points meeting the threshold.

Soil attributes like temperature, moisture, dampness, etc are studied using the data mining methodologies. Alberto et al made a detailed analysis to estimate the efficiency of the Al algorithms to support the regression model using multiple variables.

Smart farming application was developed by Gupta et al using internet of things and deep learning algorithms. Sensors were used to capture the soil attributes and processed as the raw data. This data set was sent to the cloud storage where the learning algorithms were used to process and analyze the output.

Altylar et al used a three-dimensional model to predict the crop produced using statistical tools. A GPS system was used to locate the geographical coordinates and the corresponding climatic conditions were obtained from the other weather generator applications. Root means square parameter was calculated for the obtained and the original data to check the correctness of the algorithm.

Kholsa et al proxied a two state model where the rainfall forecasting was done using Modular artificial intelligence network and crop productivity was anticipated using a support vector regression algorithm. Using the old data set future dataset was forecasted and generated.

3. DATA MINING TECHNIQUES

The general division of the data mining methodologies includes data clustering and data classification methodologies. Generally data classification is considered as the supervised learning models where the unknown set of samples are analyzed and distributed based on the given data. These sample sets are considered as the training data set and are used as the base reference model for all fun her classification. Upon vector models and the deep learning algorithms are based on the classification model which analyzes the unknown data set[23]. When the classification is based on the k-nearest neighbor, there is no available knowledge material as they use the training set. The important consideration in this method is that neighboring clusters or nodes should undergo similar classification. Thus, it has no specific decision parameter or a protocol for the classification.

When there are no available knowledge materials or the training set classification is very difficult without any leads. Here comes a clustering model for the rescue. These models are for the division of the available data set into cluster sets. The most popular technique is the k-means algorithm. In this algorithm the numbers of the clusters are decided by the k value. When a part of information is given to the algorithm the mean value that is the central cluster is calculated and the remaining clusters are designed. The major drawback of this method is the value identification of the K. The increased computational complexity of this algorithm makes it expensive ten. Other data mining techniques like component analysis, multivariate regression model, etc are also implied in the agricultural sector.

Deep learning algorithms and the neural models are developed based on the activity of the human neural system. Similarly, the artificial neural network analyzes the large number of data sets and identifies the patterns of the cluster sets. They are capable of processing the incomplete and inadequate data sets and can find solutions for any type of problems and dynamic attributes. They are characterized to be nonlinear and multivariate. The additional benefits of this algorithm are that they are analytically strong and have a good training set. Hence, they can produce logical and efficient results through this data set].

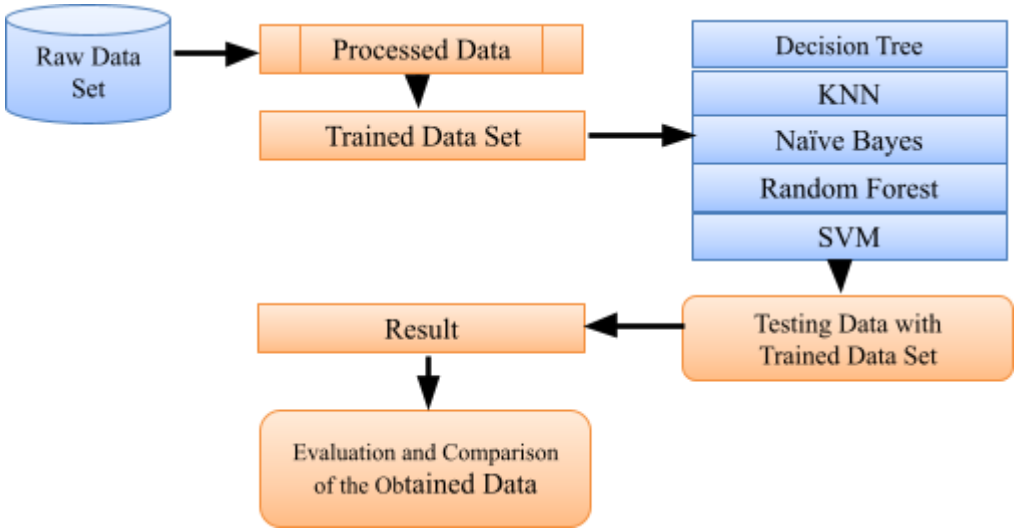


Fig.2. Generalized model of the forecasting framework

Decision tree based algorithm is applied in the classification of the data clusters. This technique is used in risk control and management applications. This algorithm comes under the supervised learning models which is generalized through the observations.

Another algorithm was designed based on the Bayes theorem which is known as the Naive Bayes model. This system model is totally dependent on the probability model. This is a supervised learning algorithmic model and has a partial training set to estimate the parameters which are required for the classification.

The generalized model of the forecasting framework is illustrated in Fig.2. The raw data are processed and cleaned to obtain the training set. This training set is then processed through any of the deep learning algorithm and the result is obtained. The obtained result set is then compared with the existing systems for evaluation.

4. METHODOLOGY

Data mining as a classification and clustering process holds a crucial role in the forecasting of the climatic condition and crop production in the agricultural sector. Different algorithms are used for the prediction of different parameters. The pollution and the atmospheric changes are anticipated by the k-means algorithmic model, the moisture and the precipitation in the air are estimated and predicted by the KNN algorithm. SVM algorithmic

models are used as the weather generator. Thus, these algorithms find their important role in the prediction of the soil parameters and the attributes [28].

In recent times with the enormous growth of science in the field of agriculture along with the plant produce huge data sets were also harvested. Though these data are very large to analyze they produce reliable predictions of productivity. Not only the dataset regarding weather change and climatic conditions are needed but also huge data regarding the soil classification, pest control, choice of fertilizer and its quantity are needed to predict the produce.

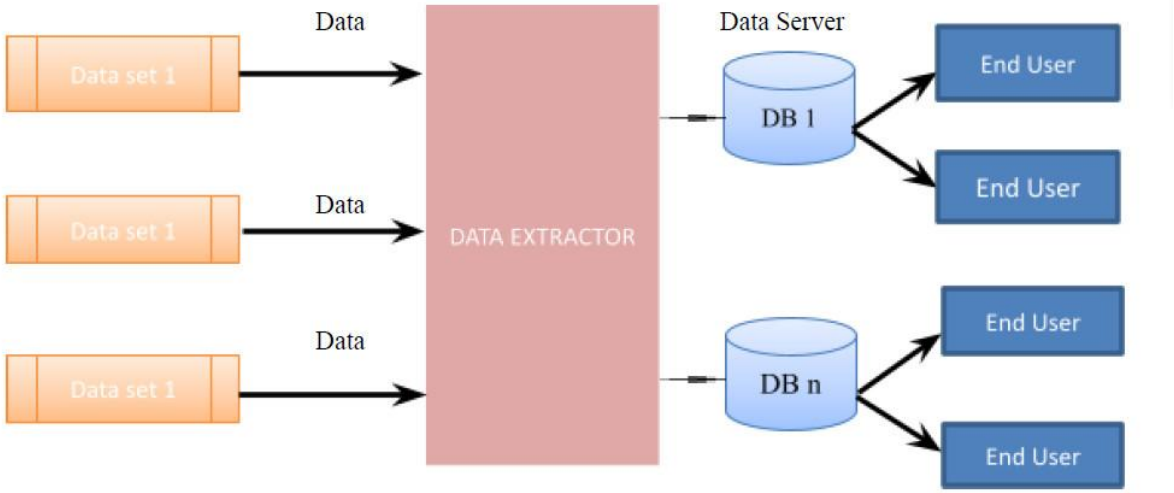


Fig.3.Framework of the Data Management System

Xu et al proposed an effective data management framework as illustrated in Fig.3. Here the input data is obtained from the system based on the customized rules which vary with the application of the system, then fed to the data extractor. Data extractor cleans the data and pre-process it based on the storage unit. Then it is distributed to the entitled end users through the data servers. There are many components involved in the data collection and data management. Repositories are responsible for the storage of all types of data collected from the system. It holds sole responsibility for data collection, storage, validation and distribution. Depending upon the application the data set is then sent to the corresponding collaborative extractor where the data is pre-processed based on the algorithm type. Then it is distributed to the destined users through a proper network. Generally the data collected for the agriculture purpose is always non stable and inadequate. Hence eta mining techniques are used to regulate the nonlinear data into atomic values. Spatial data mining techniques involve the use of data corresponding to the geographical coordinates and maps.

The ultimate goal of the spatial data mining techniques in the agro based industries is to anticipate and forecast the weather changes and the corresponding climatic condition which affect the growth of the crop and to estimate the plant production as an integrated function of the soil and weather analysis. There are many models and supporting equipment which are being used by the agro based researchers and the scientists used to validate the obtained data. The generalized prediction model used in the agriculture sector to predict the crop produce and the climatic changes are illustrated in the Fig.4.

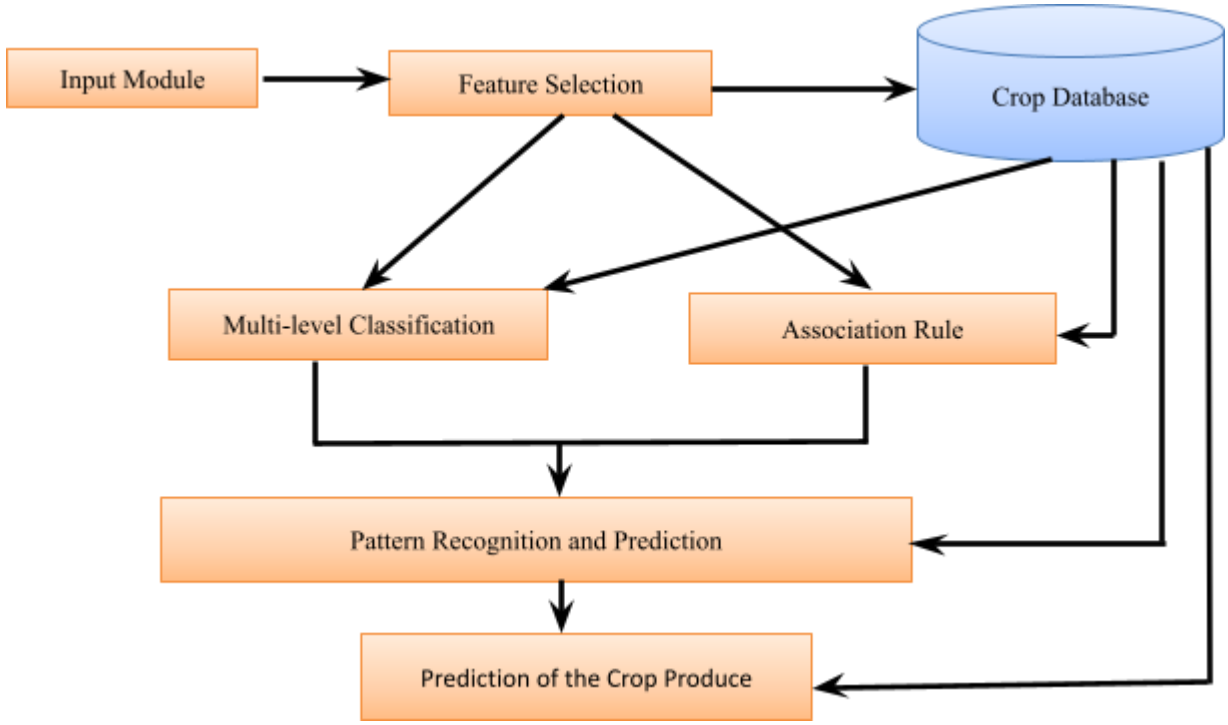
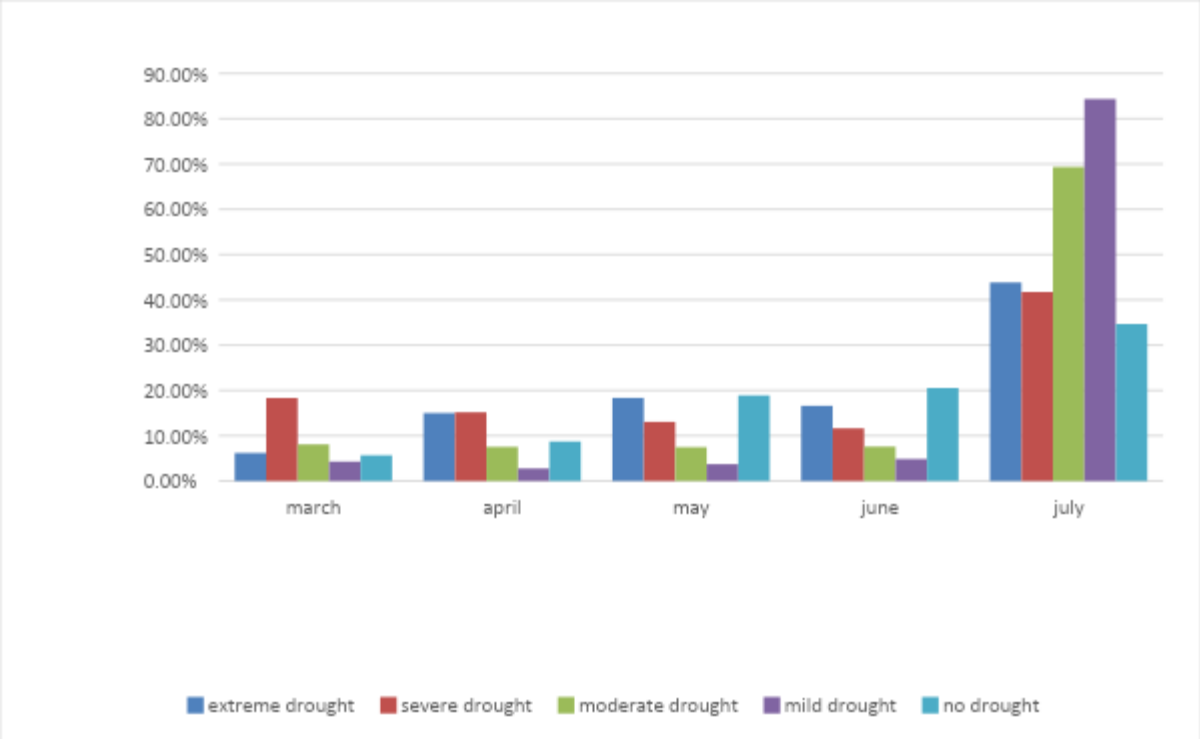


Fig.4.Generalized prediction architecture

The prediction framework has an input system where the raw data are taken from the field. This includes the type of the soil, prevailing climatic conditions, geographical location, economic status, type of the crop, etc. The data collector has equal responsibility to the analyzer to collect each and every fine detail. With the spatial data input the detail about the crop is then fed to the database. This database has a set of estimated data corresponding to the required attributes like amount of moisture, rainfall, irrigation, fertilizer needed for the crop depending upon the upcoming changes in the weather. Then the classification and the distribution algorithms are used for the clustering of the similar set. Then the anticipated result is applied to the crops to improvise the produce. Thus, the effective crop prediction algorithm improvises and augments the production with less expense and avoids the loss.

5. CONCLUSION

Thus, with the consideration of the previous studies and the researches the author proposes a nov'e1 model which includes the use of the hybrid algorithm where the spatial data mining along with the deep learning algorithms are used to predict the climatic conditions and other requirements for the crop growth. In this proposal, an effort has been made for mining the dataset using various spatial data mining techniques and algorithms. In order to perform spatial data mining, many decisions need to be made regarding the choice of data, tools, algorithm and methodology, according to research. Principal Component Analysis (PCA) has been used in feature selection of time series data while Association Rules are used to identify the relationship between SPI and VCI, and occurrence of agricultural drought in the study. Classification techniques as Naive Bayes, Decision Tree, Random Forest, Neural.



Analyzing the extent of drought in the Tamil Nādu State of India Using Vegetation condition index and Standardized Precipitation Index

Network and Support Vector machines are applied on different combinations of SPI, VCl, and rainfall feature set to compare out the best satisfaction model in agricultural monitoring. Regression techniques such as Linear Regression, Decision Tree, Random Forest, Neural Network and Support Vector Machine are also applied on different combinations of SPI, VCI, and rainfall feature sets. Time-series Analysis is used to learn patterns/trends from largely available remote sensing dataset for prediction of six years agricultural condition of study area.

REFERENCES

1) Sulaiman, M. A. (2020) "Evaluating Data Mining Classification Methods Performance in the Internet of Things Applications". Journal of Soft Computing and Data Mining, I(2), 11 - 25.

2) Nathgosavi, V ., &Pati1, S. (202 I). "A Survey on Crop Yield Prediction using Machine Learning" (No. 5238).EasyChair.

3) Chicho, B. T., Abdulazeez, A. M., Zeebaree, D. Q., & Zebari, D. A. (2021). "Machine Learning Classifiers Based Classification for IRIS Recognition". Qubahan Academic Journal, I (2), 106- 1 1 8.

4) Mahmood, M. R., &Abdu1azeez, A. M. (2019, April). "A different model for hand gesture recognition with a novel line feature extraction. In 2019 1nternational Conference on Advanced Science and Engineering" (lC OASE) (pp. 52-57). IEEE.

5) Paul, M ., Vishwakarma, S. K., &Verma, A. (2015, December). "Analysis of soil behaviour and prediction of crop yield using data mining approach". In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 766- 77 I). IEEE.

6) Patel, H., & Patel, D. (2014). "A brief survey of data mining techniques applied to agricultural data". International Journal of Computer Applications, 95(9).

7) Ramesh, D., &Vardhan, B. V. (2015). "Analysis of crop yield prediction using data mining techniques". International Journal of research in engineering and technology, 4(1), 47-473.

8) Van Klompenburg, T., Kassahun, A., & Catal, C. (2020) . "Crop yield prediction using machine learning: A systematic literature review". Computers and Electronics in Agriculture, 177, 105709.

9) Winkler, K., Gessner, U., and Hochschild, V. t20l7). Identifying Droughts Affecting Agriculture in

10) Africa Based on Remote Sensing Time-series between 2000-20 1 6: Rainfall Anomalies and Vegetation Condition in the Context of ENSO. Remote Sensing, 9(8), 83 I

11) Maria Rossana C. de Leon, Eugene Rex L. Jalaok, "A Prediction Model Framework for Crop Yield Prediction", Asia Pacific 1ndustrial Engineering and Management System, 2020.

12) Ponce-Gues'ara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.Dominguez Limaico,

13) 13. H. M ., Suarez-Zambrano, L.E., Rosero-Montalvo, P.D., Alxarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data techniques (First trial). 20 17 1EEE Second Ecuador Technical Chapters Meeting (ETCM).

14) Jheng, T.-Z., Li, T.- H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output.20 1 8 27th Wireless and Optical Communication Conference (WoCc).

15) Manjunatha, M., Parkavi, A. (2018). Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. 201 8 International Conference on Current Trends towards Con-verging Technologies (1CCTCT).

16) Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B, "Generalized Software Tools for Crop Area Estimates and Yield Forecast "2010.

17) Maria Rossana C. de Leon, Eugene Rex L. Jalaok, "A Prediction Model Framework for Crop Yield Prediction", Asia Pacific Industrial Engineering and Management System, 2013.

18) Lansigan, et.al., "Analysis of Climatic Risk and Coping Strategies in Two Major Corn Growing Areas in the Philippines" 2010.

19) Han I , Kamber M. Data mining: concepts and techniques. Massachusetts: Morgan

20) Kaufmann Publishers; 200 I.

21) Ester M, Kriegel HP, Sander I, Xu X. A densit y-based algorithm for discovering clusters in large spatial databases with noise. In: Paper presented at international conference on knowledge discovery and data mining. 1996.

22) Gonzalez-Sanchez Alberto, Frausto-Solis Juan, Ojeda-Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction. Span I Agric Res. 20 14: I 2t2):313-28.

23) Gupta, G., Setia, R., Meena, A., &Jaint, B. t2020, June). Environment Monitoring System for Agricultural Application using IOT and Predicting Crop Yield using Various Data Mining Techniques. 1n 2020 5th International Conference on Communication and Electronics Systems (ICC ES) (pp. 10 I 9- 1025). IEEE

24) Terliksiz, A. S., &Altylar, D. T. (2019, July). Use of deep neural networks for crop yield prediction: A case study of soybean yield in Lauderdale County, Alabama, USA. In 20 19 8th International Conference on Agro-Geo-informatics (Agro-Geo-informatics) (pp. I -4). IEEE.

25) Khosla, E., Dharavath, R., &Priya, R. (2019). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environment, Development and Sustainability, I -22.

26) Suleiman, M . A. (2020). "Evaluating Data Mining Classification Methods Performance in the Internet of Things Applications". Journal of Soft Computing and Data Mining, I (2), I -25.

27) Paul, M., Vishwakarma, S. K., &Verma, A. (2015, December). "Analysis of soil behaviour and prediction of crop yield using data mining approach". In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 766- 77 I). IEEE.

28) Ta her, K. 1., Abdulazeez, A. M., & Zebari, D. A. t202l) "Data Mining Classification Algorithms for Analyzing Soil Data". Asian Journal of Research in Computer Science, I 7-28.

29) Bhosale, S. V., Thombare, R. A., Dhemey, P. G., &Chaudhari, A. N. (2018, August) "Crop yield prediction using data analytics and hybrid approach" In 20 18 Fourth International Conference on Computing Communication Control and Automation (ICC UBEA) (pp. I -5) IEEE.

30) Yahia, H. S., & Abdulazeez, A. M. (202 I)"Medical Text Classification Based on Convolutional Neural Network: A Review" International Journal of Science and Business, 5(3), 27-4 I .

31) Xu, S. and Zhang, W. (2004). "PBKM: A secure Knowledge management Framework". NSF/NSA/AFRL Workshop on Secure Knowledge Management' 04.

32) Gandhi, N., and Armstrong, L. J. (2016, December). A review of the application of data mining techniques for decision making in agriculture. In Contemporary Computing and Informatics (lC3l), 20 1 6 2nd International Conference on (pp. I -6) IEEE.