International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Prediction of Suitability of Soil for Different Crops using Spatial Data Mining

Prathik A, Anuradha J, Uma K

Abstract: ***The main Objective of Data mining in agriculture is to improvise the productivity based on the data observed and timelines of cultivation. Spatial Data mining, a key to capture the data by proposing sensors on a particular geographical location and observe various parameters to enhance the productivity based on the statistical analysis of data collected. In general, Data mining is an anticipating measurement and prognosticates the various data sets and mutate into useful data sets which can be applied on various applications. In this paper, data mining is applied in bridging the soil conditions to the applicable crop for cultivation in enhancing the productivity and multiple crops cultivation for enriched productivity based on the data sets acquired. A Statistical analysis resulted from a backend algorithm with the data sets and displayed as dashboard with the forecasted productivity. A Grid based clustering algorithm is adhered at the backend for performing analysis on the collected data sets results crop selectivity & productivity timelines. Geographical analysis forms a grid pattern with multiple data sets as matrix results in multiple crop selectivity based on the soil conditions and analyzed data sets obtained from various sensor parameters on a particular location. Data visualization is performed after the algorithmic process at the backend and data stored in the cloud server. Spatial Survey & Collective data Sets analyzed with the algorithm are used to elevate the Crop Selectivity and productivity on a soil based on the Biological Predicts, defoliant and manure usage timelines yields Improved Monetary generation.***

Keywords: Spatial-data mining, Grid based clustering algorithm, Biological Predictions, Geographical Survey.

I. INTRODUCTION

Data Mining with Spatial Vector Quantities approaches the real world application Constraints with the help of geographical extractions and tropical conditions. A Sophisticated forecasting method improvises the timeline for the crop under cultivation and increases the productivity of the crops. The ultimate problem statement of real world multiple data sets from various phenomena are in transforming it into information that enhances the Application and Solves the real world constraints and thus data mining comes into the transformation layer. Capturing data under different and various phenomena and disciple into information groups which are key factors for the application.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Prathik. A, Research Associate, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, E-mail: prathik0505@gmail.com.

Anuradha. J., Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, E-mail: januradha@vit.ac.in.

Uma. K, Associate Professor, School of Advanced Science and Engineering, Vellore Institute of Technology, Vellore, India,

E-mail: kuma@vit.ac.in.

Data sets analyzed via data mining has no constraints and limitations in which any interlinked/non interlinked data sets from the different phenomena can be grouped and formed as a constructive matrix for the application[2]. Prognostic technique most commonly known as predictive investigation method is far better concerned in data mining application than any other techniques. This technique figure out the biological conditions, defoliant & manure usage for the crops over a period of time which in terms yields higher productivity ratio compared with the previous cultivation, before the data mining comes into the picture [3]. In General Application development datasets are from various phenomenal databases like terrestrial database, Object specific database and interactive media database are into the role but sophisticated application includes any one of the above mentioned database will be taken as key value for the development process. Geographical survey datasets or spatial data mining can be in different pattern like distinctive and segregated instructions, abstraction & Summarization of embossed data sets or grouping, Geographical datasets [4]. The main object is to derive the original data from the phenomena's and transforming it into application oriented key value data sets [5]. The Distinctive / Descriptive mining summarize the antecedent data sets and explore to study the behavior of data sets in forthcoming events. In Segregated instructions/ predictive relies on the probabilistic survey in anticipating the solutions based on the current data sets and from probabilistic survey results [6]. Data affiliations are characterized into four Class, Cluster, Association, consecutive pattern. Soil Quality predictive analysis along with weather prediction for a timeline helps farmers in analyzing the soil conditions for cultivation, defoliant & Manure usage. It also helps in irrigation system for the concerned cultivated crop. It is always difficult for the farmers to waste little location in their cultivation space because of the crop cultivated and other crops seems no match with the irrigation system [7]. It could be a hazard if multiple crops cultivated on a same space with the different irrigation level over the crop productivity period. In order to avoid and minimize this erroneous data mining improvises the concept of productivity with precision farming by various backend algorithmic techniques which on later analyzing transforms in to a useful value which can be used by the farmers to take actions on their timelines of crop growth period. The number *n* is much smaller than the matrix of data sets collected and it should be declared with a value before the algorithm initialization. In general, the number *n* defines the number of groups can be formed from the collected data sets [8].

The whole application is based on the identification of dissimilar and similar group of crops based on the data sets evolved and finding a solution in bridging the dissimilar data sets with a co instance with the similar data sets. All the data's will be stored in the cloud server for various multiple statistical processes and for the dashboard process.

II. LITERATURE REVIEW

Data mining in precision farming involves various techniques and algorithms in which each algorithm defines on its own way of improvising the accuracy of predictive forecasting. The various algorithms and techniques were analyzed and taken into consideration for the proposed system. This survey is mainly to identify the pros. Of Grid based clustering algorithm among all other techniques and to analyze the other techniques strengthening the application of Soil and crop selectivity based on the forecasted data sets. The future concentration of the researchers is to enhance the learning system which can be applied to enrich the profiles of soil for increased productivity [9]. The Soil test carried out a Geo-location has multiple data sets which have all the properties of soil like Nutrients, Biological impacts, Moisture content, pH range, energy transformation, Ammonia & Nitrate levels and other chemical contents.

A. Productivity Ratio

A research has been carried out by Mr. Reddy. et.al on the productivity yields based on the cultivation level of crops like bean and canes and reported that the statistical analysis carried out using lilliefor method is non hypothetical analysis of data distribution. The results found to be in a state where the statistical analysis cannot be taken into consideration for all the cases and it is distribute in normal [10]. This Analysis found to be useful information in Risk benefit analysis for crops like beans & canes.

B. Geo-location Data Mining

Geo location based data collectivity research by yashoyardhan.et.al [11] reveals that data sets relevant to the statistical analysis are collected from multiple locations. Preprocessing of data includes cleansing & integration. Cleansing refers to the extraction process carried by extracting relevant data and eliminating the unreliable data sets. Decision making is done by the technique applied on the data sets and patterns. Perception and assessment of patterns is the way of depicting the knowledge acquired on data sets evolved [10].

C. Soil Aggregation

Soil Aggregation/ Cluster is examined and analyzed by Banumathi et.al. In this Fuzzy based algorithm has been examined on a large data set analysis to discover the formation of patterns with the data sets. The pattern formation has a random effect on the data sets in defining the group formation [12]. Whereas the data sets has found the forecasting crops for the different climatic conditions. Rupali et.al [13] analyzed by k-mean algorithm technique which they found a way increasing the productivity ratio and proficiency of growth rate of crops with minimal erroneous. The data sets are grouped on the characteristic standards and they are unified based on the objects of their own type. The algorithm found to be iterative with less number of data sets in front and they require huge data's for the process. Topographical changes are the key incidents in beholding the extraction point from the natural data set exchanges. Statistical data extraction or knowledge extraction is incomplete due to the missing data sets by real world constraints. Ragel et.al has found a solution to solve the missing data problem in statistical analysis of knowledge extraction. The missing data sets can be resolved by using compatible data sets associated with the key elements of extracted data [14]. Investigation on productivity timeline of cotton crop by means of cognitive computing in fuzzy logic is done by Papagerorgiou et.al [15] results in the cognitive mapping of data sets by means of fuzzy algorithmic technique. This algorithm fine tunes the application by including the unexpected data matrix via similar process of statistical analysis with a human description of input data and found to be effective on uncertain events. Xiaoqin et.al [16] has analyzed and done a predictive forecast method in relating the observed data sets for a particular geographical location and forecasted the crop selectivity for the periodic cultivation as well the timelines with increased productivity. The algorithm precludes the data sets driven by the sensors on the geo location and based on the forecast of climatic condition observed by the GIS, the crop prediction for the soil can be done. The spatial importance in data mining especially on the agriculture is the key data set for the precision farming and a team of research from Luke et.al [17] analysed the incorporation of spatial data into the statistical data sets for the increased crop productivity along with the caution of omitting the data. In Specific the caution of omitting the key data sets defines the application to be ineffective for the suitable application. Bayesian model of prediction finds high accuracy in spatial data extraction and also predicts data considering all the phenomenal constraints than any other mining methods. Biological changes results the crop growth and productivity and to predict or forecast the growth of crop, this climatic change has to be taken into consideration while modelling the statistical analysis. A simulation under eight climatic conditions was examined by Dimitrio et.al [18] in which the data sets are taken for the statistical analysis model and formed a patter relates with the crop productivity and growth. Computing the resemblance between the phenomenal data's mining with pattern observation was analysed by Raymond et.al [19] with IR method of approximation is the enhanced method of data collectively which defines the distance between the patterns observed from the data sets relates the efficiency of relevant to the irrelevant data sets. The bridging between these two will be done on the Bayesian model and the computation process duration not varying to the maximum extent and the iteration will be four the max. Density based clustering algorithm was analyzed by Majumdar et.al [20] defines the algorithm sweeps out the external irrelevant data sets observed during the extraction. Calculating the minimum *eps* value produced by this algorithm is the key element to sweep out the external irrelevant data sets. K cluster algorithmic technique helps in achieving the key value.

III. SPATIAL DATA MINING

Spatial data extraction has a specific attributes Structure Synopsis has infinite data sets which are represented in vector quantities like Geographical forecasting & Structure array. Geographical mining has wide applications. The associated functionalities of data mining are data, knowledge and information. The geographic phenomena data sets are not definitive in structure. The patterns are formed on a size basis. The size dependency is to group the data sets and to solve spatial mining problems. The Spatial data extraction process can be branched into data input, processing and output. The input data can be constructed in a table format governed by the representation of cartons spatial, where samples can be represented as table.1

Table-I: Input Representation based on cartons spatial

Sid	*A1*	*A1*	...	*An*

The *Sid* represents the sample id, *A1* represents the attributes of the sample which are geo referenced and non geo based samples and the relation between the geo synchronized and non geo synchronized data can be easily identified

Non Geo synchronized data sets are (a) similar to the conventional data extraction (b) arithmetic & segregation whereas the Geo Synchronized data sets are Structural attributes, Structural data illustration methods. The distance between the patterns can be found by various algorithmic techniques used to find the maximum span between the referential data patterns is by shortest path algorithm, negotiable circle center with reference points, voronoi spatial separation. The output of data representation is forecasted with periodic timeline interpreted data sets, cluster and pattern formation associated with tabular representation. For the interpretation the data sets feed should be of either steady state or dynamic responsive type. The various techniques and the associated key features are depicted in the table-2

Table-II: Patterns and their Statistical Analysis

Patterns	Point Process	Lattice	Geo spatial analysis
Forecast	✓	✓	
Bias			✓
Grouping	✓	✓	
Summary	✓	✓	✓
Alliance	✓	✓	

The representation of data sets and its associate functionalities in differentiation with traditional and spatial data extraction is depicted in the table no.3 given below.

Table –III: Forecasting Methodology

Traditional	Geo-location
$x = Y\alpha + \varepsilon$ $Ki(Bi\|Y) = \frac{Kr(Y\|Bi)Kr(Br)}{Kr(Y)}$	$x = \rho Wx + Y\chi + \varepsilon$ $Ki(Bi\|Y, Bn) = \frac{Kr(Bi) * Kr(Y, Bn\|B}{Kr(Y, Bn)}$

A. Data Visualization in geographical data extraction

Evolution of multi-dimensional data interpretations, Data visualization with a key value meaning to the observer is hypothetical. Geo-location depicts the key value on the patterns created by the data sets, which helps in data extraction key value search. The outcome of data extractions are reflected, governing the supplementary clues and appreciable investigation phenomena. This kind of creative pattern found to be effective for the application development. Data set analysis over statistical methodology represents the topological data extraction frequency exists over a period of time scale.

B. Data Sets Alliance

- *Geo-location Set Relations*

Geo-location based data set relations are the key parameter in spatial data mining and it is achieved based on the survey theory. The best survey theory is field network math which is otherwise known as Regional connectivity calculation (RCC). It holds the geo-location data sets and applicable field sets. The qualitative data set alliance are of complex volume and thus are interlinked to each other.

- *Directional Relativity*

Directional relativity is applicable to the data sets that are in symmetric in nature, multi-dimensional are represented by the centre point. The geo-location is classified into four with the reference with the directional pattern like North, West, south, East and centre. The data set with *z(c)r* represents the geometric point centre of object z and r.

- *Radius relativity*

Radius relativity is the proper means of volumetric measurement. Similar to directional relativity, centre point of the objects is the key element for radius relativity. Both z and r are geo-location centered and it is lesser than the value of *P* which is for the near field vectors and for the far field it is vice-versa.

- *Volumetric Relativity*

Assuming the geo-location data sets are calculative with S_n. The calculative structure of nth order portrays the dimensional structure size. The size S_2 defines the area and depicted by comparative symbols and they are further extensive based on the application.

- *Tool for Data Extraction*

The Geo-location data sets can be treated as normal data sets by means of preprocessing. The architecture has three elements: Knowledge and Data banks, Statistical analysis and visualization. The knowledge and data banks combines' three database systems 1) Geo-location Data set database lies on the cloud 2) Spatial knowledge extraction database with cluster algorithm 3) Non Geo-location based database integrated with Geo-location database system fine tuned by the statistical analysis. This enables the internal key relation between the data sets of Geo-synchronized and non geo-synchronized system In our project the sampling information of the soil are collected by the sensor which has the information of soil chemical contents , Biological data sets associated with geo location are displayed in a dashboard for the visualization.

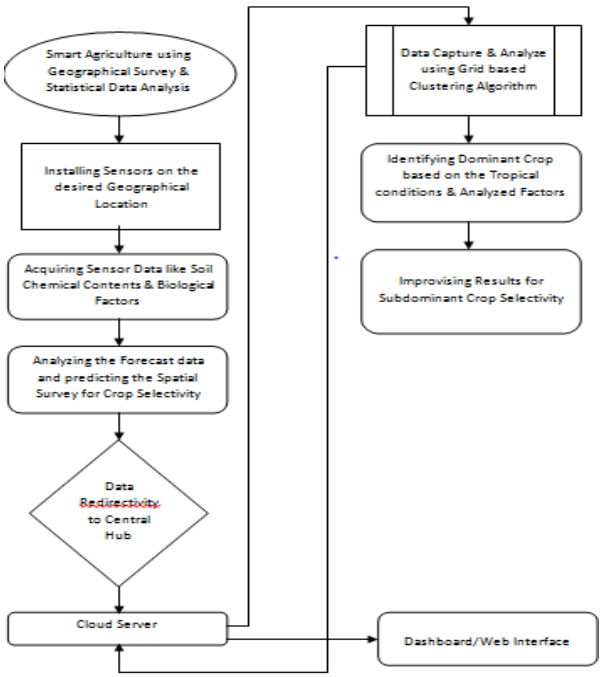

Fig 1: Proposed Methodology

IV PROPOSED METHODOLOGY

The Proposed Methodology for the spatial based Data extraction and forecasting methodology includes multiple integrations from the physical layer to the application layer. The whole ideology can be classified into three segments for the better understanding. It includes Physical Layer, Processing Layer and Application layer which is briefly explained below.

A. Physical Layer

The Physical layer includes the smart sensor connectivity for the Data extraction. The smart sensors are used to measure the soil chemical contents such as nutrient supplements for the soil growth which defines the crop selectivity at the preliminary stage. The nutrient supplements are classified into primary, secondary and tertiary supplements which specify the crop growth and productivity. The Primary supplements are nitrogen, Phosphorous, potassium content, secondary supplements are calcium, magnesium and tertiary supplements or micro nutrients are iron, zinc, copper, boron, molybdenum, chloride, manganese levels which are the key parameters for the crop selectivity and soil cultivation. The ambient sensors are also integrating along with the existing GPS & Soil moisture sensors in order to detect the light intensity levels at the particular crop cultivated location to monitor the ambient light source levels at the different climatic conditions. This enhances the irrigation system for agriculture on the timely basis which protects the crop from multiple instances of extinctions.

B. Processing Layer

The Processing layer collects the data from various sensors and stores in the cloud server. The cloud server is mostly an off the shelf server with multiple instance fast processing. The application layer is also integrated in the same cloud server for the fast responsive and to visualize the data in the graphical format. The data collectivity is by the database which is secure and further more the application and interlink layer between the physical layers and processing layer is done by the AES encryption system which beholds the data from antimalware systems. The data sets from the database are re-routed to the application interface. Before the data re-directivity the Grid based cluster algorithm is interfaced for the statistical analysis.

C. Application Layer

The application layer is interfaced with the dashboard functionality which the data sets after processing with algorithm, the data sets are grouped as pattern with the forecasted data sets which are visualized as graphical representation. The graphical representation consists of table sets and in the form of rectangular 2D plot system. The plots can be downloaded by the viewer and its open accessible by any user. This forecasted data results the Productivity ratio, integrated map functionality and with data assets such as timelines of crop growth, yields, dormant and manure time period, irrigation system and levels, irrigation timelines and other agriculture parameters which are useful for the farmer for the precise agriculture with greater efficiency. The dashboard is bootstrapped, so that the farmer can view from his android phones and in future scope app development will be taken into consideration to ease the website loading on the web browsers. The Application includes predictive analysis for the smart farming with maximum key added value information to the end user.

D. Grid Based Clustering Algorithm

Grouping the data based on their characteristics known as cluster and the objects which has same classification comes under similar with high priority and objects under different data sets are low prioritized in classification. Grouping yields patterns formation on their knowledge phase and by omitting the dissimilar objects. If the value of n is greater, the cluster formation will be more and it affects in omitting the effective data sets as aberration and it make to lose the useful information which results in improper pattern formation and result estimation. If the value of n is less, the group formation will be small in size and it adds the external noise elements which affects the accuracy and efficiency of the prediction methodology. Therefore to avoid this situation, the value of n is set to 2 in which the group formation results of group-1 results better results than the group-2. The distance between the data sets is considered to be Euclidean measurement which can be analyzed by the clustering algorithm. Weather Prediction & Soil parameter estimation for crop selectivity are briefly explained with the algorithm in the below segments.

E. Weather Prediction

Advance Weather prediction methodology over a period of timeline between the crop cultivation to the crop harvesting period plays a key role in the crop maintenance and productivity ratio. The Indian metrology departments predicts the weather conditions and it is also lively available with the Google analytics can be easily integrated to the application by a simple API. The data sets from the metrology department is fine tuned with various uncertainty conditions by the algorithm.

The Predicted weather data sets are considered as matrix element which can be represented as $W_{j,k}$ obtained from the metrology department with j represents the month and k represents the year. In similar, predicted weather data by the accuweather from the online interface can be represented as $X_{j,k}$. With these two data matrix sets $W_{j,k}$ & $X_{j,k}$ are analyzed under different statistical analysis. The data sets are tried with different combinations assuming the values for j & k. An additional parameter along with month and year, a prediction parameter p is also taken into consideration. The prediction parameter is assumed to P values which is the maximum combination of the data sets. From the various combinations, the values can be equated. For example, the value of p=2 defines W_{jk}, X_{1k} can be defined from the table set which are analyzed under different permutations and combinations. The key is to predict the data for p for neighbor year close to the target value of P.

S.No	F_{jk} Values
1	W_{jk}
2	W_{jk}, X_{1k}
3	W_{jk}, $k
4	W_{1k}, X_{jk}, $k
5	W_{jk}, X_{jk}, $k
6	W_{j-1k}, W_{jk}, W_{j+1k}
7	W_{j-1k}, W_{jk}, W_{j+1k}, P_{jk}
8	X_{jk}
9	X_{jk}, $k
10	X_{j-1k}, X_{jk}, X_{j+1k}

The data set after analysis is a duplication of the weather predicted in the past. The combinational prediction data sets finds span between the destined year and the other parameters can be calculated from the formula given below.

$$S_{jk} = \sqrt{\sum_{x=1}^{X} (W_{jkp} - W_{ixq})^2, \$k \neq x}$$

F. Soil Parameter Estimation

The soil parameter can be estimated by various tools which are developed in the recent times especially in dealing with Crop selectivity and soil nature for multiple crop cultivation technique. On analysis, the data can be transformed into patterns which are used to find the span between the pattern sets. If the pattern sets are in the range of 25 & 75 defines the soil has 25% clay content and 75% enriched soil content. The Soil parameters should be more in terms of representation to define the soil contents as well other defining the limitations of usage for other parameters. This can be represented as,

$$S_j = \sqrt{\sum_{k=1}^{m} S_k (X_{jk} - X_{ij})^2}$$

The Span between the patterns can be calculated by

$$\dot{S} = \sum_{k=1}^{K} X_k Y_k$$

Where X_k defines the weight of the neighboring soil in relating with the target. The weights of the patterns are associated with the distance in which the whole analysis can be done on the Matlab with the algorithm to analyze the soil parameters and weather prediction analysis

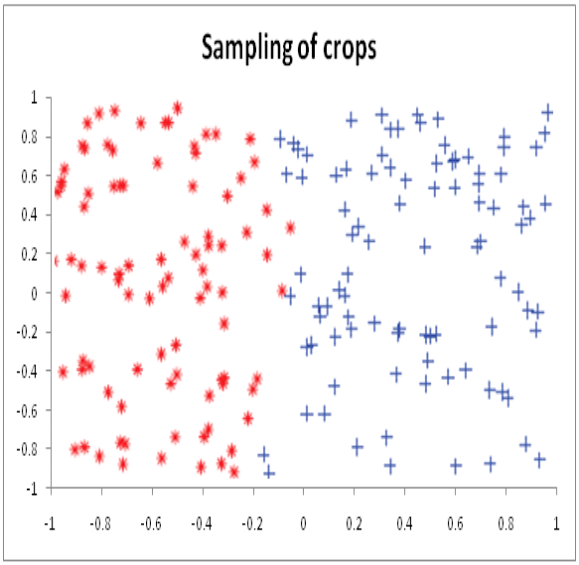

Fig 2. Sampling of crops based on the basic parameters from the data sets

```
% checking the SPAN
for kk = 1:m,
for ii = ktrain-1:-1:kk,
if dist(ii) > dist(ii+1),
res = dist(ii); dist(ii) = dist(ii+1); dist(ii+1) = res;
res = ind(jj); ind(jj) = ind(ii+1); ind(ii+1) = res;
end
end
```

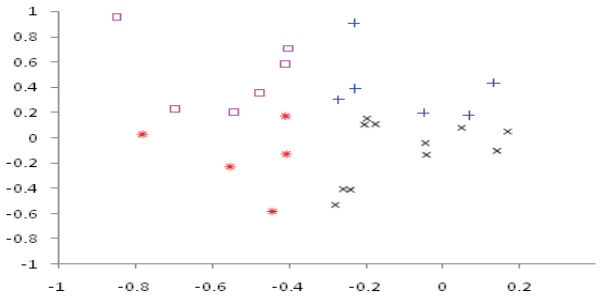

Fig 3. Analyzing the span between the data sets

```
% unknown sample data set
for i = 1:m,
score(i) = 0;
end
for clu = 1:min(k,ntrain),
score(gtrain(ind(clu))) = score(gtrain(ind(ii))) + 1;
end
```

for i = 2:m,
if score(i) > res,
end
end

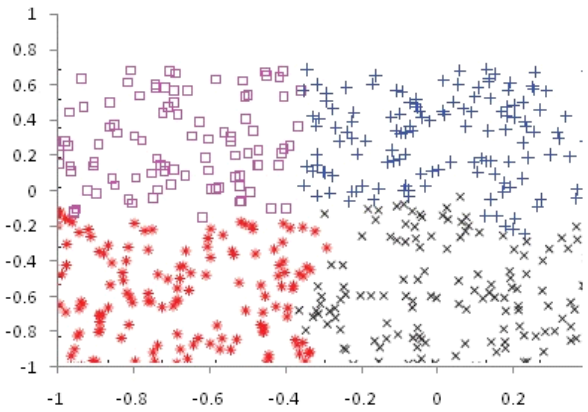

Fig 4. Identification and comparative analysis of unknown samples along with the neighborhood data sets

For identification of unknown samples the values assigned to each group and the maximum value of the each group object is depicted as object for the unknown sample sets. At the

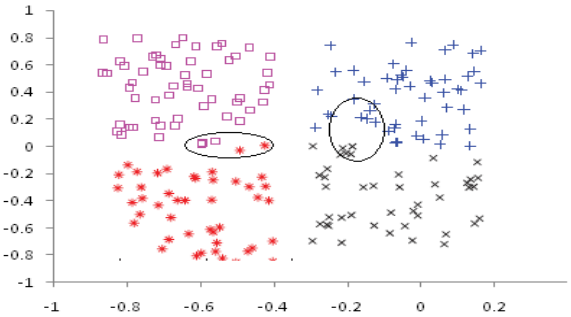

Fig 5 Cluster formation and pattern recognition with key elements

Preliminary stage all the object values are initialized to zero and then it is updated with the set of instruction sets. The clustered data sets can be depicted in the figure 5 in which the noise figure is eliminated and the data sets are grouped with the key elements. The value of *eps* is varied in accordance with the number of data taken into consideration for the cluster formation.

V. RESULTS & DISCUSSION

The Experimental evaluation done in the region of vellore district and chosen an agriculture land in anaicut taluk. This agriculture land mostly cultivates the rice & paddy crop on alternate basis. But during the experimentation time, paddy was cultivated in which it has been taken as a test bed for the comparative analysis. These data are analyzed with the backend algorithm and forecasted on the dashboard. The dashboard is created using wordpress and it is hosted on the cloud server. For instance the code has been executed from the local machine for the trial execution. The forecasted data from the tamilnadu agriculture department on crops yielding ratio along with the soil wise for each and every district which helped us knowing the basic key elements for the rice crop and in future it could help the farmers in cultivating the rice crop in their agriculture land. The district wise data sets are collected from [26] depicted in the below figure1.

Table: District-wise soil fertility status in Tamil Nadu state

S. No.	District	Rice area (x 1000 ha)	Rice yield (t/ha)	Avail N (kg/ha)	Avail. P (kg/ha)	Avail. K (kg/ha)	Zn	Cu	Fe	Mn
1	Coimbatore	7.41	2.79	205	52	340	43.5	36.6	34.7	5.9
2	Cuddlore	114.29	1.91	186	52	223	NA	NA	NA	NA
3	Kanchipuram	114.72	2.56	190	45	242	NA	NA	NA	NA
4	Kanyakumari	21.71	3.80	280	29	164	10	70	16.6	3.4
5	Madurai	71.00	3.30	224	52	346	32.9	71	5.1	15.1
7	Salem	37.33	3.68	238	41	340	34.8	15	3.9	6.4
8	Thanjavur	154.90	2.58	210	47	346	81	18	12.9	8.3
9	Tirucharapalli	79.58	2.98	242	32	291	33	21	4.4	10.1
10	Tiruvenelli	86.40	3.58	205	53	359	19.2	94	24.1	4.1
11	vellore	58.16	2.88	212	55	334	NA	NA	NA	NA
12	Dharampuri	29.02	2.94	234	39	354	8.6	16	9.2	6.1
13	Pudukkottai	95.99	2.42	-	-	-	33	21	4.4	10.1
14	Nilgiris	1.43	3.30	202	13	256	19.8	94	24.1	4.1
15	Rama' puram	127.40	1.94	212	10	290	35.9	6	16.2	6.4

Fig 6. Soil Content Data set from Agriculture Department- Districwise

These data sets are used as a data set matrix for the algorithm input and analyzed at the backend with clustering algorithm. The dashboard is developed on the Word Press platform and the API interface is done to observe the Soil moisture level, humidity, temperature and chemical content levels are displayed as a live feed at the dashboard and the google maps is integrated to show the location of agriculture filed which the data is been acquired. The website is hosted on the WordPress free online platform & the data base is interlinked with the amazon service for fast processing. https://Smartpfarm.wordpress.com is the hosted web-url link for the test run and a copy of it is shown in the figure given below,

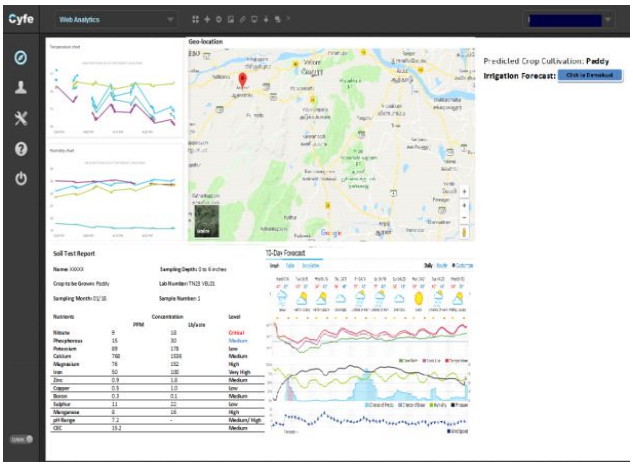

Fig 7. Dashboard of the Website Created

The data sets and the prediction methodology collects the datasets and analyses the data with the algorithm and data sets. The above mentioned soil content forecast method and weather prediction analogy data sets are converted into the graphical interactions to the user to determine the various levels.

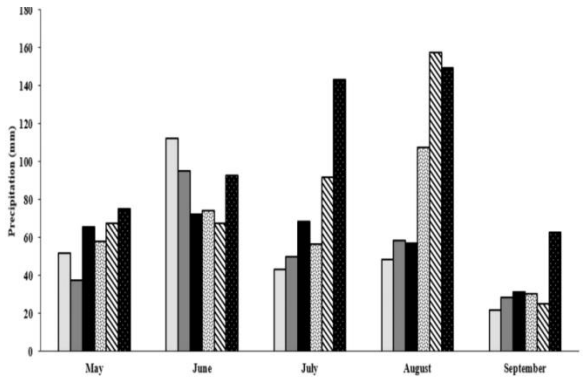

Fig 9: Precipitation prediction for a period of 4 Months

This chart depicts the various levels of precipitation levels forecasted for every 5 days in the each and every moth. The level shows low, medium, high, very high and critical levels of precipitation formation. For the limitations, the date of the month is not mentioned in the depicted image.

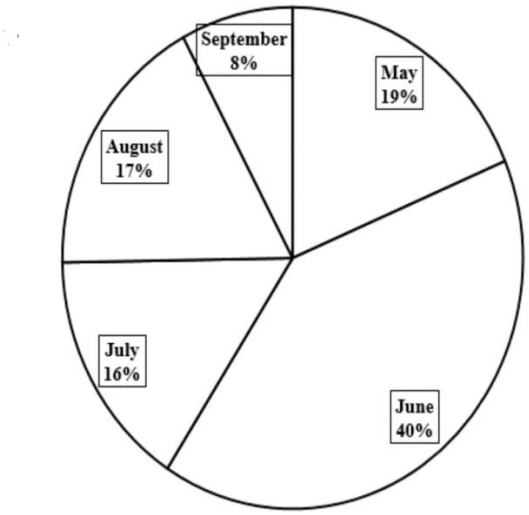

Fig 8: Growth cycle

The growth ratio is depicted in the form of chart with results the duration period and the crop growth. The crop growth along with the timelines is directly proportional to the productivity ratio. Thus the crop is cultivated during the period of may and it shows the massive growth in the month of June. The first growth of the crop is almost 40% fast growth as per the data sets predicted and it has been followed by the farmers on the crop growing period. The Forecasted data sets helps in improvising the growth level of about 10% higher than the traditional methodology within a month span of time. The rest time period is for the manure & defoliant usage.

VI. CONCLUSION

Therefore the predicted data sets shows effective in the productivity and crop protection than the traditional methodology. The traditional methodology does not help farmers on the crop protection from the external natural events, climatic change and other forms of crop destructions. This forecast method helps to increase the yield 10-15% higher than the traditional methodology and also results farmers with the sign of warning from natural climatic events. This method of agriculture make the less usage of other resources like manure and defoliants which makes the crops to grow naturally with the green farming and not more the chemical crops which are harmful and results health hazards. The cluster algorithm finds application in predicting the soil and helps in crop selectivity for the particular soil. Multiple crops can be cultivated in a single cultivation space. Mostly the dual crop cultivation occurs in the agricultural land by cultivating the secondary crop on the corners of the main crop cultivation. This makes the usage of extra manure and defoliants by the secondary crops. This helps in utilizing the agriculture space 80-90%effectively. The algorithm is more accurate in predicting the soil for the crop and also helps in finding solution for the maximum utilization with higher yield and crop protection.

REFERENCES

1. Ramesh Babu Palepu1, Rajesh Reddy Muley, An Analysis of Agricultural Soils by using Data Mining Techniques, International Journal of Engineering Science and Computing, October 2017, pp 15167-15177.
2. Ester M., Kriegel H.-P., Sander J.: Spatial Data Mining: A Database Approach, Proc. 5th Int. Symposium on Large Spatial Databases (SSD'97), Berlin, Germany, 1997, pp. 47-66.
3. Ramesh A.Medar and Vijay.S.Rajpurohit, "A survey on Data Mining Techniques for Crop Yield Prediction", international journal of advance research in computer science and management studies, Vol.2, Issue.9, pp. 59-64, Sep 2014.
4. M.C.S.Geetha, "A Survey on Data Mining Techniques in Agriculture", International journal of innovative research in computer and communication engineering, Vol.3, Issue.2, pp. 887-892, Feb 2015.
5. Ch.N.Santhosh Kumar, V. Sitha Ramulu, K.Sudheer Reddy, Suresh Kotha and Ch. Mohan Kumar, "Spatial Data Mining using Cluster Analysis", International journal of computer science and information technology, Vol.4, No.4, pp. 71-77, Aug 2012.
6. Raorane A.A. and Kulkarni R.V., "Review- Role of Data Mining in Agriculture", International journal of computer science and information technologies, vol.4, No.2, pp. 270-272, 2013.
7. Khan Mohammad A.,Md. Zahid-ul Islam and Mohsin Hafeez (2011). Evaluating the Performance of Several Data Mining Methods for Predicting Irrigation Water requirement. Proceedings of the Tenth Australasian Data Mining conference(AusDM2012),Sydney, Australia.199-207.
8. Mucherino et al., Data Mining in Agriculture, Springer Optimization and Its Applications 34, 47, Springer Science
9. Ramesh and K. Ramr Classification of agricultural land soils: A data mining approach International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011379
10. Rainfall variability analysis and its impact on crop productivity Indian agriculture research journal 2002 29,33.,8) SPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST by Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B
11. D.Rajesh, ―Application of Spatial Data mining in Agriculture‖ International Journal of Computer Applications, Volume 15, February 2011.
12. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 2, February 2012) 275 ―Survey on Data Mining ―VibhaMaduskar and Prof. yashovardhankelkar
13. Amrender Kumar, weather based crop forecasting techniques, Agricultural Statistics Research Institute.
14. Anwiti Jain, Anad Rajavat, Rupali Bhartiya ―Design, Analysis and Implementation of Modified K-Mean Algorithm for Large Data-Set to Increase Scalability and Efficiency, IEEE Explore December 2012
15. A Ragel, B Crémilleux -Treatment of missing values for association rules, Pacific-Asia Conference on Knowledge Discovery, 1998

16. E.I. Papageorgiou, A.T. Markinos and T.A. Gemtos, "Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application", Vol.11, pp. 3643-3657, 2011.
17. Luke Bornn and James V. Zidek, "Efficient stabilization of crop yield prediction in the Canadian Prairies", International journal of agricultural and forest meteorology, Vol.152, pp. 223-232, 2012.
18. Dimitrios Voloudakis, Andreas Karamanos, Garifalia Economou, Dionissios Kalivas, Petros Vahamidis, Vasilios Kotoulas, John Kapsomenakis and Christos Zerefos, "Prediction of climate change impacts on cotton yields in Greece under eight climatic models using the AquaCrop crop simulation model and discriminant function analysis." Agricultural and water Management, Vol.147, pp. 116-128, 2015.
19. Raymond T. Ng and Jiawei Han, Member, CLARANS: A Method for Clustering Objects for Spatial Data Mining Ieee transaction on knowledge and data engineering, vol. 14,
20. Majumdar et al.- Analysis of agriculture data using data mining techniques: application of big data, Journal of big data, DOI 10.1186/s40537-017-0077-4
21. Kriegel H.-P., Pfeifle M., Schönauer S.: Similarity Search in Biological and Engineering Databases, IEEE Data Engineering Bulletin, Vol. 27, No. 4, 2004, pp. 37-44.
22. Anselin, L. 1995. Local indicators of spatial association — LISA. Geographical Analysis 27: 93–115.
23. H. Miller and J. Han(eds), Geographic Data Mining and Knowledge Discovery, Taylor and Francis, 2001
24. Openshaw, S., 1994, Two exploratory space-time-attribute pattern analysers relevant to GIS. In A. S. Fotheringham and P. A. Rogerson (eds.), Spatial Analysis and GIS,(London: Taylor and Francis), pp. 83-104
25. Ester M., Kriegel H.-P., Sander J.: Spatial Data Mining: A Database Approach, Proc. 5th Int. Symposium on Large Spatial Databases (SSD'97), Berlin, Germany, 1997, pp. 47-66.
26. Soil information system: http://14.139.94.101/fertimeter/DistTN.aspx

AUTHORS PROFILE

Prathik. A is a Research Associate in School of Information Technology and Engineering, VIT University in Vellore, Tamil Nadu. He received M.Tech in the field of Information Technology-Networking from VIT University,Vellore in the year 2015.He is currently doing PhD in VIT University, Vellore, Tamil Nadu, India.His research interest includes Spatial Data Mining, communication networks, coding theory, algorithms, computation, operations research, circuit design, and communication network addressing and data base management.

Dr Anuradha. J is working as Associate Professor in School of Computer Science and Engineering, VIT University in Vellore, Tamil Nadu. She has more than 15+Years of Teaching Experience. She is currently guiding 6 PhD scholars. Her research interest includes Machine Learning, Classification, Computational Intelligence,Spatial Data Mining, and coding theory, Soft Computing, Clustering Algorithm, Rough Set Theory, Feature Selection and Algorithms.

Dr Uma. K is working as Associate Professor in School of Advanced Sciences, VIT University in Vellore, Tamil Nadu. She has more than 20+ years of teaching Experience. She is currently guiding PhD scholars. Her research interest includes Engineering, Applied and Computational Mathematics,Mathematical Analysis, Numerical Mathematics, Real and Complex Analysis, Measure Theory, Harmonic Analysis, Constructive Approximation.