International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I1P224
Volume 69 Issue 1, 152-165, January 2021

Data Mining of Paddy Cultivation Patterns And Water Resource Management In Late Samba Season of Tamilnadu

B.Vinoth, M.C.A., M.Phil., (Ph.D.), Dr. N.M.Elango

[1]*Research Scholar.* [2]*Associate professor,*
[1]*Sathyabama Institute of Science and Technology (Deemed to be University)*
[2]*VIT University, Vellore*

vinothkarthick84@gmail.com

I. ABSTRACT

The Earth is borne to influence its water contents mostly of its Biosphere, the living organisms, habitat, birds, men, and lives due to fluctuating parameters of Earth, and it has 71 % water and 29 % land. When the Sunlight shine-up, the water available from the surface of Oceans, Rivers, and ponds get transformed to vapor and goes up to Space in the form of clouds. The clouds are called Cumulonimbus, Nimbus, Cirrus, etc. On this Earth, among the countries of the Asian continent, India has the largest population for its credibility. Agriculture is the main backbone of India, and especially the state of Tamilnadu is the second largest food producer of India. Amount various paddy cultivable seasons, and Late Samba is most vulnerable for its yield towards food production. As yield is the main target of farmers, this paper addresses exhuming hidden paddy cultivable patterns, which optimizes the season scenario.

Keywords: *Data mining, Late Samba season, WEKA tool, Modified Assignment problem.*

II. INTRODUCTION

Agriculture is the Primary Moving Force (Primary Sector) of the Indian Economy. In India, 49 % of the people population participate in Agriculture. Agriculture is bred as a private sector industry in India. Many Land Reform measures being brought to force have changed the trend of agriculture and the lifestyle of primitive artisans. If a Farmer has a small piece of land, introducing Modern measures of machinery, e.g., Tilling Machines, Power tillers, Tractors, Harvest machines, Crop Plantation Machine more yield could be realized. By improving the Economic strata of the Agricultural industry, the industry in India addresses the domestic requirement and also contributes hugely to the country's exchequer by way of foreign economy.

In the early 1960s, the advent of the Green Revolution and its measures have changed the scenario of Agriculture in India drastically, with which a huge amount of Wheat and Rice were produced, and that tune marked the surplus state of Agricultural commodities. The Scientist Norman Borlaug doing extensive research in the 1960s has identified key factors which can impact the state of Agricultural resource. Factors such as yield variety, chemical fertilizers, irrigation patterns, chemical pesticides, germicides, herbicides and Credit Storage, Marketing / Distributions, etc., are the viable entities.

Revolution: 1. Socio-Economic Impact, Ecological impact, Critical Ecological Crisis, Toxic Level in Food Chain. The Market Economy is based on various factors, First Supply and Demand. The market fluctuations, Inflation, Stagflation, Rising factors of exports are totally dependent on the course of Nature conditions.

III. EXPLANATION

The Paddy Seeds of Data are abstract from all the districts of Tamil Nadu. These are sowing in the good lands with irrigation facilities or rainwater with Natural Fertilizer. We should plough the Lands before we are sowing the Seeds. For Growth, the Seeds need Water, Land, Warmth, Air, and fertilizer. Photosynthesis is responsible for food production in plants, and their yield is the inferred turn up from the plants. The yields are not in constant factor, some seasonal changes like Flood, Cyclone, and Famine causes damage to the crops. Beyond, the Farmer gets earnings. The Farmer's life is in a vicious circle of poverty. Hence he wants some scientific Techniques in Data Mining and modified assignment problems to get a high yield in paddy cultivation. There are nine paddy cultivable seasons in Tamil Nadu, from which the Late Samba season is taken into consideration. Table 2 shows the Late Samba season details.

RESOURCE OF TAMILNADU:

Physiographic: South India, being bounded mostly by coastal coverage and one side by soil land, is so popular for agrarian cultivation and popularly known to be the Deccan plateau in which Tamilnadu is so vulnerable with its delta soil. Adding nurture to this factor of the delta, the river

Cauvery with its alluvial irrigation pattern is so viable for paddy cultivation, and hence Thanjavur district and others so selected here for the analysis are producing more food grains from time immemorial.

Rainfall: Tamilnadu state receives most of its water from North-East monsoon rains. The South-West wind also contributes a little to the western districts of the state beyond the water catchment received during hot weather season due to summer rains. The Mean Annual Precipitation of rain is 900 mm in the plains; in Coastal areas, it is more than 1000 mm, and with the hill ranges of the Nilgiris, it would be of the tune of 1666mm. The soil moisture regime of coastal plains and hills is finely humid for the paddy cultivation pattern of life.

SOIL:

The soil has the characteristic of contribution as to contain represent in the form of Acid, Base, and Salt, represented in the form of PH [12]. The PH value of 0 to 6 ranges is of Acidity. PH value 7 is neutral (Base). The PH value 8-14 is salty (alkalinity). PH value 7 is normal and ideal for plant cultivation. The air content of the soil is vital. It is helpful in the process of oxidation, which involves nitrogen fixation too. Soil Water depending upon the texture of the soil, water moves downward by percolation; the size of soil is between 0.02 and 0.002 mm. The structure of the soil is in the form of sand, silt, clay, and humus link together. The waterlogging capacity is differing depending upon the types of soil. The problems concerning the Indian soil types as to Soil erosion, the declining pattern of the fertility of the soil, salinity, and alkalinity, waterlogging, and desertification are taken into consideration. Table 1 shows the properties of soil.

Table 1 shows the properties of the soil of Tamil Nadu

Sl.No	Soil Property	Black soil	Red soil	Alluvial soil	Laterite soil
1	Clay (%)	35.5	18.2	34.5	43.6
2	Silt (%)	13.8	6.1	15.8	5.6
3	Fine sand (%)	19.6	23.6	19.6	19.0
4	Coarse sand(%)	24.7	50.7	28.0	29.2
5	Iron oxide (Fe_2O_3)	3.9	6.0	3.5	10.6
6	Organic carbon (%)	0.3	0.1	0.6	2.3
7	Available nitrogen (ppm)	48.0	85.0	119.9	139.0
8	Available Phosphosrus (ppm)	10.0	12.0	35.0	25.0
9	Water soluble potassium (ppm)	9.0	12.5	10.0	50.5
10	C.E.C meq/100 gm	54.2	11.7	16.9	8.5
11	Base saturation (%)	94.3	93.2	97.4	65.6
12	PH	7.2	6.6	7.6	4.5
13	E.C. (mmhos/cm)	0.2	0.2	0.4	0.2
14	Exchasngeablepotassium (meq/100gm	0.4	0.1	0.1	0.3
15	Exchangeable calcium (meq/100 gm)	22.8	5.6	10.6	3.9
16	Calcium carbonate (%)	1.1	-	0.8	-

FERTILIZER:

Is water is only needed for the crop plantation? No, the ideal land for cultivation, the optimal seed for germination, natural fertilizer for nitrification, and organic farming pattern for ideal yield output. While plough or plow, the multi-grain based nutrients are needed to the plants. Artificial fertilizers, poison pesticides de-fertilize the soil. The soil gets revived due to cattle farming and exposure to direct sunlight. The poisonous pesticides destroy the good pests of the land. The Earthworm is also destroyed; it is a friend to the Farmer. The Vermin is added to soil contents just for establishing air ventilation in the long run.

The Rats also damage the crops. The snakes, owls, Kooten eat the rats. Thus there is a balancing pattern in nature itself for complying with food needs in the whole of nature and its species diversification. Nutrients to crop patterns can be classified as 1. Leaf nutrients 2. Gray nutrients 3. Sulfate nutrients 4. Nitrate nutrients 5. Gypsum 6. Zinc 7. Rhizobium. The climate, warmth are needed for photosynthesis. Due to photosynthesis, the crops can grasp the water, nutrients, and deliberate the CO2, O2 during the process. In Tamilnadu, agrarian people follow both the traditional pattern of farming (organic farming) and the advanced agri-tech based method of farming.

IV. Related Work

Tamilnadu adopts many seasonal paddy cultivating patterns, of which, Late Samba season with its seed varieties, total area irrigation, water resources, farmers' investments, and total harvesting days. All these factors are taken into consideration and analysis of data and lead to finding hidden value, optimal solution and that of the decision, the prediction is going to be used for implementation leading to future enhancement.

Farmers have the tendency to choose seed variety based on prevailing climatic conditions and availability of water resources. To characterise seed selection pattern, 15 seed varieties viz ADT38, ADT39, TRY 1*, TRY 3*, ADT (R) 46, ADT (R) 49, CO (R) 48, CO (R) 49, CO (R) 50, White Ponni, ASD 19, TPS 2, TPS 3, ASD 18, MDU 5 have been chosen. Viable areas of cultivation have also been featured, precisely by means of districts of paddy cultivation viz Thanjavur, Nagapattinam, Tiruvarur, Tiruchirapalli, Kanchipuram, Tiruvallur, Coimbatore, Tirupur, Erode, Karur, Perambalur, Ariyalur, Kanyakumari, Pudukottai, Madurai, Dindigul, Theni, Tirunelveli, Thoothukudi all from Tamilnadu state are considered has been shown in table 2.

Table 2 shows the Late Samba season details

MONTH OF SOWING	SEASON		DURATION OF GROWTH / RIPING (in days)			DISTRICTS									
DEC - JAN	LATE SAMBA		<135			Thanjavur, Nagapattinam, Tiruvarur, Tiruchirapalli, Kanchipuram, Tiruvallur, Coimbatore, Tirupur, Erode, karur, Perambalur, Ariyalur, Kanyakumari, Pudukottai, Madurai, Dindigul, Theni, Tirunelveli, Thoothukudi.									
SEEDS OF THE SEASON	ADT 38	ADT 39	TRY 1	TRY 3	ADT (R) 46	ADT (R) 49	CO (R) 48	CO (R) 49	CO (R) 50	White Ponni	ASD 19	TPS 2	TPS 3	ASD 18	MDU 5

Rice cultivation is the major industry of Tamilnadu state, and especially medium and long span varieties are mostly chosen by the farmers due to their optimal yield outcome. Late Samba season is often proliferated by most of the farmers belonging to 19 major districts of Tamilnadu. Farmers get irrigation for their fields by way of rivers, canals, bore wells, etc.; the winter season is ideal for plant crops in terms of Growth and pollination. Farmers are belittled due to seasonal changes, irrigational scarcity, and pest impacts. So there is a need to analyze cultivating routines and to address sufferings and hardships faced by farmers by using data mining techniques, which would certainly tell upon the standards of farmers.

Table 3 shows the investment for hector in agriculture

1 HECTOR LAND CULTIVATION COST FOR PADDY	
SEED RATE	RS. 2800
LAND PLUGING FOR SEED SOWING	Rs. 600
Fertilizer for seed Sowing	Rs. 150
Land plugging for using a tractor	Rs. 4000
Machine Farm implements	Rs. 6600
Hand Weeding	Rs. 2500
Hand Weeding	Rs. 2500
First Fertilizer	Rs.2000
Second Fertilizer	Rs.1500
Third Fertilizer	Rs.2000
Fourth Fertilizer	Rs.1000
Fifth Fertilizer	Rs.1500
First Insects Fertilizer	Rs. 1500

Second insects fertilizer	Rs. 1500
Third insects fertilizer	Rs.1000
Machine Harvesting	Rs.3000
Labor Cost	Rs.4200
Total	Rs.43700

Often every domain of operation is analyzed for its data availability and manipulative operations due to data mining and automated learning patterns. So also, the industry of agriculture gets a booming boost due to data analysis and implication of results into the field. Data collected from the field, cleansed, and made to be applied to data mining engines and algorithms would certainly augment the standards of poor-men farmers who are deprived of such analyses otherwise.

The following table has data pertaining to late samba season in which 'T' represents a "True" state indicating a preferred variety of seed by the districts captioned, and 'F' represents the case wherein no such selection is made.

Table 4: shows the data collection for Late Samba Season.

LATE SAMBA	Thanjavur	Nagapattinam	Tiruvarur	Tiruchirapalli	Kanchipuram	Tiruvallur	Coimbatore	Tirupur	Erode	karur	Perambalur	Ariyalur	Kanyakumari	Pudukottai	Madurai	Dindigul	Theni	Tirunelveli	Thoothukudi
ADT 38	T	T	T	F	F	F	T	T	F	T	T	T	F	T	T	T	T	F	F
ADT 39	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
TRY 1*	T	T	T	T	T	T	F	F	T	T	T	T	T	T	F	F	F	T	T
TRY 3*	T	T	T	T	T	T	F	F	F	F	F	F	F	F	F	F	F	F	F
ADT (R) 46	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
ADT (R) 49	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
CO (R) 48	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
CO (R) 49	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
CO (R) 50	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
I.White Ponni	F	F	F	F	T	T	F	F	F	T	T	T	T	F	T	T	T	T	T
ASD 19	F	F	F	F	F	F	F	F	F	T	T	T	T	T	F	F	F	F	F
TPS 2	F	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
TPS 3	F	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
ASD 18	F	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
MDU 5	F	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F

V. REVIEW OF LITERATURE

A range of Data mining techniques is quite applied to the agricultural Farmer's domain to characterize farmers working patterns, which often lead to viable and vulnerable results. All such works are cited in the reference literature. To quote some of them here, enlisted are Journal of "Application of Data Mining Techniques for Prediction of Crop Production in India"[1], "A Review on Data Mining Techniques for Fertilizer Recommendation"[2], "New Proposed Method for Solving Assignment Problem and Comparative Study with the Existing Methods" [3]. All these works pivotally

consider the application of data mining techniques for getting more yield of paddy by farmers.

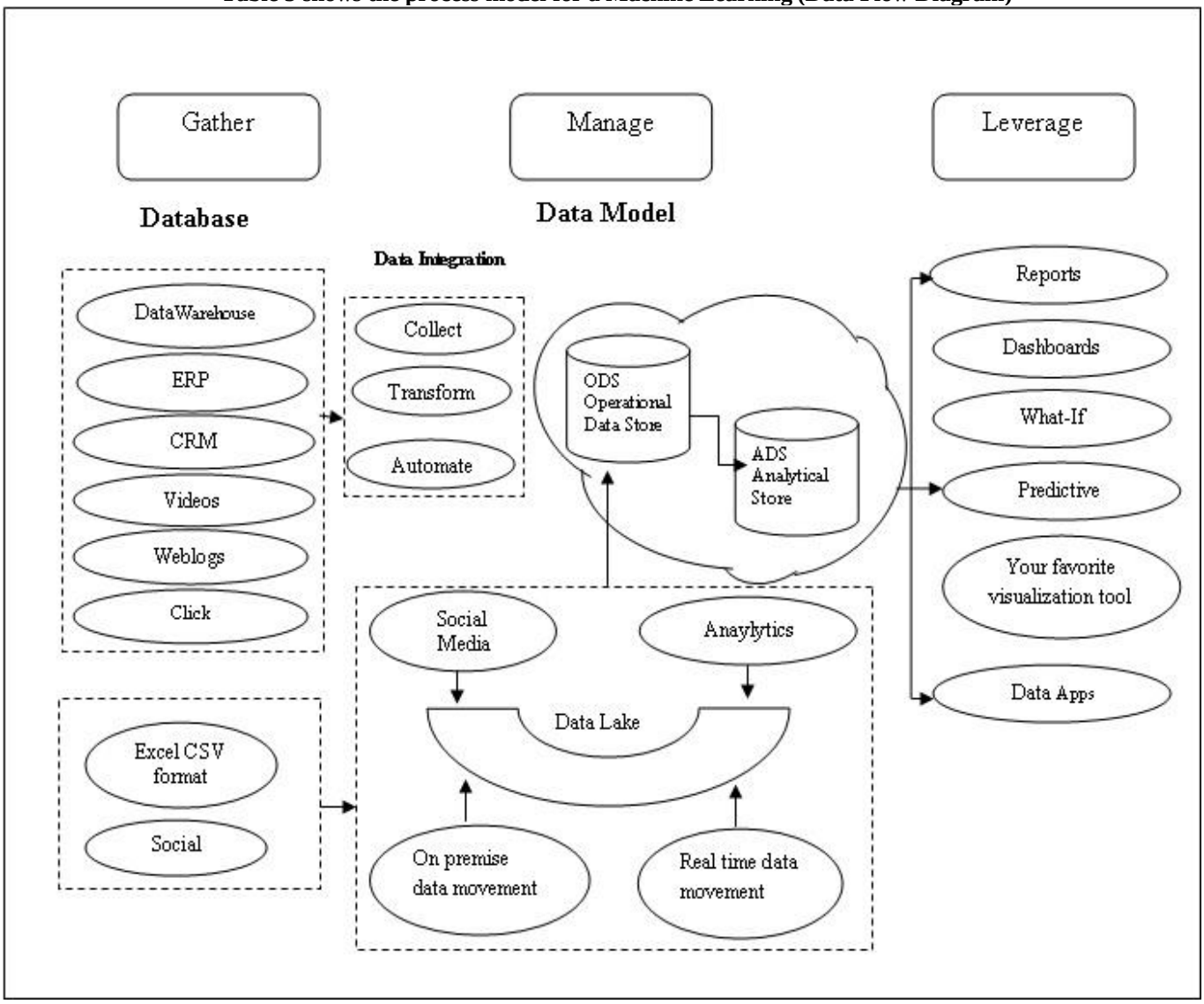
VI. Methods

Data mining Techniques:

Agriculture is often a dependent industry on parameters such as rainfall, temperature, and humid conditions, PH of soil, and impact of fertilizer being used. Dating back to 16 BC, India has contributed to 99% of world food production level, and now a small decline to this has made to set India to occupy 2nd largest position in the world.

A work [5] states that this production rate would further decline down by 23% due to some biological factors. Data mining is a current research trend to address any sector, including agriculture, to exhume viable research results [4]. Data mining proves to be potential in evolving patterns by discovering large databases has being depicted in the table5.

Table 5 shows the process model for a Machine Learning (Data Flow Diagram)



Agriculture, being the only industry in existence during time immemorial, the socio-economic strata of Farmer's family had a high impact from the ups and downs of the agricultural industry, and their pattern parameters such as atmospheric condition, people's cultural values, and technology being employed for breeding agriculture. Though the monoculture pattern of agriculture has dominated the industry for large scale grain production, usage of fertilizer and pesticides have curbed the Growth of pests soil gets polluted due to such usage, what lost is our traditional pattern of breeding agriculture and our ancient seed varieties [7]. To fill up the

above-cited depletion, more yield obtained adopting new methods in agriculture and supplementing this field with the results from Data analytics due to Data mining techniques, have a sheer impact in the field and tell upon the standards and living pattern of farmers involved in agriculture. Here data is analyzed due to classification and clustering [6].

Classification techniques lead to the design of a decision support system for farmers to make them move forward in agriculture.

A. TECHNIQUES

The most viable technique of data mining viz apriori method is used in the earmarked domain of agriculture for analysis frequency seed patterns, which have compatibility of sowing in the fields of the selected districts confronting climatic conditions too. Here optimal-yield-giving seed varieties are chosen by the said method.

B. Formulation of Apriori algorithm

In the Late Samba season, it is more important for farmers to augment agriculture routines with consideration of climatic conditions, scarcity of water resources, and to give out good yield for tuning agro income. The apriori algorithm is the well-known algorithm based on association rule mining, which would produce definitive results by exploring (K + 1) – item set and then the method iterates to find frequent K- item sets that meet the minimum support level along with factors of Support, Confidence, and Lift which are given below (1 - 3) [8].

$$\text{Support}(A, B) = P(A \cup B) = \frac{\Sigma \text{ The Processing Contains of A and B}}{\Sigma \text{The Total Processing}} \quad (1)$$

$$\text{Confidence}(A \rightarrow B) = P(A \mid B) = \frac{\Sigma \text{ The processing Contains A and}}{\Sigma \text{The Total Processing contains A}} \quad (2)$$

$$\text{Lift}(A \rightarrow B) = \frac{\text{SUPPORT } (A \cap B)}{\text{SUPPORT } (A)\text{SUPPORT } (B)} \quad (3)$$

Shows the General Apriori Algorithm

T: Transactional database
C_k: Candidate item set of size k
L_k: Frequent item set of size k
s: Support
Apriori (T, s)
L1←
{Large 1 – item set that appears in more than an equal to s Transactions}
K← 2
While $L_{K-1} = \emptyset$
C_K← Join (L_{K-1})
For Each Transaction t in T
For each candidate c in C_K
If(c c t) then count [c] ←Count [c]+I
End for
EndFor
$L_{K-1} = \emptyset$
For each candidate c in Ck // Prune
If (count[c] >= s) then
L_k ←(c) U{c}
End If
LK =?
End For
k← K+1
End While
Return L_K
End Apriori

C. J48 Algorithm:

The simple algorithm for decision support due to classification is by means of J48. This discriminates the decision variable attribute from others clearly and classifies data instances. This method is highly notorious for its precise prediction through classification. Here the method is capable of classifying all data instances by using predictive measures through machine learning way. The targeted new data and values are subjected to a classification process through a decision tree approach in which the internal nodes are denoted by dissimilar attributes, and all the possible data are tested and are denoted as branches. Finally, the terminal nodes are nominated by the dependent variable, and thus the prediction is ascertained [10]

D. Naive Bayes Algorithm:

This algorithm based on the Bayesian rule on conditional probability on the dataset attributes defines decision making in a simpler way. The attributes are analyzed separately, in which equal importance is given for all [9].

Results of the models:

Data analysis models are often characterized by performance parameters such as accuracy, True Positive Value, False Positive Value, Precision, Recall, F-Measure. Also, the accuracy of the classifier model is fully dependent on the number of correctly classified instances. The two decision values based on the decision trees model through all the above trade-off values are tabulated and compared here below

Accuracy value: Algorithms that reach 100% predictions against the level of deviation that may fall up to 51% of holistic measure tabulated to be powerful and efficient [4].

$$\text{Accuracy} = \frac{\text{TP} + \text{TNNumber of correct predictions}}{Total\ number\ of\ predictions}$$

equation (1)

And accuracy other terms are truly positive, false positive,

a true negative and false negative is calculated as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{TP + TN + FN + FP}$$

equation (2)

TRUE POSITIVE RATE: TP data is the positive percentages that are correctly classified instances as given in class.

FALSE POSITIVE RATE: FP data is the positive percentages that are incorrectly classified instances as given in class.

PRECISION: It is the positive prognostic worth value obtained as being the number of relevant instances from the retrieved instances domain.

RECALLVALUE: It is the vulnerable rate of the ratio between the relevant instances and actually retrieved instances.

F-MEASURE: In WEKA, the f-measure due to machine learning is calculated as a two-class classifier against the generated values of f-measure obtained from 3 different data analytics models viz, NB classifier, f-measure for J48, and a weighted f-measure of these two algorithms.

The learning model is based on 10 fold cross corroboration from the training set and is based on the various parameters like Accuracy, Total Positive value, False Positive, Recall, Precision, F-measure, and hence results are classified based on accuracy and are shown below.

E. Solution due to Assignment method

Optimality criterion-based problems are related to solving by using assignment problems to get optimized results subject to the available constraints. Hitherto the field of agriculture and its own applications can be addressed by using the same operational research arena. Several attribute factors of this domain also characterize an optimality criterion to maximize the yield outcome. The assignment problem on M X N matrices in which an M refers to M objects/jobs and N refers to N entities/person top whom an optimal assignment could be done with. Normally an assignment problem is used to find the minimum cost value of optimization; here Hungarian algorithm is applied for the maximization problem arena by augmented modification.

Here what is done is that the maximum of the matrix is subtracted from each and every member of it, and hence the matrix is changed to follow the normal assignment problem. The data confined to 19 different districts of Tamilnadu for 15 varieties of seeds are marked as matrices, and a pivotal value is considered for arriving at the modified assignment problem version. Values confined to rainfall, investment amount, harvesting days, and soil-related parameters including pH, soil contents like mineral strength as to vital nutrients are considered. The results of analysis characterized due to the following min-max equation form subject to the impost constraints are shown below. The formula for the assignment problem to solve the matrices of the yield cost on X is

$$X(cost) = \sum_{i=1}^{n}\sum_{j}^{n} a_{ij}Y_{ij} \sum inv_{[i]} \sum hac_{[i]} \sum rai_{[i]} \sum So_{[i]}$$

Where $Y_{ij} =$

$$\begin{cases} 1, & \text{if the ith person is assigned jth work} \\ 0, & \text{if the ith person is not assigned the jth work} \end{cases}$$

With the constrains as shown below

i) $\sum_{i=1}^{n} X_{ij}=1;\ j=1,2,3\ldots n,$ { i.e., i^{th} person will do only one work

ii) $\sum_{j=1}^{n} X_{ij}=1;\ i=1,2,3\ldots n,$ { i.e., j^{th} person work will be done only by one person

The data pertaining to the above scenario are input from the Agricultural portal of Aduthurai Rice Research Institute, and hence the problem was analyzed for finding the highest yield value with assigning the jth seed variety to ith district domain.

VII. Results and Discussion

A. Apriori algorithm Results

The WEKA mining tool is used for the analysis of data, and hence the results are shown. In the first phase, seed varieties are analyzed for selection to various districts. Then the apriori algorithm produces the frequent item patterns in the form sets, and the whole results are tabulated below.

Table 6: shows the output of the LATE SAMBA season using the Apriori algorithm:

WEKA TOOL USING LATE SAMBA USING APRIORI ALGORITHM		
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1		
Relation: LATE SAMBA SEED		
Instances: 19		
Attributes: 16		
=== Run information ===		
Minimum support: 0.95 (18 instances)		
Minimum metric <confidence>: 0.9		
Number of cycles performed: 1		
Generated sets of large itemsets:		
=== Classifier model (full training set) ===		
SEEDS /INPUT	ITERATIONS	Best rules found:
ADT 38	Size of set of large itemsetsL(1): 5	1. ADT (R) 46=T 19 ==> ADT 39=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
ADT 39	Size of set of large itemsetsL(2): 10	2. ADT 39=T 19 ==> ADT (R) 46=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
TRY 1*	Size of set of large itemsetsL(3): 10	3. CO (R) 48=T 19 ==> ADT 39=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
TRY 3*	Size of set of large itemsetsL(4): 5	4. ADT 39=T 19 ==> CO (R) 48=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
ADT (R) 46	Size of set of large itemsetsL(5): 1	5. CO (R) 49=T 19 ==> ADT 39=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
ADT (R) 49		6. ADT 39=T 19 ==> CO (R) 49=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
CO (R) 48		7. CO (R) 50=T 19 ==> ADT 39=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
CO (R) 49		8. ADT 39=T 19 ==> CO (R) 50=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
CO (R) 50		9. CO (R) 48=T 19 ==> ADT (R) 46=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
I.White Ponni		10. ADT (R) 46=T 19 ==> CO (R) 48=T 19 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
ASD 19		
TPS 2		
TPS 3		
ASD 18		
MDU 5		

Following discussion is drawn out of the data analysis due to the apriori method. Traditional seed analysis by farmers has resulted in the selection of only one type of seed to be best, whereas apriori analysis has resulted in the case that three different varieties if selected, could be optimal in yield for farmers out of range of fifteen seed varieties. As of the apriori algorithm, the frequent itemsets favored ADT ®46, ADT 39, and CO ® 49 apart from the CO ® 50 with less frequency in selection.

B. J48 algorithm results:

J48 algorithm is used for the design of a decision support system. Normally farmers analyze all vital data like availability of water resource, seed variety, monsoon pattern, and others, which decide to dictate reaping high yield in Kg/Ha on paddy cultivation pattern in specific areas. The entire data quoted above are now subjected to the j48 algorithm to decide as to which area and which seed has a high yield with its pertinent factors.

Table 7: shows the output of the j48 algorithm

WEKA TOOL USING LATE SAMBA USING J48 ALGORITHM

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: LATE SAMBA

Instances: 19

Attributes: 20

=== Run information (TOTAL IRRIGATED AREA('000 ha))===

Test mode: 10-fold cross-validation
Time is taken to build the model: 0 seconds
Number of Leaves: 7
Size of the tree: 8

J48 pruned tree
91.8 = 69.2: 23.4 (1.0)
91.8 = 116.0: 77.0 (1.0)
91.8 = 92.6: 33.0 (1.0)
91.8 = Tirunelveli: Thoothukudi (1.0)
91.8 = F: F (7.0)
91.8 = T : T (1.0)
91.8 = T: T (7.0)

=== Classifier model (full training set) ===

DISTRICT	TOTAL IRRIGATED AREA('000 ha)	SEEDS	Stratified cross-validation	Summary
Thanjavur	167.6	ADT 38		
Nagapattinam	123.4	ADT 39		
Tiruvarur	166145	TRY 1*		
Tiruchirapalli	102.3	TRY 3*		
Kanchipuram	122.3	ADT (R) 46		
Tiruvallur	93.8	ADT (R) 49		
Coimbatore	91.1	CO (R) 48		
Tirupur	11.1	CO (R) 49		
Erode	173.3	CO (R) 50		
Karur	14.9	I.White Ponni		
Perambalur	12.4	ASD 19		
Ariyalur	29.1	TPS 2		
Kanyakumari	20.3	TPS 3		
Pudukottai	104.9	ASD 18		
Madurai	71	MDU 5		
Dindigul	19.6			
Theni	14.7			
Tirunelveli	91.8			
Thoothukudi	18			

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.417	23.4
0	0	0	0	0	0.417	77.0
0	0	0	0	0	0.417	33.0
0	0	0	0	0	0.389	Thoothukudi
0.714	0.583	0.417	0.714	0.526	0.714	F
0	0	0	0	0	0.361	T
0.714	0.167	0.714	0.714	0.714	0.774	T
Weighted Avg.						
0.526	0.276	0.417	0.526	0.457	0.654	

The above results have concluded with a viable implication that the Thoothukudi district has reached the highest yield outcome when compared to the remaining 19 districts with consideration to all 15 seed varieties. Based on the results raised above, we infer a vulnerable result that the True Positive value has reached a level to be 0.714, which is more appreciably higher than the corresponding False Positive value.

C. Naïve Bayes algorithm results:

As Naïve Bayes algorithm having its base on Bayesian theorem is another method that supports the design of decision support system and the results of the algorithm are often viable for interpretation, hitherto the domain of concern having a selection of seed patterns for late samba is analyzed, and the results are enlisted as of below.

Table 8: shows the output of the Naïve Bayes algorithm:

WEKA TOOL USING LATE SAMA USING NAIVE BAYES ALGORITHM				
Scheme: weka.classifiers.bayes.NaiveBayes				
Relation: Late Samba				
Instances: 19				
Attributes: 20				
=== Run information (TOTAL IRRIGATED AREA(‘000 ha))===				
Test mode: 10-fold cross-validation Time is taken to build the model: 0 seconds				
=== Classifier model (full training set) ===				
DISTRICT	TOTAL IRRIGATED AREA(‘000 ha)	SEEDS	Stratified cross-validation	Summary
Thanjavur	167.6	ADT 38		
Nagapattinam	123.4	ADT 39		
Tiruvarur	166145	TRY 1*		
Tiruchirapalli	102.3	TRY 3*		
Kanchipuram	122.3	ADT (R) 46		
Tiruvallur	93.8	ADT (R) 49		
Coimbatore	91.1	CO (R) 48		
Tirupur	11.1	CO (R) 49		
Erode	173.3	CO (R) 50		
Karur	14.9	I.White Ponni		
Perambalur	12.4	ASD 19		
Ariyalur	29.1	TPS 2		
Kanyakumari	20.3	TPS 3		
Pudukottai	104.9	ASD 18		
Madurai	71	MDU 5		
Dindigul	19.6			
Theni	14.7			
Tirunelveli	91.8			
Thoothukudi	18			

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0.056	0.5	1	0.667	0.972	23.4
0	0	0	0	0	0.972	77.0
1	0	1	1	1	1	33.0
1	0	1	1	1	1	Thoothukudi
0.857	0.083	0.857	0.857	0.857	0.988	F
0	0	0	0	0	0.222	T
0.857	0.167	0.75	0.857	0.8	0.905	T
0.789	0.095	0.724	0.789	0.751	0.917	Weighted Avg.

Here also, results have predicted that the Thoothukudi district has reached a maximum value in terms of yield, with its corresponding True Positive value as being 0.857. The results being produced here are very similar to that of the results raised in the J48 algorithm, with only subtle differences as depicted in the comparison shown below.

Table 9: shows the comparison of J48 and Naïve Bayes results

ALGORITHMS	Instances: seeds, productivity, rainfall, total water resource, districts, and total irrigation area	Attributes: values of seeds, productivity, rainfall, total water, no: of districts, and total irrigation area	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error	Time is taken to build a model	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic
.J48 -C 0.25 -M 2	19	20	0.1353	0.327	61.1829 %	99.019 %	0.02 seconds	52.6316 %	47.3684 %	0.25
Naive Bayes Classifier	19	20	0.0731	0.2277	33.0366 %	68.9582 %	0 seconds	78.9474 %	21.0526 %	0.6984

Based on the results tabulated above, the following conclusion is drawn between the classifier method models. A number of correctly classified instances, being the accuracy value, shows the correct prediction method, and therefore the method due to Naïve Bayes having the highest record of accuracy with 78.9474% is the best prediction method.

D. Modified Assignment problem results:

Synthesizing the seed analysis for optimal yield by means of the modified assignment method as an operational research problem has given very valuable and valid results, which are codified here for reference and implication. In the set method, matrix roles are considered to be various districts that have paddy as the primary cultivating variety, and the columns are arranged with various seed varieties that were considered by the domain. For easy reference, numerical codes are assigned to all districts, and rows and seed varieties are given with alphabetic assignments.

Districts are numbered as 1, 2, 3, … and seeds are considered as A, B, C …

The whole assignment of rows and columns with districts versus seed patterns areas 1. Thanjavur 2. Nagapattinam, 3. Tiruvarur, 4. Tiruchirapalli, 5. Kanchipuram 6. Tiruvallur 7. Coimbatore, 8. Tirupur, 9. Erode 10. Karur 11. Perambalur, 12. Ariyalur, 13. Kanyakumari, 14.Pudukottai, 15. Madurai, 16. Dindigul, 17. Theni, 18. Tirunelveli, 19. Thoothukudi as rows and columns of seeds as A= ADT 38, B=ADT 39, C=TRY 1,D= TRY 3, E=ADT (R) 46, F=ADT (R) 49, G=CO (R) 48, H=CO (R) 49, I=CO (R) 50, J=White Ponni, K=ASD 19, L=TPS 2, M=TPS 3, N=ASD 18, O=MDU 5. Beyond these, values pertaining to rainfall, investment amount, harvesting days, and soil values are inducted. This assignment algorithm is coded in a python program, and hence the results are shown below.

Table 10:: **Shows the input value of every seed average yielded value for every district and other values**

LATE SAMBA	ADT 38	ADT 39	TRY 1*	TRY 3*	ADT (R) 46	ADT (R) 49	CO (R) 48	CO (R) 49	CO (R) 50	White Ponni	ASD 19	TPS 2	TPS 3	ASD 18	MDU 5	INVESTMENT	RAINFALL
Thanjavur	6200	5000	5255	5833	6656	6173	9625	6286	6338	0	0	0	0	0	0	47425	318
Nagapattinam	6203	5001	5251	5832	6652	6172	9622	6281	6339	0	0	0	0	0	0	47200	286
Tiruvarur	6202	5002	5252	5831	6653	6172	9621	6289	6338	0	0	0	0	0	0	47300	296
Tiruchirapalli	0	5003	5253	5830	6651	6170	9624	6282	6334	0	0	0	0	0	0	47500	294
Kanchipuram	0	5004	5254	5833	6653	6173	9621	6288	6338	4500	0	0	0	0	0	47100	491
Tiruvallur	0	5002	5255	5831	6654	6175	9623	6282	6332	4501	0	0	0	0	0	47000	252
Coimbatore	6201	5001	0	0	6655	6174	9622	6287	6335	0	0	0	0	0	0	47800	190
Tirupur	6195	5005	0	0	6657	6173	9622	6283	6332	0	0	0	0	0	0	47100	155
Erode	0	5000	5251	0	6658	6178	9621	6289	6338	0	0	0	0	0	0	47900	230
karur	6196	5001	5253	0	6651	6179	9625	6284	6339	4503	5800	0	0	0	0	48000	214
Perambalur	6197	4999	5254	0	6652	6172	9623	6285	6334	4502	5801	0	0	0	0	47000	291
Ariyalur	6200	4998	5255	0	6653	6173	9621	6281	6338	4499	5803	0	0	0	0	47800	392
Kanyakumari	0	5000	5251	0	6655	6175	9621	6285	6331	4498	5801	4615	5600	7000	5000	47500	477
Pudukottai	6198	4997	5254	0	6652	6177	9623	6289	6335	0	5804	0	0	0	0	47200	351
Madurai	6204	4996	0	0	6654	6152	9625	6287	6338	4501	0	0	0	0	0	47500	336

Dindigul	6202	5000	0	0	6658	6177	9622	6281	6339	4502	0	0	0	0	0	47000	295
Theni	6198	5002	0	0	6655	6173	9623	6284	6337	4499	0	0	0	0	0	46900	158
Tirunelveli	0	5003	5251	0	6652	6178	9624	6285	6335	4498	0	0	0	0	0	46500	142
Thoothukudi	0	5000	5255	0	6655	6172	9624	6286	6338	4500	0	0	0	0	0	46800	75
HARVESTING DAYS	135	125	135	135	135	137	135	135	130	140	127	125	140	110	5833		

Figure 1 shows the output of the maximum yielded seed with its area outcome

.

```
-29337  -24838  -24838  -24838  -24838  -24838  31073
        -24838  -29841  -30089  -24838  -31490  -31016  31074   -31123  -31173
-29336  -24838  -24838  -24838  -24838  -24838  31073
        -24838  -29838  -30093  -24838  -31493  -31010  31074   -31124  -31176
-29338  -24838  -24838  -24838  -24838  -24838  31073


District:seeds  :Yield/Hect:Invstment Rs.: Rainfall MM : No:of days harvesting
2       :I      :3286   :       47200   :     286       :     130
 Available soils and their values for better Yield=>
Red loam Soil, alluvium Soil, PH:6.5-7.5(NEUTRAL) 89%, NITROGEN: <280 kg/ha(LOW)
 100%, PHOSPHORUS: >>22 kg/ha(HIGH) 81% and POTASSIUM: 118-280 kg/ha(MEDIUM) 88%


3       :H      :3336   :       47300   :     296       :     135
 Available soils and their values for better Yield=>
Red loam Soil, alluvium Soil, PH:6.5>7.5(NEUTRAL) 85%, NITROGEN: <280 kg/ha(LOW)
 81%, PHOSPHORUS: >22 kg/ha(HIGH) 91% and POTASSIUM: >280 kg/ha(HIGH) 54%

Total Assignment Cost = 12214 RS.
```

The analysis due to the assignment problem approach has given its useful results to have its say and rank on the agriculture domain as follows. Important results achieved due to districts versus seed patterns along with factors like rainfall, investment, harvesting days, and PH value of the soil having determinants like Nitrogen, Phosphorus, and Potassium. The matrix analysis as given the pivotally equation as "Optimum Seed= All Districts [i] + Seed_ Management[j] + Soils[i] + Rainfall[i]".

Results

- District 2 Nagapattinam, with the seed of I = CO (R) 50, the average seed yield value of 6338 Kg/Hec, the average rainfall of the area is 286 MM with the harvesting days of 130 and with the soil of Red Loam soil, alluvium soil and PH values of 6.5 - 7(Neutral) 89%, with the Nitrogen of < 280 Kg /Ha (Low) 100%, Phosphorus of > 22 Kg/Ha (High) 81% and Potassium of 118 – 280 Kg/Ha (medium) 88%.
- District 3 Tiruvarur, with its seed of H = CO (R) 49, this seed has an average yield of 6286 Kg/Hec, with the rainfall of 296 MM, harvesting days of the seed is 135 days, and with the soil of Red Loam soil, alluvium soil and PH values of 6.5 > 7.5(Neutral) 85%, with the Nitrogen of < 280 Kg /Ha (Low) 181%, Phosphorus of > 22 Kg/Ha (High) 91% and Potassium of >280 Kg/Ha (High) 54%.

VIII. DISCUSSION

Based on figure 1, the following results of inference. CO (R) 50 and CO (R) 49 are the highest yielding seed varieties among others. Yet other algorithms recommended that ADT ®46, ADT 39, and CO ® 49 are the most productive seed varieties selected by the farmers from all areas, and Nagapattinam, Tiruvarur, and Thoothukudi districts are the high yield recording districts. Thus the problem so embarked here is analyzed threadbare, and hence very useful results have been drawn with notions of inference for pure selection and implication by agricultural practitioners and farmers.

IX. CONCLUSION

Farmers' Growth decides the economical Growth of any country, having agriculture as the primary industry. Therefore this work based on agricultural data analytics is

so productive in terms of its giving useful and viable results of implication to be practiced by agrarian farmers, which have a sure say in the determination of economical growth pattern and socio-economic trendsetting. Future cultivating patterns in agriculture could be so determined by advancements witnessed in the agricultural industry, and hence the deterministic factors on agricultural yield could be enhanced into all the above algorithmic methods to cater to the needs of the day, to be a promising culture in data analytics of the agricultural domain.

X. References

[1] Suvidha Jambekar, Shikha Nema, Zia Saquib, Application of Data Mining Techniques for Prediction of Crop Production in India, International Journal of Innovations & Advancement in Computer Science, 7(4)(2018).

[2] Jignasha M. Jethva1, Nikhil Gondaliya2, Vinita Shah3, A Review on Data Mining Techniques for Fertilizer Recommendation, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | 3(1) | ISSN: 2456-3307

[3] Humayra Dil Afroz1, Dr.Mohammad Anwar Hossen2, New Proposed Method for Solving Assignment Problem and Comparative Study with the Existing Methods, IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-5728, p-ISSN: 2319-765X. Ver. IV 13(2) (2017) 84-88

[4] P. Grace Sharon, Application of Data Mining In Agriculture, International Journal of Computer Science and Information Technology Research ISSN 2348-120X (Online) 2(3)(2014) (158-162).

[5] U. Singh1, K. Garg A Review Technique of Data Mining in the Agronomy field, International Journal of Computer Sciences and Engineering.6(5)(2018).

[6] E.Manjula* S.Djodiltachoumy, Analysis of Data Mining Techniques for Agriculture Data, International Journal of Computer Science and Engineering Communications, 4(2)(2016) 1311-1313.

[7] D. Sabareeswaran[*1], A. Edwin Robert[2] A survey on data mining techniques in agriculture, Indian Journal of Innovations and Developments. 5(8)(2016) 2277-5382, ISSN (Print).

[8] NP Dharshinni Analysis of Accuracy K-Means and Apriori Algorithms for Patient Data Clusters, et al., J. Phys.: Conf. Ser. 1230 012020, Journal of Physics: Conference Series. (2019).

[9] Jignasha M. Jethva1, Nikhil Gondaliya2, Vinita Shah3 A Review on Data Mining Techniques for Fertilizer Recommendation, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | 3(1) | ISSN: 2456-3307.

[10] Kusum Lata, Sajidullah S. Khan Experimental Analysis of Machine Learning Algorithms Based on Agricultural Dataset for Improving Crop Yield Prediction, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 9(1)2249 – 8958.

[11] Viviliya B, Vaidhehi V The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 9(2)(2019) 2278-3075.

[12] Jagdish Kumar, Achila Singh, Manoj Mishra, Deepti Pal, and HG Prakash. Effect of Bt Hybrids On Growth, Yield Attributes, And Seed Cotton Yield In Hirsutum Cotton SSRG International Journal of Agriculture & Environmental Science 8.1 (2021): 17-19.