

# Agriculture Analysis Using Data Mining And Machine Learning Techniques

Mr. S. KAVIARASAN<sup>1</sup>, R. TEJA<sup>2</sup>, M.SRITEJA<sup>3</sup>, SHEIK. MOHAMMED SHOAIB<sup>4</sup>

<sup>1</sup>ASSOCIATE PROFESSOR, DEPT. OF COMPUTER SCIENCE AND ENGG., PANIMALAR INSTITUTE OF TECHNOLOGY, CHENNAI, INDIA.

<sup>2,3,4</sup>STUDENT, DEPT. OF COMPUTER SCIENCE AND ENGINEERING, PANIMALAR INSTITUTE OF TECHNOLOGY, CHENNAI, INDIA.

## Abstract

Agriculture is a significant application in India. The cutting edge advancements can change the circumstance of ranchers and decision making in agrarian field in a better way. Python is utilized as a front end for investigating the horticultural informational collection. Jupyter Notebook is the information mining device used to foresee the yield creation. The boundary remembers for the dataset are precipitation, temperature, reference crop, evapotranspiration, region, creation and yield for the season from January to December for the years 2000 to 2018. The information mining procedures like K-Means Clustering, KNN, SVM, and Bayesian organization calculation where high exactness can be accomplished.

**Keywords:** Bayesian Network, Support Vector Machine, K- Nearest Neighbour, K- Means Clustering.

## 1. Introduction

Farming is the significant wellspring of the Indian Economy. Step by step, the populace increments. So the interest of food increments. To dispose of these circumstances ranchers, farming researchers, and scientists are going after for better harvest yield.

The breaking down cycle of covered up designs as per different points of view for characterization and changed over into significant data is called as information mining in which information is orchestrated specifically regions like information archive. The proficient examination utilizing information mining methods assist ranchers with taking choices. These informations help them to decrease of expenses and expanding the creation rate. The information mining measure incorporates following advances: Extracting, changing and stacking information in a vault and dealing with the information in

multidimensional data sets. Information mining providedata admittance to investigators utilizing application programming. The investigated information is effortlessly addressed utilizing charts.

## 2. Classification Techniques SVM Classification

Backing Vector Machine calculation is unmistakable information examination approach and it is utilized for characterization and relapse procedures. Here the information focuses have been plotted utilizing n-dimensional space with the worth of specific attributes as the worth of a particular organize. The order is finished by tracking down the hyper-plane line that separate the classes independently.

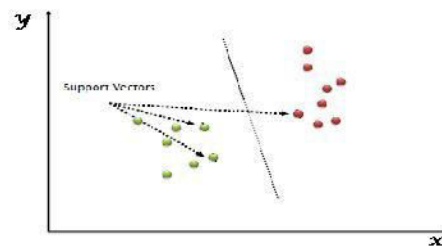


Figure 1 SVM Implementation diagram

In figure 1, the SVM Implementation chart contains Support Vectors and hyper-plane. Backing Vectors are The co-ordinates of a particular class is known as support vectors. Hyperplane line is utilized to isolate the two classes..

In figure 2, the ideas of harvest regarding environment, precipitation, choices, and so on, are considered to make the ranchers aware of anticipate the yield.

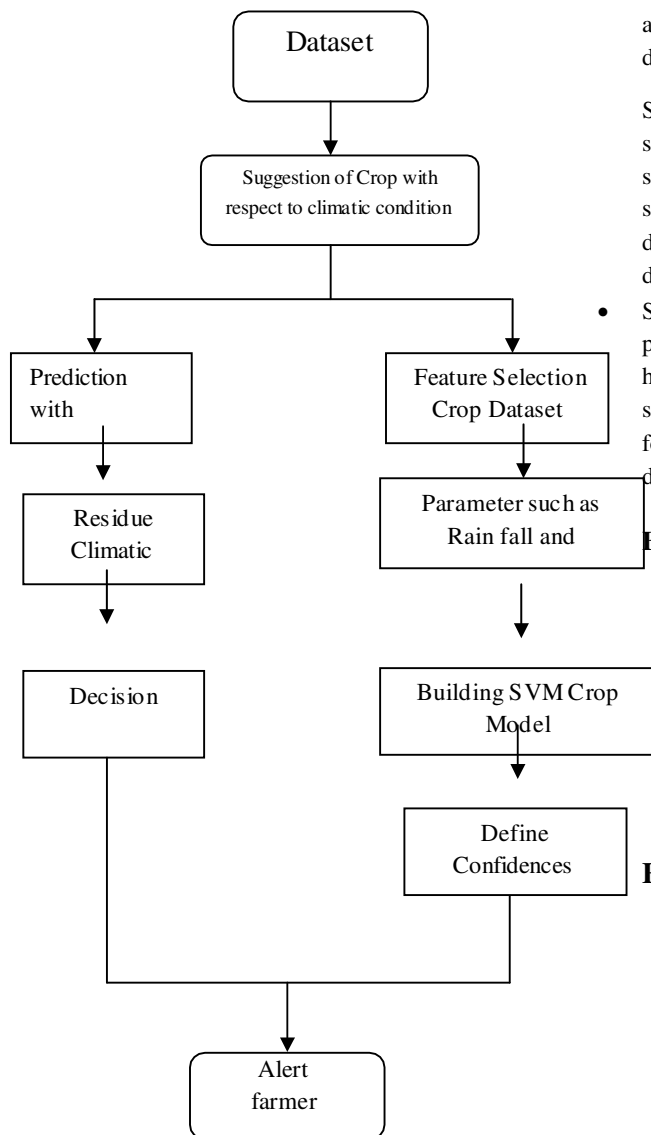


Figure 2 SVM Classification system flowdiagram  
SVM Algorithm

- Step 1: Read the harvest dataset.
- Step 2: Create the information casing and extraction highlight of Crop creation, year, temperature, mean precipitation and mean temperature dataset.
- Step 3: Create the SVM class utilizing e107 bundle and direct, non-straight and bit arrangements model.
- Step 4: Predication crop for temperature: The principal stage read information casing and setrundownofyield

each year, region and temperature. The subsequent stage applied help vector networks to forecast state for crop dataset. The SVM results compute on relapse design dataset.

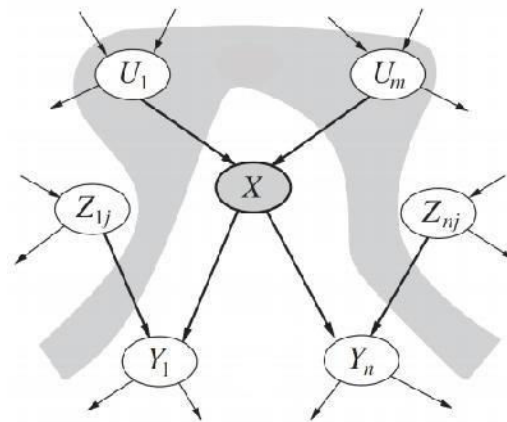
Step 5: Predication crop for year of creation: The main stage read information edge and set rundown of yield per season, precipitation and temperature. The subsequent stage applied help vector grids to forecast state for crop dataset. The SVM results compute on relapse design dataset.

- Step 6: Predication crop for Rainfall mean worth: The primary stage read information edge and set rundown of harvest per season, precipitation and yield. The subsequent stage applied help vector networks to forecast state for crop dataset. The SVM results determined on relapse design dataset and straight hunt.

### Bayesian Network

A Bayesian organization is a non-cyclic diagram comprises of edges and hubs with bearings in which each edge addresses to a contingent reliance, and every hub addresses to a special arbitrary variable. The probabilistic graphical model that utilizes Bayesian deduction for calculations are called as Bayesian organizations.

Figure 3 Bayesian Network



In Figure 3, Bayesian Network is represented using an acyclic graph. It consists of nodes and edges.

A Selective Attribute network is a smaller, graphical model of a likelihood dispersion which allocates a likelihood to each occasion of interest. For instance, in the harvest identification, a Selective Attribute

organization can be utilized to ascertain the likelihood worth of a specific yield expectation given the season showed by a rice crop dataset.

Particular Attribute Classification is a fundamental undertaking in information investigation and example acknowledgment that requires the development of a classifier, that is, a capacity that allocates a class name to occasions depicted by a bunch of qualities. The acceptance of classifiers from informational indexes of pre-grouped occasions is a focal issue in AI. Various ways to deal with this issue depend on different utilitarian portrayals, for example, choice trees, choice records, neural organizations, choice diagrams, and rules. given C at whatever point  $\Pr(A|B, C) = \Pr(A|C)$  for all potential upsides of A, B and C, at whatever point  $\Pr(C) > 0$ .

Bayesian organization B, that encodes a dissemination PB (Crop\_Year, Crop\_Production, Crop\_Temperature, Crop\_Rainfall, Crop\_n), from a given preparing set. We would then be able to utilize the subsequent model so that given a bunch of properties a1. . . an, the classifier dependent on B returns the name c that boosts the back likelihood  $PB(c|Crop\_Yea\_a1, Crop\_Production\_a2, . . . , Crop\_an)$ . Note that, by inciting classifiers as such, we are tending to the principle concern communicated in the presentation: eliminate the inclination by the freedom suspicions installed in the gullible Bayesian classifier.

Organization with a moderately decent MDL score that performs ineffectively as a classifier and to comprehend the conceivable inconsistency between great prescient precision and great MDL score, we should reconsider the MDL score. Review that the log probability term in Equation 2 is the one that actions the nature of the learned model, and that  $D = \{Crop\_u1, . . . , Crop\_uN\}$  means the preparation set. In a characterization task, each Crop\_ui is a tuple of the structure Crop\_hai1, . . . ,Crop\_ain,Crop\_cii that appoints qualities to the properties Crop\_A1, . . . ,Crop\_An and to the class variable Crop\_C. We can change the log probability work as

$$LL(Crop\_B|Crop\_D) = \frac{1}{N} \sum_{i=1}^N \log PB(Crop\_ci|Crop\_ai1, . . . ,Crop\_ain) + N \sum_{i=1}^N \log PB(Crop\_ai1, . . . , Crop\_ain)$$

$$=1 \log PB(Crop\_ai1, . . . , Crop\_ain)$$

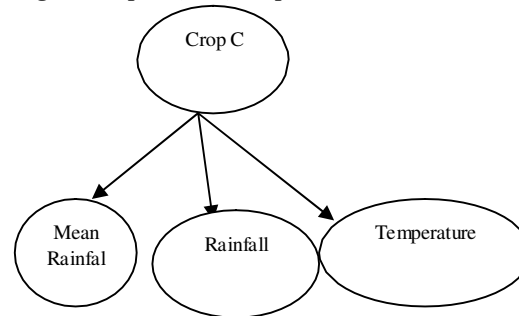


Figure 4 Structure of the Selective Attribute

In Figure 4, The credits are autonomous from each other. Ex: Mean precipitation, precipitation, temperature are doesn't depaend upon each other. These three are rely upon the harvest C.

### Selective Attribute Networks Algorithm

- Step 1: Read the Crop dataset.
- Step 2: Create the information list from Crop dataset and highlight extraction for expectation crop subtleties.
- Step 3: Create the Selective Attribute (Bayesian) net utilizing Neuralnet bundle.
- Step 4: To make a Rice, Coconunt, Arecanut, Black pepper and Dry ginger harvest base net utilizing model2network work in Jupyter Notebook.
- Step 5: To peruse the dataset and dole out the information into CA, CS, Ck, Cw, CB, CL and CE object variable. The item contains Crop year, Crop creation, Crop Area, Crop mean Temperature, Crop mean Rainfall, mean Crop Temperature and Rainfall esteems, and regions subtleties is associated the Bayesian organization.
- Step 6: The yield order rule and likelihood esteems dole out the Bayesian net.
- Step 7: To make custom Bayesian net utilizing Bayesian hypothesis in Rice, Coconunt, Arecanut, Black pepper and Dry ginger yield, and so on,

- Step 8: To detect the Bayesian Rule for Rice, Coconunt, Arecanut, Black pepper, Dry ginger crop, etc..., Return the accuracy values.
- Step 9: Repeat the process Step 3 to Step 8.
- Step 10: To exactness calculate the TP, TN,FP and FN values.

## 2.5 KNN (K Nearest Neighbours) Classifier

K-Nearest Neighbors is one of the grouping calculations in Machine Learning. It is otherwise called

as administered learning model and apathetic calculation as a result of example learning. It is utilill for various applications like example acknowledgment information mining and

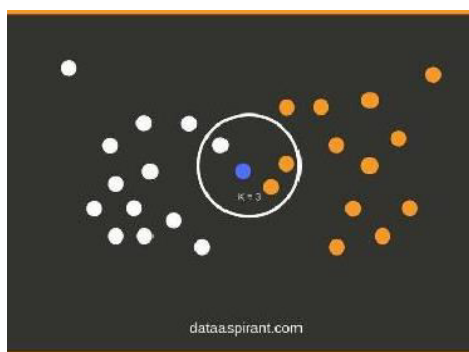
interruption location. Execution is basic for little informational collections. The preparation information needn't bother with any knowlegde about the construction of information before the investigation.

Disadvantage of this classifier is discovering the closest neighbor for each sample. Lot of space is

required when the preparation information is huge. The distance between test information and the preparation information ought to be determined for each test information. In this way, the testing needs a great deal of time. There are two stages in this classifier: Training stage: Save the models, Prediction stage: Get the test occasion and discover the preparation set.

### Figure.5 KNN Classifier Diagram

In Figure.5 The **white and orange** circles are two different classes. There are 26 circles. The blue circle is the



target circle used for prediction. Here, the value of  $k$  is three. The Euclidean distance is used to calculate equal distance. The classes are close when the similarity score is less. In this image, we have calculated the minimum distance.

## 3. Clustering Technologies

### K-Means Clustering

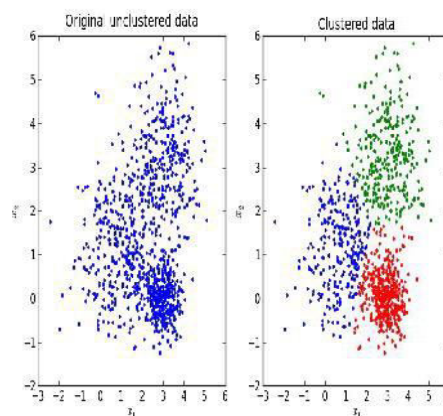
K-Means Clustering is one of the bunching method that interaction a gathering of information focuses into few groups. For example, the items in a mall are bunched in different categories (medium, large, X-L are grouped as the size of the dress). This is a subjective strategy for parting a gathering of information.

A quantitative approach is used to qualities of the items measure unique  $k$  bunches the characteristics of the products. In number of information focuses must be divided. The aim of this methodology is to appoint a bunch to each data point. K-implies calculation plans to discover the group positions that minimize the *distance* from the data points to the cluster.

### ALGORITHM

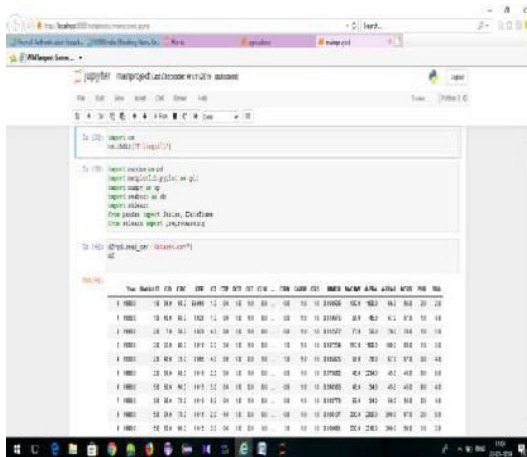
STEP1: The  $k$  bunch centroids are utilized to name new information

STEP2: Marks are need for the preparation information



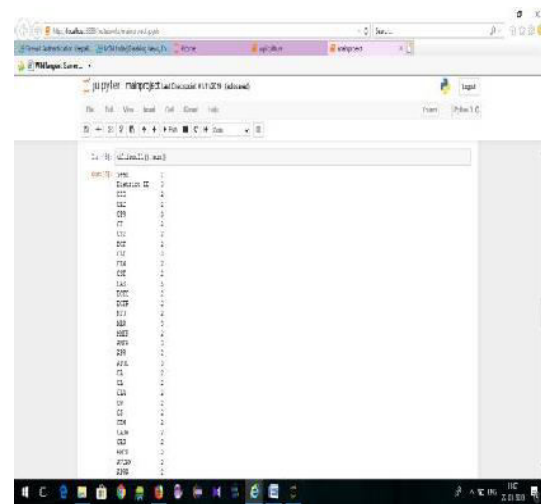
**Figure.6 Unclustered data and Clustered data**

In Figure.6 The unclustered information comprises of in general dataset focuses together and the bunched



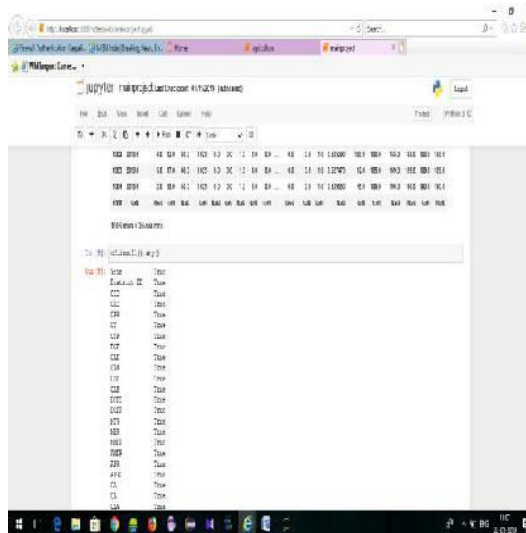
**Figure.7 Importing the dataset**

In Figure.7, Dataset in dominate sheet is imported as csv record utilizing python language in Jupyter Notebook.



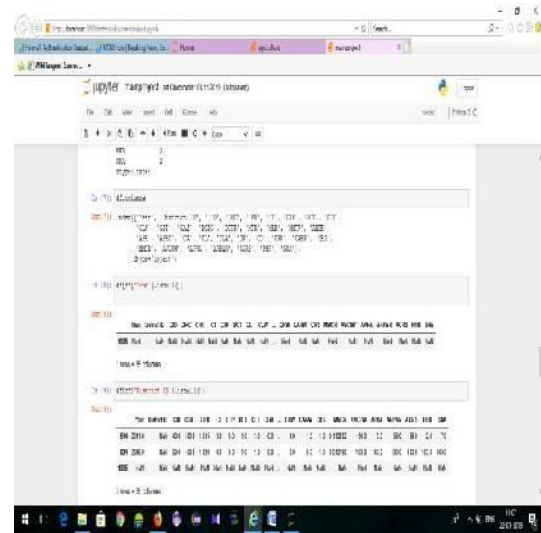
**Figure.9 Sum of null values**

In Figure.9 Amount of invalid qualities in each line is checked.



**Figure.8 Checking null values**

In Figure.8 Null values are boolean checked using values.

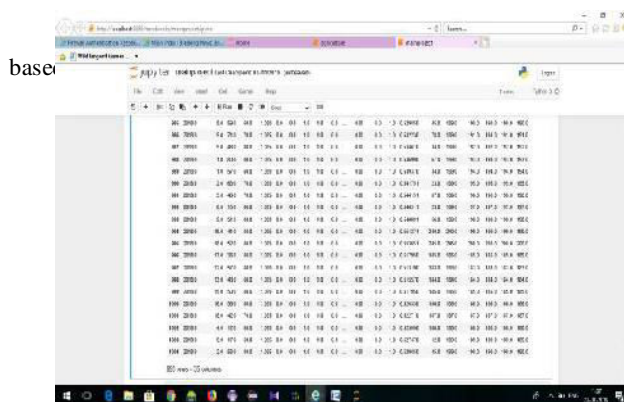
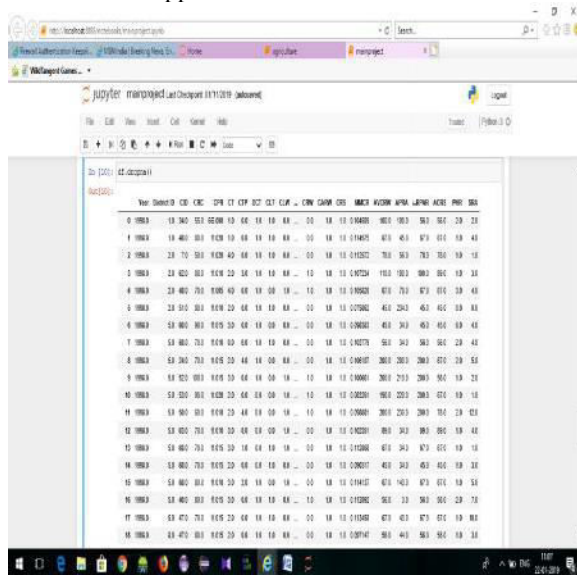


**Figure.10 Column names and Place of null values in each row**

In Figure.10 Segment names are shown and spot of invalid qualities in each line is addressed as NaN.



farming is an upcoming research field. Productive procedures can be developed and used for solving complex agricultural problems using data mining. Future enhancement of this agriculture analysis is to predict the crop yield using these techniques. It is helpful for making crop decisions for ranchers and government organizations. In future, the ANN and NN classification approach can be used to improve the better classification



**Figure.12 After Dropping null values**

In Figure.12 After dropping values of the null the actual number of rows in reduced. is

## 4. Conclusion

In this analysis, we used some of the common data mining techniques in the field of agriculture. Some of these techniques, such as the k-means, k nearest neighbor, SVM, and bayesian network are discussed and an application in agriculture for each of these techniques is presented. Data mining in

execution of the harvest yield expectation. It could comprehension of the great dimensional between complex yearly and occasional climatic examples which decide crop yield helps the two ranchers and other leaders to have the option to anticipate the impacts of dry season and other climatic conditions. References

- [1] [1] Veenadhari S, Misra B, Singh CD. Information digging methods for anticipating crop efficiency—An audit article. In: IJCST. 2011; 2(1).
- [2] [2] Jain A, Murty MN, Flynn PJ. Information grouping: an audit. ACM Comput Surv. 1999;31(3):264–323.
- [3] [3] Berkhin P. A study of grouping information mining method. In: Kogan J, Nicholas C, Teboulle M, editors. Gathering multidimensional information.
- [4] Berlin: Springer; 2006. p. 25–72V. Arulkumar. "An Intelligent Technique for Uniquely Recognising Face and Finger Image Using Learning Vector Quantisation (LVQ)- Template Key Generation," International Journal of Biomedical Engineering and Technology 26, no. 3/4 (February 2, 2018): 237-49.
- [5] Han J, Kamber M. Data mining: concepts and techniques. Massachusetts: Morgan Kaufmann Publishers; 2001.
- [6] 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

- [7] 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.
- [8] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise.

In: Paper presented at International conference on knowledge discovery and data mining. 1996

- [9] C.V. Arulkumar et al., "Secure Communication in Unstructured P2P Networks based on Reputation Management and Self Certification", International Journal of Computer Applications, vol. 15, pp. 1-3, 2012.
- [10] 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).
- [11] ability of machine learning methods for massive crop yield prediction. Span J Agric Res. 2014;12(2):313–28