



# Research Journal of Pharmaceutical, Biological and Chemical Sciences

## Remote Sensing Image and Data Mining in Agriculture.

Neelanjan Bhattacharyya\*, Anthanna Sandeep, and Venkatesan M.

School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India

### ABSTRACT

Data Mining is emerging research field in Agriculture crop yield analysis. In this paper we examine about the Exchange of vitality from one living space to the next can impact biological community capacities. Information mining systems are being used, for example, Neural Networks, and Support Vector Machines and K-closest neighbor (KNN). Populace element and environmental change impact sly affect farming Being a carbon SINK, because of a higher rate of carbon dioxide retention, or a carbon SOURCE, because of deforestation and timberland fire, backwoods adjust environmental change and climatic CO<sub>2</sub>. Expanded CO<sub>2</sub> levels directly affect trim yields and, at last, sustenance security. Distinguishing and refining the adjustment activities while keeping up nourishment security and natural maintainability is important. Late utilization of remote detecting has been the essence of the improvement of satellite images in different data layer and different data mining techniques have exhumed measurements used in guaranteeing the nature of information from gathering to estimation.

**Keywords:** Neural Network, Vector Machine, Source, Sink, K-closest neighbor.

*Corresponding author*



## INTRODUCTION

Analysts in different worldwide projects have concentrated on the potential utilization of remotely detected information to acquire precise data ashore surface procedures and conditions. These studies have exhibited that quantitative evaluation of the dirt vegetation-air exchange procedures can prompt to a superior comprehension of the connections between product development and water administration. Data ashore surface can now be acquired at an extensive variety of spatial (5–5000 m) and fleeting resolutions (0.5–24 days) [1]. In any case, despite the fact that extensive advance has been made in the course of recent years in research applications, remotely detected information remain underutilized by honing water asset managers. In This paper we recognize the genuine issues that remote detecting could understand, yet with extra innovative work. As new water turns into an undeniably rare asset, all chances to better oversee water utilizes, especially in flooded agribusiness, must be taken. An imperative test in the field of water assets is to use the opportune, objective and exact data gave by remote detecting. Remote detecting give all around planned systems to dissecting harvest and timber land status over huge regions continuously [4]. The goal of this unique issue is to display the most recent improvements in remote detecting in ranger service and horticulture that may discover better answers for maintainable woods and cropland administration. In the common frameworks, for example, Mediterranean locales, and numerous other comparable frameworks, the two biological communities are principle drivers of economical environment and our prosperity, where the nature of soil, water and air is regularly changed by extreme dry spells and backwoods fires [3]. Data Mining is the way toward removing valuable and important data from vast arrangements of information. Information Mining in farming field is a generally novel research field. In this paper depict a review of Data Mining methods connected to rural and their applications to rural related ranges. Yield expectation is an essential with a specific end goal to take care of agrarian issue. Any rancher is occupied with knowing how much yield he is going to anticipate. Previously, yield expectation was performed by considering rancher's experience on specific field and harvest [2]. Consider that information are accessible for quite a while back to the past, where the comparing yield forecasts have been recorded. In any of Data Mining methods the preparation information is to be gathered from some time back to the past and the assembled information is utilized as a part of terms of preparing which must be abused to figure out how to arrange future yield expectations [5].

## DATAMINING TECHNIQUES

For this venture reason we may chip away at various data mining technique. Data Mining methods are mostly separated in two gatherings, grouping and bunching procedures. Arrangement methods are intended for grouping obscure specimens utilizing data gave by an arrangement of ordered examples. This set is typically alluded to as a preparation set as it is utilized to prepare the arrangement method how to play out its grouping [7]. For the most part, Neural Networks, and Support Vector Machines ,these two arrangement systems gain from preparing set how to characterize obscure examples.

Another grouping system, K-Nearest Neighbor, does not have any learning stage, since it utilizes the preparation set each time an arrangement must be performed. A preparation set is known, and it is utilized to group tests of obscure arrangement. The essential suspicion in the K-Nearest Neighbor calculation is that comparable specimens ought to have comparable order [6]. The parameter K demonstrates the quantity of comparable known specimens utilized for doling out a grouping to an obscure example. The K-Nearest Neighbor utilizes the data as a part of the preparation set, however it doesn't remove any govern for ordering the other.

Another way is clustering. In the occasion a preparation set not accessible, there is no past information about the information to characterize. For this situation, grouping methods can be utilized to part an arrangement of obscure examples into bunches. A standout amongst the most utilized grouping system is the K-Means calculation [9]. Given an arrangement of information with obscure order, the point is to discover a segment of the set in which comparable information are assembled in a similar group. The parameter K assumes a critical part as it determines the quantity of groups in which the information must be parceled. The thought behind the K-Means calculation is, given a specific parcel of the information in K groups, the focuses of the bunches can be figured as the method for all examples having a place with a groups [7]. The focal point of the bunch can be considered as the agent of the group, in light of the fact that the inside is very near all examples in the group, and along these lines it is like every one of them. There are a few hindrances in utilizing

K-Means strategy. One of the burdens could be the decision of the parameter K. Another issue that necessitates consideration is the computational cost of the calculation [6]. There are other Data Mining procedures factually based strategies, for example, Principle Component Analysis (PCA), Regression Model and Biclustering Techniques have a few applications in agribusiness or rural-related fields.

## APPLICATION

There are several applications around data processing and remote sensing. Data processing has varied applications in Agriculture fields, weather outlook fields and decisive average downfall, prediction of pollution etc. For instance, k-mean rule is employed to forecast the pollution in air [5]. On the opposite hand, k-nearest rule is employed to forecast completely different variables within the weather and the way they're ever-changing on the daily basis. Data processing techniques are applied to review sound recognition issues. As an example, it is accustomed to classify the sound of birds and alternative completely different sounds. On the opposite hand, K-Nearest Neighbor approach is used to judge forest inventories and to estimate forest variables for analyzing satellite representational processes. It can also be accustomed to classify eggs as fertile and for PC Vision to acknowledge cracks in eggs [7]. Some one works on classifying dish sauce and uses SVMs for detective work, weed and chemical element stress in corn. Data processing techniques are typically accustomed to study soil characteristics. As an example, the K-Means approach is employed for classifying soils together with GPS-based technologies. A K-Means approach is employed to classify soils and plants and SVMs to classify crops. Apples are checked for mistreatment completely different approaches before causing them to the market [8]. Once more, K-Means approach to research colour pictures of fruits as they run on conveyor belts. X-ray pictures of apples to observe the presence of water cores, and a neural network is trained for discriminating between sensible and unhealthy apples. Supervised bi-clustering technique to a data set of wine fermentations with the aim of choosing and discovering the options that are accountable for the problematic fermentations and conjointly exploit the chosen options for predicting the standard of latest fermentations. Similarly, sensors are accustomed to smell milk, that's classified for mistreatment SVMs.

## OVERVIEW OF DATA

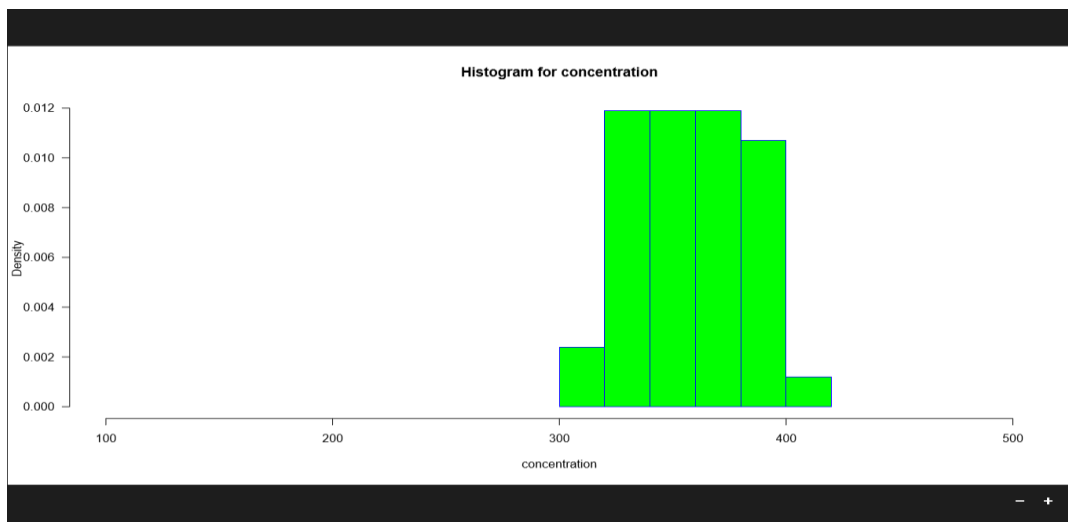
The data available in this paper are taken from the year 1975 to 2016. Here we are taking the year-wise CO<sub>2</sub> concentration and growth rate of soyabeans in Madhya Pradesh. We are considering Madhya Pradesh as it is considered to be the highest production area of soyabeans followed by Maharashtra. Madhya Pradesh produces the highest and it contributes 75 percent of overall production of soyabeans. In this paper we are collecting the data of last 40 years and showing how CO<sub>2</sub> concentration is increasing in each year and how it is affecting on crops. Here we are taking soyabeans as crop and showing growth rate of soyabeans decreasing year by year. During collecting the CO<sub>2</sub> concentration we came to know it increases 2.12 ppm per year for last 10 years i.e. from 2005 to 2016. Before that it increases 1.95 ppm per year for another last 10 years i.e. 1995 to 2004. Like this we have collected the data for last 40 years. And we also collected the data for soyabeans that how the growth rate of its decreasing per year because of increasing rate of soyabeans. It can be shown that growth rate of soyabeans decrease 0.30 per year for last 10 years i.e. from 2005 to 2016. And it decreases 0.22 per year for another last 10 years i.e. 1995 to 2004. Like this we calculated the decreasing growth rate of soyabeans for increasing rate of CO<sub>2</sub> concentration. This data of growth rate is collected for Madhya Pradesh and CO<sub>2</sub> data is collected for its worldwide concentration. Here we have taken three parameters - Year, CO<sub>2</sub> concentration and growth rate.

## RESULT ANALYSIS

Here we are analyzing the result on the basis of collecting data. Here we are representing result in the form of Histogram. We are forming Histogram one for the CO<sub>2</sub> concentration and another one for Growth rate. We are plotting the Histogram with respect to the probability density function. We are plotting the concentration and growth rate in the x-axis and taking the probability density function in the y-axis on two different histograms and representing. We have done all this analysis with the help of data mining and remote sensing. In data mining we have used k-means algorithm in order to get proper clustering. K-mean algorithm is the easiest but very effective clustering method that we have employed for clustering on the basis of year. Here we have done four clustering on the basis of year in which each cluster contains the data of ten years. In

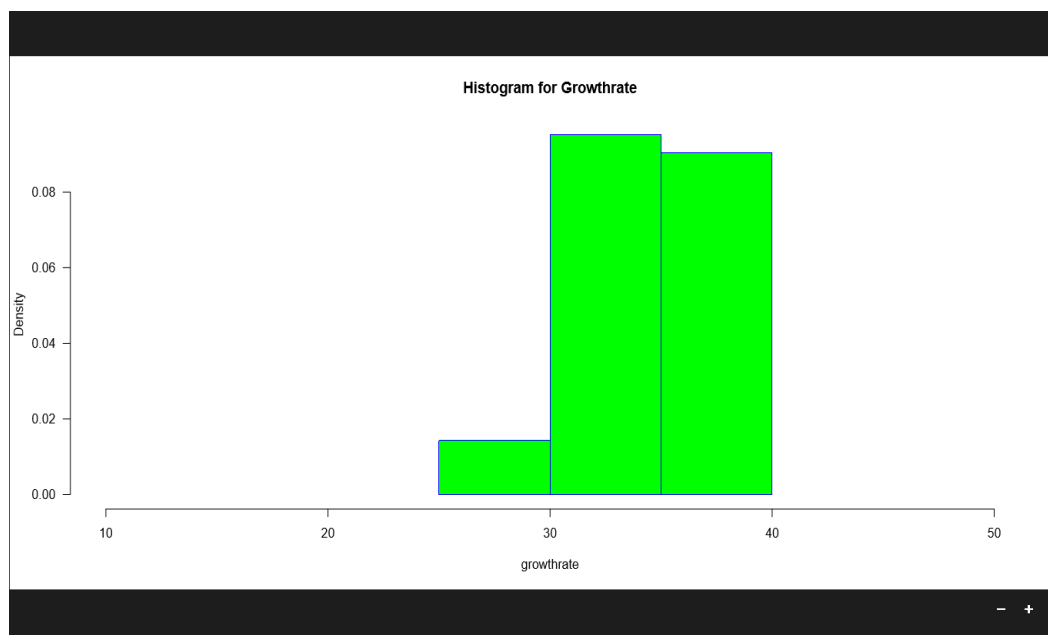
this result it's easily can be shown how growth rate is changing with respect to the probability density function for last 40 years and concentration is changing rapidly.

In the Fig1 Histogram of co2 concentration with respect to the probability density function is given and through this its clearly can be how the concentration increases year wise for last 40 years. First in the beginning it was near about 300ppm but it continuously increases and even cross the 400ppm marks in 2016. Similarly we can describe the decreasing result of growth rate of soyabeans through this Histogram and its as follows:



**Fig1: Histogram of CO2 Concentration With Respect to the Probability Density Function**

In this Fig2 through Histogram, it can be describe that how growth rate is decreases year wise for last 40 years. At the very beginning from where we have taken data it was near about 40 but as the year progress and co2 concentration increases in the air, growth rate decreases and reaches near about 26 in the recent year. In this paper through this diagram we are analyzing our result that how much crop is affected on the increasing rate of CO2 and for this purpose we have taken soyabeans and its decreasing growth rate for Madhya Pradesh and result is describe as above.



**Fig2: Histogram of Growthrate With Respect to the Probability Density Function**

## FUTURE SCOPE

In this paper what we have analyze it can be improved in near future. Instead of that we can make a comparative study of different crops and how much there growth rate is affecting for increasing rate of CO<sub>2</sub> concentration. On basis of that result we can analyse that which crop is affecting more. And if it's not sufficient then instead of taking only one parameter, we can take few more parameter. Like we can include photosynthesis rate and respiratory rate of different crops on different CO<sub>2</sub> concentration. It will help us to elaborately and much more clearly to describe the condition. It can also be further elaborate by taking the data for different state for different crops. It will give us clear view and idea to analyze the effect of increasing CO<sub>2</sub> concentration. And It all can be done with the help of data mining and remote sensing.

## CONCLUSION

In this paper bound data processing techniques were adopted so as to estimate crop yield analysis with existing knowledge. The applications that use the K-Means approach, utilize solely the fundamental formula, where as several different enhancements area unit on the market.

Some data processing techniques haven't none the less been applied to agricultural issues. As associate degree example, Bi-clustering techniques could also be utilized for locating vital info from agricultural-related sets of knowledge. The K-Means formula is in a position to partition the samples in clusters, however no concerns area unit created on the compounds that area unit accountable for this partition. Biclustering will provides this type of knowledge. The longer term work aimed toward the analysis of the whole set of knowledge and can be dedicated to appropriate methods for up the potency of the projected formula.

## REFERENCES

- [1] G. Ruß, Data Mining of Agricultural Yield Data: A Comparison of Regression Models, Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P. Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, 24–37, Springer, 2009.
- [2] Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. Data Mining with neural networks for wheat yield prediction, Proc ICDM 2008, 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.
- [3] P. Wagner and M. Schneider. Economic benefits of neural network-generated site-specific decision rules for nitrogen fertilization. In J. V. Stafford, editor, Proceedings of the 6th European Conference on Precision Agriculture, 775–782, 2007.
- [4] A. Mucherino, A. Urtubia, Feature Selection for Datasets of Wine Fermentations, I3M Conference Proceedings, 10th International Conference on Modeling and Applied Simulation (MAS11), Rome, Italy, September 2011.
- [5] Kovacevic, B. Bajat, B. Gajic, Soil Type Classification and Estimation of Soil Properties using Support Vector Machines, Geoderma 154(3–4), 340–347, 2010.
- [6] Sam Y. Sung, Member, IEEE Computer Society, Zhao Li, Chew L. Tan, and Peter A. Ng, “Forecasting Association Rules Using Existing Data Sets”, iee transactions on knowledge and data engineering, vol 15, no.6, November/December 2003.
- [7] McCarthy, J. and Atkinson, D. (2008) “Innovative Uses of Data Mining Techniques in the Production of Official Statistics,” US Department of Agriculture, National Agricultural Statistics Service, RDD, Paper for United Nations 2009 Statistical Commission Session on Innovations in Official Statistics.
- [8] Boryan, C. (2009) “Remote Sensing of Agriculture: NASS’ Cropland Data Layer Program,” US Department of Agriculture, National Agricultural Statistics Service, RDD Presentation to George Mason University Digital Remote Sensing Seminar, Fairfax, VA.
- [9] Cecere, W. (2008) US Department of Agriculture, National Agricultural Statistics Service, RDD Report 08-in preparation, Fairfax, VA.