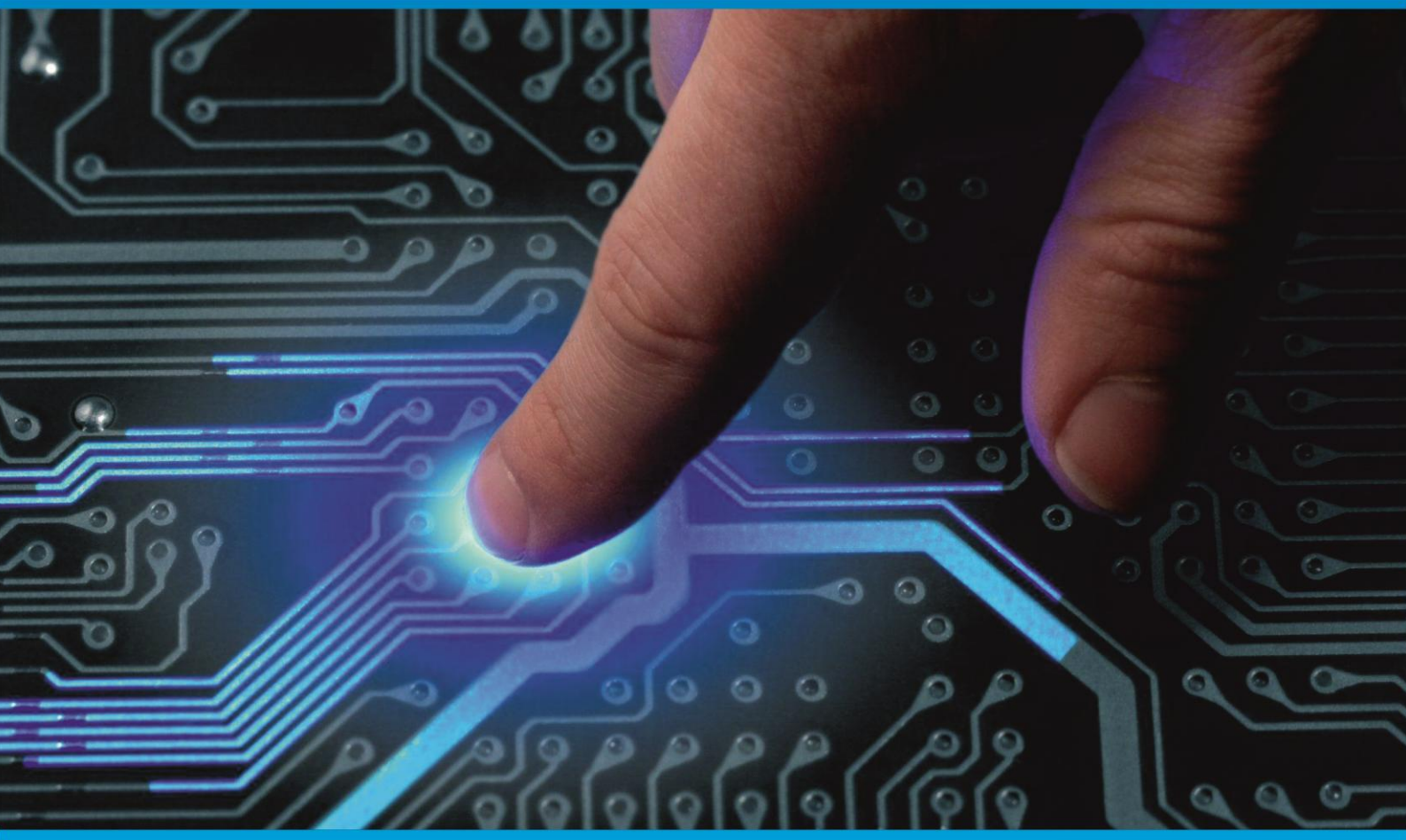
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

Agriculture Analysis Using Data Mining & Machine Learning Techniques: A Survey

Simran Dhar[1], Ankita Balasaheb Tate[2],Vaibhavi Sunil Vanarase[3], Vrushali Lalaso Mane[4], Prof.Harsha Bhute[5]

B.E (Computer Engineering), SAE, Kondhwa, Pune, India[1,2,3,4]

Professor, Department of Computer Engineering, SAE, Kondhwa, Pune, India[5]

ABSTRACT: Agricultural statistics and forecast is an important resource that the government has not explored commensurate to its impact. The aim of our proposed work is to make this process computerized by implementing principles of data mining and analytics. More specifically, our project aims at targeting the social issue of drought, analyzing data based on crop produce, amount of rainfall, agricultural inputs, irrigation, and similar factors for every crop in the state of Maharashtra.

Based on the extensive research carried out through this project, effective countermeasures and suggestions will be given, which if implemented expeditiously, can help tackling the problem of drought in state. Data can be mined and analyzed to find various trends and relations, such as – contrast between total irrigation area and type of crop; total principal and non-principal crop amount versus district-wise rainfall etc.

KEYWORDS: information technology, data mining, cluster, agriculture, regression, functional requirements; non-functional requirements; text normalization; feature extraction; machine learning; support vector machines

I. INTRODUCTION

Raw data in agriculture is very diverse. It is necessary to collect and store it in an organized form and integrate it for the creation of agricultural information system. Therefore, there is a need of utilization of information and communications technologies, which will enable the extraction of significant data from agriculture in an effort to obtain knowledge and trends [2]. It will also help in the elimination of manual tasks. Easier data extraction directly from electronic sources and its transfer to secure electronic system of documentation will reduce the production cost, higher yield and higher market price [3].

Data mining technique will provide information about crops and enable agricultural enterprises to predict trends about customer conditions or their behaviour. It analyzes the data from different perspectives and helps in finding relationships in seemingly unrelated data [4]. The computational needs of agriculture data and how data mining techniques can be used as a tool for knowledge management in agriculture should be considered by researchers. Data warehouses can be prepared to hold agriculture data, which makes transaction management, information retrieval and data analysis much easier [5].

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 [4, 6]. Most of the existing systems are hardware based which makes them expensive and difficult to maintain [8]. Also they lack to give accurate results. Some systems suggest crop sequence depending on yield rate and market price. 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 [9]. 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 [11].

The purpose of this review is to introduce briefly the techniques of Data Mining and to outline its use in agriculture and allied sectors. Being much more superior to the conventional data analysis techniques used in agricultural research, data mining can open a new opportunity to explore and increase the development in agriculture [12]. Decision support systems are one of the major tools for IT infrastructures in order to generate wealth from extracted information, further converted into results. Collecting, maintaining, and analyzing large amounts of data; however many of the tasks that involve significant technical challenges, expenses, and organizational commitment [9]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

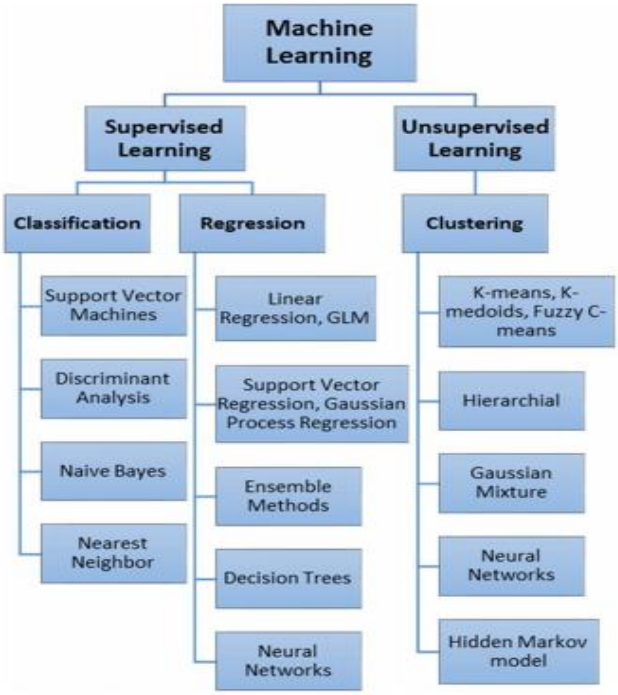


Figure1.1: Classification of Machine Learning

Agriculture plays a critical role in the global economy. Pressure on the agricultural system will increase with the continuing expansion of the human population. Agri-technology and precision farming, now also termed digital agriculture, have arisen as new scientific fields that use data intense approaches to drive agricultural productivity while minimizing its environmental impact. The data generated in modern agricultural operations is provided by a variety of different sensors that enable a better understanding of the operational environment (an interaction of dynamic crop, soil, and weather conditions) and the operation itself (machinery data), leading to more accurate and faster decision making [14].

Machine learning (ML) has emerged together with big data technologies and high-performance computing to create new opportunities to unravel, quantifies, and understands data intensive processes in agricultural operational environments [16].

Much existing work has been devoted to understanding, decomposing, managing, formalizing and reasoning over qualities of typical non-ML software. Such qualities are often included as part of non-functional requirements. These include well-studied NFRs such as performance, reliability, maintainability, and usability, but also security, privacy, and customer satisfaction.

ML encompasses over a dozen algorithm types: Regression, Bayesian, Instance-based, Deep Learning, and Neural Networks), with many more specific algorithms (e.g., Logistic Regression, Linear Regression, and Naıve Bayes. Most work on ML topics provides examples and algorithm details, including performance results, but do not focus on a wide range of NFRs or quality aspects [14].

The proposed system applies machine learning and prediction algorithm like Multiple Linear Regression to identify the pattern among data and then process it as per input conditions. This in turn will propose the best feasible crops according to given environmental conditions [18].Thus, this system will only require the location of the user and it will suggest number of profitable crops providing a choice directly to the farmer about which crop to cultivate. As past year production is also taken into account, the prediction will be more accurate.

1.1 HISTORY

Machine learning (ML) makes machines independent and self-learning component. Researchers applying machine learning algorithms to solve various real word problems in various domains. Nowadays agriculture affects by various

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

factors such as global warming, climatic changes, lack of manpower, etc. To help the farmers from the above factors and increase agriculture production, recently many machine learning techniques are utilized in the agricultural field [23]. In this paper, we studied different applications of machine learning techniques in the agriculture domain. We classified applications of machine learning algorithms in agriculture by four categories namely, machine learning in plant monitoring, machine learning in soil analysis, machine learning in detection (or) prediction process in agriculture, machine learning in animal monitoring. We also analyzed the important features of machine learning applications in agriculture [19].

Maharashtra is divided into five geographical regions, comprising six administrative divisions—Konkan, Pune, Nashik, Marathwada (Aurangabad) and Vidarbha (Amravati and Nagpur). The Vidarbha region reported the most farmer suicides, 1,541, in 2015. Nagpur (362) and Amravati (1,179) witnessed the maximum farmer suicides in this region [26]. Vidarbha was followed by Aurangabad (1,130) that forms the Marathwada region. As many as 3,228 farmers committed suicide in Maharashtra in 2015, the highest since 2001, according to data tabled in the RajyaSabha on March 4, 2016–that is almost nine farmers every day. Vidarbha and Marathwada, with 5.7 million farmers, accounted for 83% of all farmer suicides in Maharashtra in 2015 [28].

The farmer suicides, which have remained unstoppable for past few years in eight districts of Marathwada, have crossed the staggering 400-mark in just over a four-month period in 2016. Compared to 2015, as many as 92 more farmers have embraced deaths in the first four and half months of 2016, highlighting the failure of the government schemes launched in August to curb the spate of suicides.

1.2 TECHNIQUE OF SEMANTIC ANALYSIS

Data mining has been classified into two types such that one is descriptive another one is predictive. Descriptive data mining considers the existing data, that is raw data and then make it summarized. The descriptive mining represents the characteristics of the past events and allows us to learn how they influence the future. The foundation of predictive mining depends on probabilities, it is used to predict future based on the values considered from known results. Forecasting involves using the variables or field in the database to estimate anonymous results.
to generate useful information from it and then concise this information [16]. Data mining software is considered as an analytical tool for analyzing data in many different dimensions, perspectives, or angles, and categorizing the analysis, and summarizing the relationships among these categories are identified. There are four types of data relationships:
1) **Classes Data is stored in the form of classes**: A data class consists of a data fields and some basic methods to access those data fields. The fields in a data class have similar characteristics.
2) **Clusters:**A data cluster is a labeled portion of data containing similar items. It also refers as a method of partitioning a set of data into a set of sub-classes each of which has a specific meaning, called clusters. It helps users to understand the accepted alignment or structure of the data in set. Clustering comes under unsupervised learning.
3) **Association:** It is one of the data mining techniques that determine the likelihood of the co-occurrence of data items in a data set [27]. The relationships between co-occurring data items are expressed as association rules.
4) **Sequential patterns:** It is used to finding statistically appropriate patterns between data elements where these data element values are distribute in a sequence.

Data mining provides the bridge between the different transaction and analytical systems. Using open-ended user queries data mining software analyzes relationships and patterns in stored data. Some of the analytical software include; machine learning, statistical, and neural networks [27].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

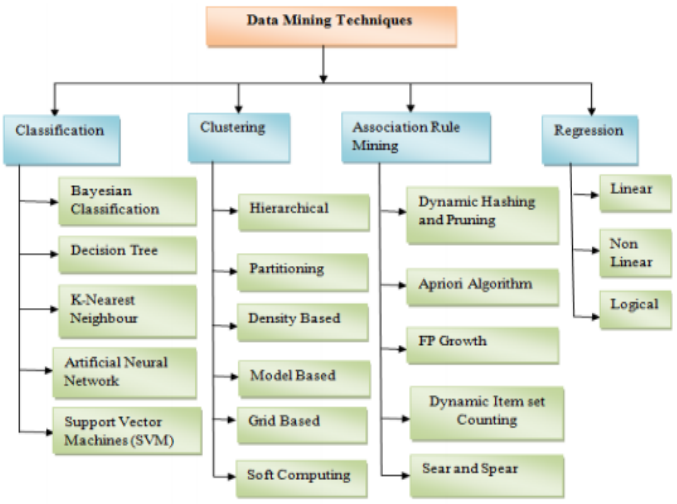


Figure1.2: Classification Data Mining Techniques [27]

Classification: Any data mining or Machine learning algorithms can follow three learning approaches: that are, supervised learning, unsupervised learning, and Semi-supervised learning. The classification algorithms are falling into the category of supervised learning. The classification is a two-step process. In the first step, by analyzing the data tuples described by their attributes within the predetermined data set we build a model. In this case class labels are predetermined. In the second step we estimate the analytical accuracy of the built-in model used for classification [22].

Regression is a data mining function which is used to predict a numeric or continues value. Both of the techniques classification and regression exhibits the similar functionality (predictive analysis), but classification classify the data into discrete sets. The regression is used to establish the relationship between dependent variables versus independent variables. It maps a data item to a real-valued prediction variable. [22] Regression also represents the variations occurred in one variable in accordance with variations in another variable. One important issue regarding regression is it describes the relationship among the variables in detail than correlation – it represents the strength of the relation between the variables. Regression tasks are often treated as classification tasks with quantitative class labels.

Association Rules: When we use association rules, we should consider two important basic rule measures such as Support and Confidence. These two measures have some threshold value to mine interesting patterns from huge collection of data. Naturally the support is the realistic value and confidence is an estimated value. The rules which are mined are not satisfying the threshold value is called uninteresting [21, 22]. To measure the interestingness of an association rule, the rule must be simple, certainty, and utility.

Clustering is an unsupervised learning technique, to find the likelihood between unlabeled data elements into groups. In clustering, the data elements with the groups are more relevant to each other and differ with data elements in other groups. Basically, clustering is two kinds; one is conventional clustering and other is conceptual clustering. The following are some characteristics should have a good clustering algorithm [21,22].

a) A cluster algorithm should easily interpretable and usable.
b) It should have high dimensionality.
c) It is able to partition constrained oriented data.
d) It should be scalable.
e) It is able to produce uninformed shaped clusters.
f) It is capable to deal the data which has different kinds of attributes.
g) It should take less domain knowledge.
h) It should handle noise data.

The clustering is used to determine a new set of groups, these groups are of significance in themselves, and their assessment is intrinsic.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

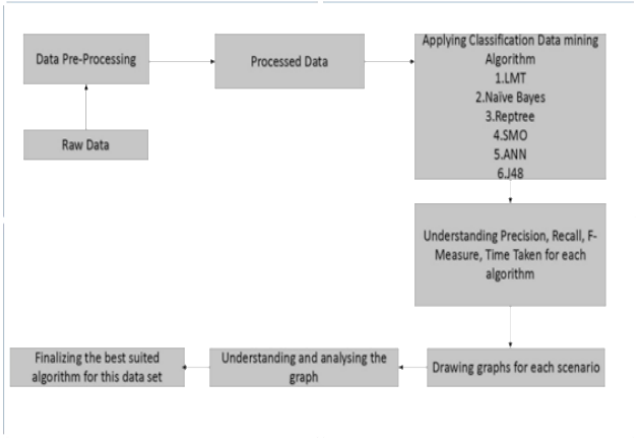
Sequence analysis algorithms: Today the most challenging area in machine learning is to extracting hidden patterns from the sequential data. This means, patterns are shared among data objects. We can perform sequential mining either on single data sequences or multiple data sequences. Sequential mining is more relevant to the temporal databases. The main limitation of sequential patterns is, there is no probability assessment followed by a pattern [24, 22].

Machine Learning Techniques:

Database/Crop Datasets:
The data that populates this database includes the plant growth parameters that were used to form the individual decision trees in the random forest. Such data include irrigation, spacing, nutrient requirements, location, temperature, and other related factors that originated from several trusted databases. &is plant growth condition database is designed to help decision making with the machine learning algorithm [27].

Method:

In this work, machine learning applied what had been used to set parameters and embedded it into a dataset on a mobile application. &e machine learning algorithm was designed to maximize land proportion. &e dataset contains parameters of some inputs that are critical for plant growth. &e machine learning algorithm defines the relationship between these input parameters and certain internally stored prediction parameters and provides a solution for the output. &e values in the database have been converted to a range system of 0 to 1; the need for conversion to the same range is due to data incoherence; data was derived from different sources and was therefore inconsistent, thus requiring a specific conversion



On the processed dataset data mining algorithms are applied and accuracy of the dataset is calculated using F-measure. Comparative study is done to analyze and understand various other algorithms. And finally best algorithm is found based on accuracy Weka tool is used for analysis of dataset [7].

II.RELATED WORK

In [1] paper investigates the author discuss about Zambia's Crop Forecast Survey (CFS) and its limitations. In this proposed method at the end of season harvest the data is collected. The data about crop yield and area of production is collected from the actual harvest from households in real time. Using this data a forecast model for yield is developed using Extreme Gradient Boost. To test the accuracy values RMSE is used. From results it is found that CFS accuracy is less compared to gradient boost.

In [2] this paper, the corn yield prediction is made using deep neural network model from environment and genotype data. The dataset consist of 148,452 samples for training. Yield performance is calculated using yield and check yield using deep neural networks. Neural network in his paper is also used for weather prediction. The model accuracy is tested using RMSE 11 % error for single hidden layer with 300 hidden neurons.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

In [3] the paper compares the accuracy in forecasting the price of agricultural products using Auto ARIMAmodel and back propagation (BP) network method. A web crawler technology is also used to get prediction of price for the crop commodities. . The data is collected about the cucumber crop. The data obtained is normalized in order to scale it down to (0, 1) range. The dataset is consist of linear data on which ARIMA algorithm is applied for forecasting. The ARIMA model does not consider the rapid changes in the prize of the crop that arises due to instability in seasons. In order to consider such changes BP network model is applied to the data. BP network model contains set of input, hidden and output layers. The number of output layers is increased in order to check whether the price prediction for long-term is accurate or not.RNN model used to process sequential data is applied on the normalized data to obtain the price. The results show that the foresting for monthly, weekly and daily price of crop ARIMA does well if it is for short term but accuracy is not proper for long term. So, BP neural network is used for such issues which give high accuracy.

Difference between Existing Techniques [29,28]

S/N	Author(S)	Year	Problem	Method	Contribution
1	Priya et al. [3]	2018	This work was concerned with the use of the random forest algorithm to generate predictions for crop yield and improvement.	The random forest algorithm was used for yield production using a dataset with four features or parameters. A training set as used to train the algorithm rules which were then applied to the remaining datasets.	The results showed that we can attain an accurate crop yield prediction using the random forest algorithm. Random forest algorithm achieves a largest number of crop yield models with lowest models. It is suitable for massive crop yield prediction in agricultural planning.
2	Jeong et. al. [4]	2016	This work aimed at examining the performance efficiency of the random forest algorithm in crop yield prediction for the wheat crop, potato crop, and maize crop.	The random forest algorithm was used to train the datasets, and the same datasets were applied to an MLR model as a benchmark for the random forest algorithm.	The work showed that the random forest algorithm is far more effective in crop yield prediction.
3	Liakos et. al. [2]	2018	This work involved a research into the use of machine learning agricultural production systems.	This work applied artificial neural networks.	This work showed that machine learning models have been used in several agriculture-related areas. Mainly in crop production and aiding management decision making processes.
4	Ming et. al. [5]	2016	This work involved classification of land cover based on image and remote sensing.	Random forest machine learning algorithm was used in the classification of image data.	Random forest is an efficient classification algorithm and performs effectively without using special selected features.
5	Nitze et al. [6]	2012	This work compared the effectiveness of several machine learning algorithms: support vector machine, artificial neural networks, and random forest.	several classifiers, Naïve Bayes for ML, random forest (RF), multilayer perceptron in case of ANN, and LibSVM for support vector machine, were used in this work for the classification of crops.	Even though classification results depended strongly on the number of images used, the SVM classifiers performed much better than the RF and ANN in most of the cases.
6	Chen and Cournede [7]	2017	This work focused on finding the most efficient way to predict the yield of corn based on meteorological records.	This work studied a new methodology named multiple scenarios parameter estimation and used the CORNFLO model.	Random forest regression was shown to be the most efficient for crop yield prediction.
7	Mitra et al. [8]	2017	This work focused on simulating and predicting crop yield for effective crop management and adequate results.	A three-layered artificial neural network (ANN) and R language were used in this work for prediction and simulation of crop yield.	The artificial neural network was effective for simulation and prediction.

III.METHODOLOGY

The design of the flow of the system is as follows:

- Raw data is collected.
- Data is segregated and standardized.
- It is fed to the database; we will be using Excel as ours.
- Its sent for pre-processing where the data is fed next to train the system.
- The algorithms are used to predict results.
- Python scripts and libraries will be used to then visualize the obtained results for a better understanding of trends.

IV.PROBLEM STATEMENT

Crop Prediction involves predicting yield of the crop from available historical available data like weather parameter, soil parameter and historic crop yield. This project focus on predicting the yield of the crop based on the existing data by using Random Forest algorithm. Real data of used for building the models and the models were tested with samples. The prediction will helps to the farmer to predict the yield of the crop before cultivating onto the agriculture field. To predict the crop yield in future accurately Random Forest, a most powerful and popular supervised machine learning algorithm is used. And also solve their problem by Instructor.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

V.AIM &OBJECTIVES

"To design and identifying specific areas that are suitable for specific crops, so as to avoid deficit of crops".

Objective:

- To design an algorithm to predict crop.
- To apply a data mining algorithm.
- To analyze the agriculture data using data mining techniques

VI.LITERATURE REVIEW

Applications of Machine Learning Algorithms in Agriculture Article · May 2020, Machine learning(ML) makes machines independent and self-learning component. Researchers applying machine learning algorithms to solve various real word problems in various domains. Nowadays agriculture affects by various factors such as global warming, climatic changes, lack of manpower, etc. To help the farmers from the above factors and increase agriculture production, recently many machine learning techniques are utilized in the agricultural field. In this paper, we studied different applications of machine learning techniques in the agriculture domain. We classified applications of machine learning algorithms in agriculture by four categories namely, machine learning in plant monitoring, machine learning in soil analysis, machine learning in detection (or) prediction process in agriculture, machine learning in animal monitoring. We also analyzedthe important features of machine learning applications in agriculture.

Machine Learning in Agriculture: A Review Konstantinos G. Liakos, PatriziaBusato, DimitriosMoshou, Simon Pearson ID and DionysisBochtis, Machine learning has emerged with big data technologies and high-performance computing to create new opportunities for data intensive science in the multi-disciplinary agri-technologies domain. In this paper, we present a comprehensive review of research dedicated to applications of machine learning in agricultural production systems. The works analyzed were categorized in (a) crop management, including applications on yield prediction, disease detection, weed detection crop quality, and species recognition; (b) livestock management, including applications on animal welfare and livestock production; (c) water management; and (d) soil management.

Data mining Techniques for Predicting Crop Productivity – A review article S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, in this paper an attempt has been made to review the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such asID algorithms, the k-means, and the k nearest neighbor, artificial neural networks and support vector machines applied in the field of agriculture were presented. Data mining in application in agriculture is a relatively new approach for forecasting / predicting of agricultural crop/animal management. This article explores the applications of data mining techniques in the field of agriculture and allied sciences.

Crop Prediction System using Machine Learning Prof. D.S. Zingade ,OmkarBuchade ,Nilesh Mehta ,ShubhamGhodekar ,Chandan Mehta, India being an agricultural country, its economy predominantly depends on agriculture yield growth and allied agro industry products. In India, agriculture is largely influenced by rainwater which is highly unpredictable. Agriculture growth also depends on diverse soil parameters, namely Nitrogen, Phosphorus, Potassium, Crop rotation, Soil moisture, surface temperature and also on weather aspects which include temperature, rainfall, etc. India now is rapidly progressing towards technical development. Thus, technology will prove to be beneficial to agriculture which will increase crop productivity resulting in better yields to the farmer. The proposed project provides a solution for Smart Agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent.

Analysis of agriculture data using data mining techniques: application of big data, JharnaMajumdar, SnehaNaraseeyappa and ShilpaAnkalaki, in agriculture sector where farmers and agribusinesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This paper focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniques like PAM, CLARA, DBSCAN and Multiple

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

Linear Regression. Mining the large amount of existing crop, soil and climatic data, and analyzing new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change.

FLOW CHART

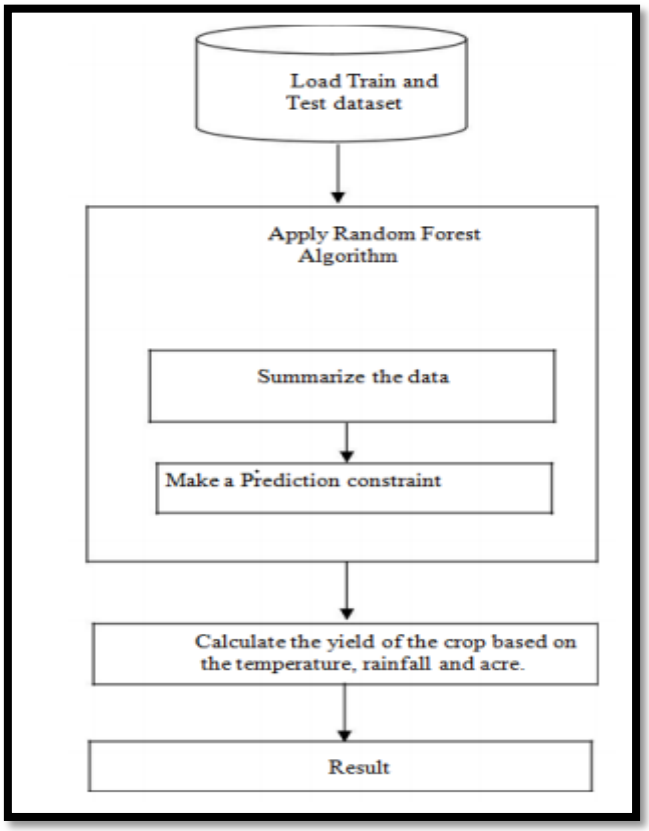


Figure1.3: Flow Chart

VII.ALGORTITHM

SVMs are binary classifiers that construct a linear separating hyper-plane to classify data instances. SVMs are used for classification, regression, and clustering. In farming, they are used to predict yield and quality of crops as well as livestock production [24].

* **Python** is a high level and effective general use programming language. It supports multi-paradigms. Python has a large standard library which provides tools suited to perform various tasks. Python is a simple, less-clustered language with extensive features and libraries. Different programming abilities are utilized for performing the experiment in our work. In this thesis, the following python libraries were used [19]
* **Pandas** - It is a python package that provides expressive data structures designed to work with both relational and labelled data. It is an open source python library that allows reading and writing data between data structures [18]
* **Numpy** - It is an open source python package for scientific computing. Numpy also adds fast array processing capacities to python [19]
* **Matplotlib**- It is an open source python package used for making plots and 2D representations. It integrates with python to give effective and interactive plots for visualization [21]
* **Tensorflow** - It is a mathematical open source python library designed by Google Brain Team for Machine intelligence
* **Sklearn** - It is an open source python machine learning library designed to work alongside Numpy. It features various machine learning algorithms for classification, clustering and regression [27].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

SVM Classification

Support Vector Machine algorithm is prominent data analysis methodology and it is used for classification and regression techniques. Here the data points have been plotted using n-dimensional space with the value of particular characteristics as the value of a specific coordinate. The classification is done by finding the hyper-plane line that differentiates the classes separately [23].

Step1: Read the crop dataset.

Step2: Create the data frame and extraction feature of Crop production, year, temperature, mean rainfall and mean temperature dataset.

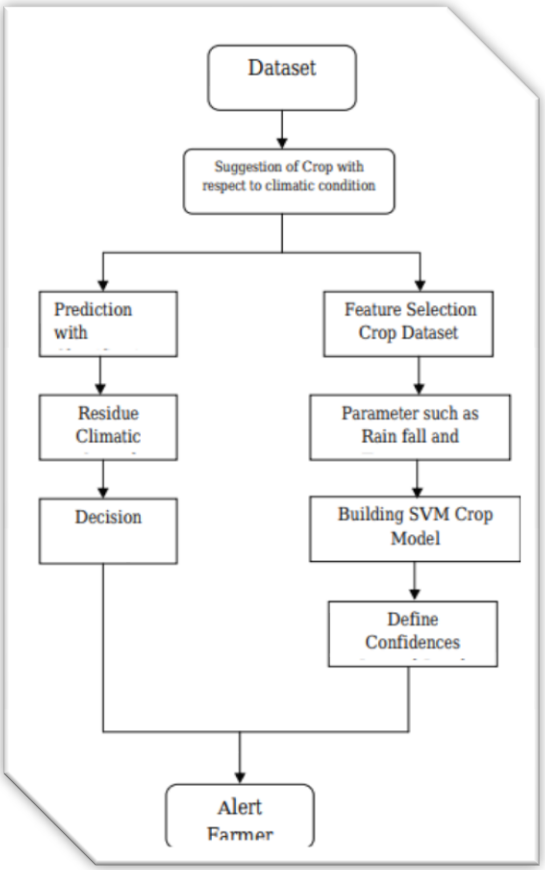
Step 3: Create the SVM class using e107 package and linear, non-linear and kernel sequences model

.

Step4: Predication crop for temperature: The first phase read data frame and set list of crop per year, area and temperature. The second phase applied support vector matrices to prediction state for crop dataset. The SVM results calculate on regression format dataset.

Step 5: Predication crop for year of production: The first phase read data frame and set list of crop per season, rainfall and temperature. The second phase applied support vector matrices to prediction state for crop dataset. The SVM results calculate on regression format dataset.

Step 6: Predication crop for Rainfall mean value: The first phase read data frame and set list of crop per season, rainfall and crop. The second phase applied support vector matrices to prediction state for crop dataset. The SVM results calculated on regression format dataset and linear search.



ARCHITECTURE DIAGRAM

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

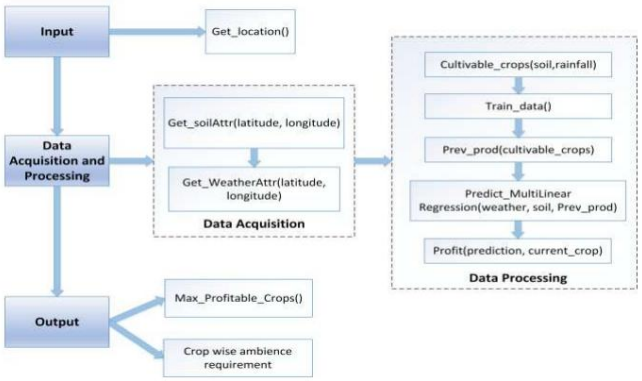


Figure1.5: Architecture Diagram [19]

Input: The prediction of crop is dependent on numerous factors such as Soil Nutrients, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system.
Data Acquisition: Depending on the current user location, the system mines the soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set.
Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop. The prediction is based on past production data of crops i.e.: identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner.
Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

VIII.DATASET

DATASET DESCRIPTION

It can be quite hard to find a specific dataset to use for a variety of machine learning problems or to even experiment on. The list below does not only contain great datasets for experimentation but also contains a description, usage examples and in some cases the algorithm code to solve the machine learning problem associated with that dataset [29].

Kaggle Datasets

This is one of my favourite dataset locations. Each dataset is a small community where you can have a discussion about data, find some public code or create your own projects in Kernels. They contain annumerous amount of real-life datasets of all shapes and sizes and in many different formats. You can also see “Kernels” associated with each dataset where many different data scientists have provided notebooks to analyze the dataset. Sometimes you can find notebooks with algorithms that solve the prediction problem in this specific dataset [29].

IX.CONCLUSION

By analyzing a number of papers we conclude that various data mining techniques such as SVM, Decision tree,KNN etc. are implemented on the input data to assess the best performance yielding method.The proposed work is to increase the crop production under the diferent climatic conditions.

ACKNOWLEDGMENT

We express our sincere thanks to Project guide, Prof. HarshaBhute, for her continuous support. We also thankful to our Head of Department of Computer Prof. B. B. GiteFor support.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

REFERENCES

[1].Zhao, Yi, Noemi Vergopolan, Kathy Baylis, Jordan Blekking, Kelly Caylor, Tom Evans, Stacey Giroux, Justin Sheffield, and Lyndon Estes. "Comparing empirical and survey-based yield forecasts in a dryland agro-ecosystem."Agricultural and Forest Meteorology 262 (2018): 147-156.

[2].Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." In INFORMS InternationalConference on Service Science, pp. 139-147. Springer, Cham, 2019.

[3].Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang and F. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," in IEEETransactions on Computational Social Systems,vol.6,no.3,pp.547

[4].N.Gandhi and L.J. Armstrong, "Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India", Proceedings of the 10th INDIACom-2016, 3rd 2016 IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, 16th to 18th March 2016.

[5].Shweta Srivastava, DiwakarYagysen,"Implementaion of Genetic Algorithm for Agriculture System", International Journal of New Innovations in Engineering and Technology Volume 5 Issue 1-May 2016.

[6].Sandeep V.Rode,Snehal S.Dahikar,"Agricultural Crop Yield Prediction using Artificial Neural Network Approach", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol.2,Issue 1, January 2014.

[7]. Pantazi XE, Moshou D, Alexandridis T, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric. 2016;121:57–65.

[8]. Veenadhari S, Misra B, Singh D. Machine learning approach for forecasting crop yield based on climatic parameters. In: Paper presented at international conference on computer communication and informatics (ICCCI-2014), Coimbatore. 2014.

[9]. . A. Kumar & N. Kannathasan, (2011), —A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining —, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3

[10]. Anwiti Jain, AnadRajavat, RupaliBhartiya —Design, Analysis and Implementation of Modified K-Mean Algorithm for Large Data-Set to Increase Scalability and Efficiency‖, IEEE Explore December 2012

[11]. N.Neelaveni, Ms. S. Rajeswari —DATA MINING IN AGRICULTURE- a Survey‖ Volume 4, Issue 4 [4] https:// docs.oracle.com/cd/B28359_01/datamine.111/b28129/market_ basket.htm#BABFDDCG

[12]. Ability of machine learning methods for massive crop yield prediction. Span J Agric Res. 2014;12(2):313–28

[13]. Ramesh D, Vishnu Vardhan B. Data mining techniques and applications to agricultural yield data. In: International journal of advanced research in computer and communication engineering. 2013; 2(9).

[14]. H. Anandakumar and K. Umamaheswari, "A bio-inspired swarm intelligence technique for social aware cognitive radio handovers," Computers & Electrical Engineering, vol. 71, pp. 925–937, Oct. 2018.

[15]. C.V. Arulkumar, G. Selvayinayagam and J. Vasuki, "Enhancement in face recognition using PFS using Matlab," International Journal of Computer Science & Management Research, vol. 1(1), pp. 282-288, 2012

[16]. X. Chen and P. Cournede, "Model-driven and data-driven approaches for crop yield prediction: analysis and comparison," International Scholarly and Scientific Research & Innovation, vol. 11, no. 7, pp. 334–342, 2017.

[17]. P. Mitra, A. Mukherjee, and A. Chandra, "Machine learning for prediction of crop yield: a case study on rain-fed area rice yield from West Bengal, India," in Proceedings of the DSANRM2017: Workshop on Data Science for Agriculture and Natural Resource ManagementInternational Institute of Information Technology (IIIT, Hyderabad), Hyderabad, India, December 2017.

[18]. P. Priya, U. Muthaiah, and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," International Journal of Engineering Sciences & Research Technology, vol. 7, pp. 1–7, 2018.

[19]. J. H. Jeong, J. P. Resop, N. D. Mueller et al., "Random forests for global and regional crop yield predictions," PLoS One, vol. 11, no. 6, Article ID e0156571, 2016.

[20]. K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[21]. Aditya Shastry, H.A Sanjayand E.Bhanushree,"Prediction of crop yield using Regression Technique", International Journal of computing12 (2):96-102 2017,ISSN:1816-9503

[22]. E. Manjula , S. Djodiltachoumy,"A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017

[23]. Mrs.K.R.SriPreethaa, S.Nishanthini, D.SanthiyaK.Vani Shree ,"CropYieldPrediction",International Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III,Issue III, March- 2016

[24]. JharnaMajumdar, SnehaNaraseeyappa and ShilpaAnkalaki, "Analysis of agriculture data using datamining techniques: application of big data"Majumdar et al. J Big Data (2017) 4:20 DOI 10.1186/s40537-017-0077-4

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904052 |

[25]. Shweta Srivastava, DiwakarYagysen,"Implementaion of Genetic Algorithm for Agriculture System", International Journal of New Innovations in Engineering and Technology Volume 5 Issue 1-May 2016.
[26]. YogeshGandge, Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction", 2017 International Conference on Electrical Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).
[27]. Bhuvana, Dr.C.Yamini (2015), 'Survey on Classification Algorithms in Data mining.' International Conference on Recent Advances in Engineering Science and Management
[28]. AakunuriManjula, Dr.G .Narsimha (2015), 'XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction', Conference on Intelligent Systems and Control (ISCO)
[29]. Tng Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms", Proceedings of the twenty-first international conference on Machine Learning.
[30]. Umid Kumar Dey, Abdullah HasanMasud, Mohammed Nazim Uddin, "Rice yield prediction model using data mining", International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.
[31]. N.Gandhi and L.J. Armstrong, "Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India", Proceedings of the 10th INDIACom-2016, 3rd 2016 IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, 16th to 18th March 2016

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details