

Enhanced and Improved Hybrid Model to Prediction of User Awareness in Agriculture Sector

A.V.S. Pavan Kumar

Research Scholar,
Department of Computer Science and Engineering,
GIT, GITAM, Visakhapatnam

Dr. R. Bhramaramba

Associate Professor,
Department of Information Technology,
GIT, GITAM, Visakhapatnam

Abstract—Agriculture is the backbone of Indian economy and is the main income source for most of the population in India. So farmers are always curious about yield prediction. Crop yield depends on various factors like soil, weather, rain, fertilizers and pesticides. Several factors have different impacts on agriculture, which can be quantified using appropriate statistical methodologies. Applying such methodologies and techniques on historical yield of crops, it is possible to obtain information or knowledge which can be helpful to farmers and government organizations for making better decision and policies which lead to increased production. The main drawbacks of Indian farmers are they do not have proper knowledge regarding crop yield based on soil necessities. So in this paper, we proposed and developed an Improved Hybrid Model (which is combination of both classification, i.e. Artificial Neural Networks and clustering approach i.e. k-means (works based on Euclidean distance)) to provide awareness, usage and prediction to each farmer that relates to classify different crop yield representation based on soil necessity. For that we collected farmer's data from standard repositories like http://www.tropmet.res.in/static_page.php?page_id=52#data and then using that data provide awareness and other parameter sequences to all the farmers in India. Our experimental results show efficient e-agriculture with respect to user awareness, usage and prediction with respect to prediction, recall and f-measure for supporting real time marketing of different agriculture products.

Keywords—Agriculture products; e-agriculture; classification; clustering; ensemble model

I. INTRODUCTION

India is described by little homesteads. More than 75% of aggregate land capitals inside the nation are under 5 sections of land. Most yields are rain sustained, with pretty much 45% of the land inundated. According to a few estimations, around 55% of aggregate populace of India relies upon cultivation. In the US, in light of the fact of overwhelming automation of agriculture, it is around 5%. India is one of the greatest makers of agrarian items and still has exceptionally less ranch profitability. Efficiency should be expanded so agriculturists can get more pay from a similar land parcel with less work. Accuracy in agriculture gives an approach to do it. Exactness cultivating, as the name suggests, alludes to the applying of exact and appropriate aggregate of remark like pee, manures, soil and so on at the best possible time to the gizzard for expanding its profitability and expanding its yields. Not all exactness agriculture frameworks offer best outcomes [7], [9]. In any case, in agribusiness it is vital that the proposals made

are exact and exact in light of the fact that if there should be an occurrence of mistakes it might prompt overwhelming material and capital misfortune [13]. Numerous inquiries about are being done, so as to achieve an exact and proficient model for trim forecast.

Improved Hybrid Model (which is combination of both classification i.e., Artificial Neural Networks and clustering approach i.e. k-means (works based on Euclidean distance)) (Ensemble approach) is one such strategy that is incorporated into such research works. Among these different machine learning procedures that are being utilized as a part of this field. This paper proposes a framework that uses the voting technique to assemble a productive and exact model. The agriculture part in India is at present confronting a troublesome stage. India is moving towards an agribusiness crisis because of deficient interest in irrigational and farming framework, absence of consideration, insufficient land administration, not given of reasonable costs to ranchers for their yields and inadequate land change in India, and so on. Sustenance creation and efficiency in India is declining while its nourishment utilization is expanding. The circumstance has additionally been exacerbating because of utilization of sustenance grains as a result of interest of bioenergizers. As India does not have ports and calculated frameworks for extensive - scale sustenance imports, the arrangement of import of sustenance grains would be difficult. In our proposed approach e-agriculture is a rising field in the association of agricultural informatics, advancement and enterprise which is focusing to farming administrations, innovation dissemination and data conveyed or created through the Internet and related advances [8]. In particular, it connects with the conceptualization, outline, improvement, appraisal and application of imaginative approaches to utilize dynamic or developing Information and Communication Technologies (ICTs) [12].

Our proposed approach is a rising technology for upgrading existing agriculture [10], [11] and sustenance security through improved procedures for learning access and to switch to utilizing data and correspondence advancements. The World Summit on the Information Society (WSIS) Plan of Action includes e-Agriculture as a locale of capacity of data and correspondence innovations (ICTs). In short e-Agriculture will associate every concerned individual beginning from ranchers to scientists together. Agriculturists can get the coveted data at any moment of time from any piece of world and they can

likewise get the assistance from specialists seeing their concern instantly by without moving anyplace.

II. REVIEW OF RELATED WORK

Satish et al. [1] states the necessities and arranging required for building up a product display for accurate cultivation is examined. It profoundly ponders the nuts and bolts of accurate cultivation. The creator's beginning from the nuts and bolts of accurate cultivation and moves towards building up a model that would bolster it. This paper depicts a model that applies Precision Agriculture (PA) standards to little, open ranches at the individual rancher and yield level, to influence a level of control over changeability. The extensive target of the model is to convey guide warning administrations to even the littlest rancher at the level of his/her littlest plot of harvest, utilizing the most open innovations, for example, SMS and email. This model has been intended for the situation in Kerala State where the normal holding size is much lower than the vast majority of India. Thus this model can be situated somewhere else in India just with a few alterations. Anshal Savla et al. [2] makes a qualified meditation of grouping calculations and their execution in yield forecast in exactness farming. These calculations are executed in an informational collection gathered for quite a while in yield expectation on soya bean trim. The calculations utilized for yield forecast in this paper are Support Vector Machine, Random Forest, Neural Network, REPTree, Bagging, and Bayes. The conclusion drawn at the end is that packing is the best calculation for yield expectation among the above expressed calculations since the mistake deviation in stowing is least with a mean outright blunder of 18985.

M.P. Singh et al. [3] demonstrates the significance of yield choice and the elements choosing the harvest choice like creation rate, showcase cost and government strategies are talked about. This paper proposes a Crop Selection Method (CSM) which takes care of the product determination issue and enhances net yield rate of the harvest. It recommends a progression of yield to be chosen over a season considering factors like climate, soil write, water thickness, edit type. The anticipated estimation of powerful parameters decides the precision of CSM. Consequently there is a need to incorporate a forecast strategy with enhanced exactness and execution. Liying Yang et al. [4] expects to tackle the pivotal issue of choosing the classifiers for the troupe learning. A technique to choose a best classifier set from a pool of classifiers has been proposed. The proposition expects to accomplish higher precision and execution. A technique called SAD was proposed in view of precision and characterization execution. Utilizing Q measurements, the reliance between most significant and exact classifiers is distinguished. The classifiers which were not picked were joined to shape the troupe. This measure should guarantee higher execution and decent variety of the outfit. Different strategies, for example, SA (Selection by Accuracy), SAD (Selection by precision and Diversity) and NS (No choice) calculation were distinguished. At last it is gathered that SAD works superior to others.

Shakil Ahamed et al. [5] propose different order techniques to group the liver ailment informational index. The paper stresses the requirement for exactness since it relies upon the

dataset what's more, the realizing calculation. Order calculations such as Naïve Bayes, ANN, ZeroR and VFI were utilized to group these sicknesses and think about the adequacy, adjustment rate among them. The execution of the models was contrasted with precision and computational time. It was reasoned that every one of the classifiers with the exception of innocent bayes demonstrated enhanced prescient execution. Multi-layer perception demonstrates the most noteworthy precision among the proposed calculations. Aymen E Khedr et al. [6] tries to take care of the issue of nourishment frailty in Egypt. It proposes a system which would foresee the generation, and import for that specific year. It utilizes Artificial Neural Networks alongside Multi-layer perceptron in WEKA to fabricate the forecast. Toward the finish of the procedure we would have the capacity to imagine the measure of generation import, need and accessibility. Thus it would settle on choices on whether sustenance must be additionally transported in or not. The dirt datasets in paper are examined and a classification is anticipated. From the anticipated soil class the edit yield is distinguished as a Classification run the show. Credulous Bayes furthermore, KNN calculations are utilized for trim yield expectation. The future work expressed is to make productive models utilizing different characterization methods, for example, bolster vector machine, main segment investigation. The advantages of the Indian agriculturists: if the market and climate data is conveyed to their cell phones [14]. Furthermore, this has been led with a randomized trial in 100 towns of Maharashtra. This administration has been sent in by a business benefit called Reuters Market Light (RML). The treated ranchers connect RML data with various choices they have made in the farming, and we find that the treatment influenced spatial arbitrage and product reviewing. Be that as it may, the size of these impacts is little. We discover no measurable noteworthy normal impact of treatment on the cost got by ranchers, edit esteem - included, trim misfortunes coming about because of rainstorms, or the probability of changing harvest assortments and development hones [15]. The information that are applicable of the required quality dependably have the capability of expanding effectiveness in all circles of movement of an Indian rancher, accordingly the developing situation of the deregulated horticulture, has brought a need and direness to guarantee it in a fundamental piece of basic leadership. In this way, investigating IT as a key device is the advantage of country like India of accepted significance. Here the data meets the Indian ranchers when all is said is done which are recorded broadly.

III. DESCRIPTION OF PROBLEM

The farming industry in Native Indian is currently experiencing a hard stage. Native Indian is moving towards an farming emergency due to inadequate investment in irrigational and farming facilities, lack of attention, worthless area management, non-given of fair prices to farm owners for their plants and inadequate area change in Native Indian, etc. Meals manufacturing and efficiency in Native Indian is decreasing while its food consumption is increasing. The situation has further been difficult due to use of food grain because of demand of bioenergy sources. As Naive Bayesian does not have slots and logistical systems for large - scale food

imports, the solution of transfer of food grain would be a challenge.

By the use of ICT, India's food manufacturing and efficiency has been increased for farming reasons. The developed countries are using technological innovation of laser in place of vehicles to plough lands. This helps in improving the use of a range of information parameter such as water, plant seeds, plant foods, etc. The issue occurs here is that Native Indian farm owners cannot pay for this technological innovation. In addition, energy and power also cause a major issue for Native Indian farm owners and choice of energy like solar panel technology sections, controlled and enhanced by ICT.

IV. PROPOSED METHODOLOGY AND IMPLEMENTATION

The proposed technique, i.e. Enhanced Hybrid Model performs information parceling with Principal segment. It parcels the given informational index into k sets. The middle of each set can be utilized as great beginning group focuses and afterward allot every datum focuses to its closest bunch centroid. The underlying centroids of the bunches are given as information. It begins by framing the underlying bunches in view of the relative separation of every datum point from the underlying centroids. The Euclidean separation is utilized for deciding the closeness of every datum point to the group centroids. For every datum indicate, the bunch which it is doled out and its separation from the centroid of the closest group are noted. For each group, the centroids are recalculated by taking the mean of the estimations of its information focuses. The strategy is relatively like the first k-implies calculation aside from that the underlying centroids are figured deliberately. The following stage is an iterative procedure which makes utilization of a heuristic technique to enhance the proficiency. Amid the emphasis, the information focuses may get redistributed to various bunches. The strategy includes monitoring the separation between every datum point and the centroid of its present closest group. Toward the start of the emphasis, the separation of every datum point from the new centroid of its present closest bunch is resolved. On the off chance that this separation is not exactly or equivalent to the past closest separation, that means that the information point remains in that group itself and there is no compelling reason to process its separation from different centroids. This outcome in the sparing of time required to register the separations to k-1 group centroids. Then again, if the new centroid of the present closest group is more inaccessible from the information point than its past centroid, there is a shot for the information point getting incorporated into another closer bunch. All things considered, it is required to decide the separation of the information point from all the group centroids. This technique enhances the proficiency by lessening the quantity of calculations.

This ANN K-Means clustering approach illustrated as follows: Several steps of data processing were followed by details pre-processing, data reduction, Data Mining, details clustering, information interpretation, neighborhood study and statistical diagnosis shown in Fig. 1.

Algorithm 1: Procedure for proposed algorithms with respect to different attributes

Algorithm	: ANN and K-Mean clustering algorithm
INPUT	: Agriculture Data set with d dimensions
OUTPUT	: K number of Clusters
Steps	
1.	Reduce the D dimension of the N data using Artificial Neural Networks (ANN) and prepare another N data with d dimensions ($d < D$).
2.	The Principal components are ordered by the amount of variance.
3.	Choose the first principal component as the principal axis for partitioning and sort it in ascending order.
4.	Divide the Set into k subsets where k is the number of clusters.
5.	Find the median of each subset.
6.	Use the corresponding data points for each median to initialize the cluster centers
7.	Compute the distance of each data-point x_i ($1 \leq i \leq n$) to all the centroids c_j ($1 \leq j \leq k$) using Euclidean distance formula.
8.	For each data object x_i , find the closest centroid c_j and assign x_i to the cluster with nearest centroid c_j and store them in array Cluster[] and the Dist[] separately. <ul style="list-style-type: none"> a. Set Cluster[i] = j, j is the label of nearest cluster. b. Set Dist[i] = $d(x_i, c_j)$, $d(x_i, c_j)$ is the nearest Euclidean distance to the closest center.
9.	For each cluster j ($1 \leq j \leq k$), recalculate the centroids;
10.	Repeat
11.	for each data-point <ul style="list-style-type: none"> 11.1 Compute its distance from the centroid of the present nearest cluster 11.2 If this distance is less than or equal to the previous nearest distance, the data-point stays in the cluster
12.	Else
13.	For every centroid c_j
14.	Compute the distance of each data object to all the centre
15.	Assign the data-point x_i to the cluster with nearest centroid c_j
16.	For each cluster j ($1 \leq j \leq k$), recalculate the centroids;
17.	Until the convergence criteria is met.

Performance evaluation for different crop yield parameters are like pH, TA, TDS, MG and CA with feasible environments sequences with respect to soil parameters shown in Table I. Following table shows effective results of classification measurements like precision, f-measure and recall with data representation may appear effective utilization soil analysis in agriculture sector for real time applications.

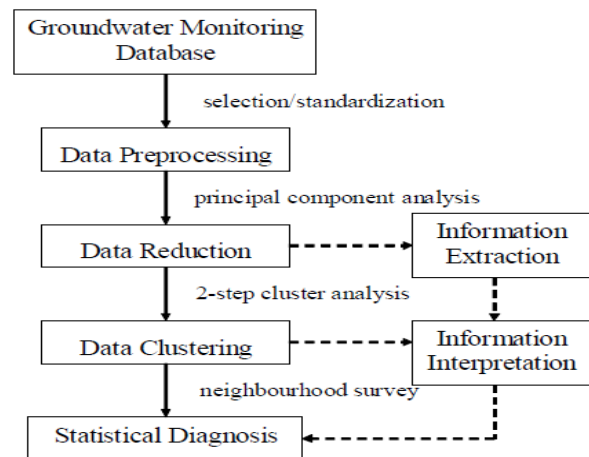


Fig. 1. Step by Step Procedure of Proposed Approach.

TABLE I. BASIC PARAMETER SEQUENCES FOR DIFFERENT SOIL PARAMETERS

Performance Measures	Accuracy	Precision	Recall	F-Measure
PH	0.95	0.83	0.84	0.84
TA	0.89	0.88	0.82	0.81
TDS	0.87	0.86	0.81	0.85
MG	0.85	0.89	0.84	0.84
F	0.87	0.76	0.84	0.82
CA	84.59	0.85	0.85	0.78

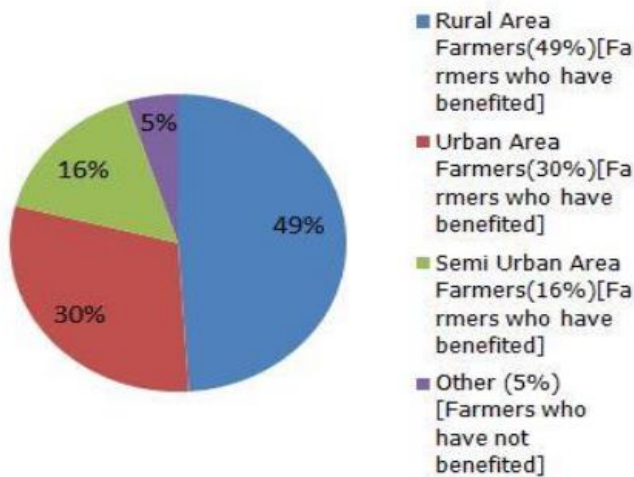


Fig. 2. Information of Different Soil Parameters at Different Areas in India.

Based on these parameters, performance evaluation of different formers at different parameter sequences is shown in Fig. 2.

A. Data Collection

The data set containing the ground specific features which are gathered from Poly test Labs ground examining lab, Pune, Maharashtra, Indian. In addition, similar resources of general vegetation information were also used from Marathwada University. The vegetation regarded in our design includes groundnut, impulses, pure cotton, vegetables, bananas, paddy, sorghum, sugarcane, and cilantro. The number of illustrations of each vegetation available in the training information set is proven. The attributes regarded where Detail, Structure, Ph, Soil Shade, Permeability, Water flow and drainage, Standard water having and Break down. The above mentioned factors of ground play a major part in the crop's capability to eliminate water and nutritional value from the ground. For vegetation development to be possible, the ground must provide appropriate atmosphere for it. Soil is the core of the origins. The water having potential decides the crop's capability to process nutritional value and other nutritional value that are turned into ions, which is the form that to obtain can use. Structure decides how permeable the ground is and the convenience of air and water action which is essential to prevent the vegetation from becoming water

logged. The stage of acid or alkalinity (Ph) is expert varying which affects the accessibility of ground nutritional value. The action of microorganisms present in the ground and also the stage of exchangeable metal can be impacted by PH. The water holding and drainage figure out the infiltration of origins. Hence for the following reasons the above mentioned parameters are viewed for selecting vegetation.

V. EXPERIMENTAL RESULTS

The suggested program can be experimentally confirmed in conditions of clustering efficiency. Evaluation can be made centered on parameters such as Time and Precision. A past method of E-agriculture does not use any information exploration methods. The proposed program uses information exploration way of clustering strategy to team the data of farm owners. The clustering efficiency can be calculated in following conditions namely precision, recall, F-measure.

Precision value which is calculated relies on the recovery of information at real beneficial forecast, incorrect beneficial. In health care information precision is calculated as the portion of good outcomes came back that are appropriate and shown in Fig. 3.

$$\text{Precision} = \frac{TP}{TP+FP}$$

TP-True positive FP-true negative

Recall value is measured relies on the recovery of details at real beneficial forecast, incorrect adverse shown in Fig. 4. In healthcare details precision is measured the amount of beneficial outcomes came back that are Remember in this perspective is also known to as the True Positive Amount. Remember is the portion of appropriate circumstances that are recovered,

$$\text{Recall} = \frac{TP}{TP+FN}$$

FN – false negative

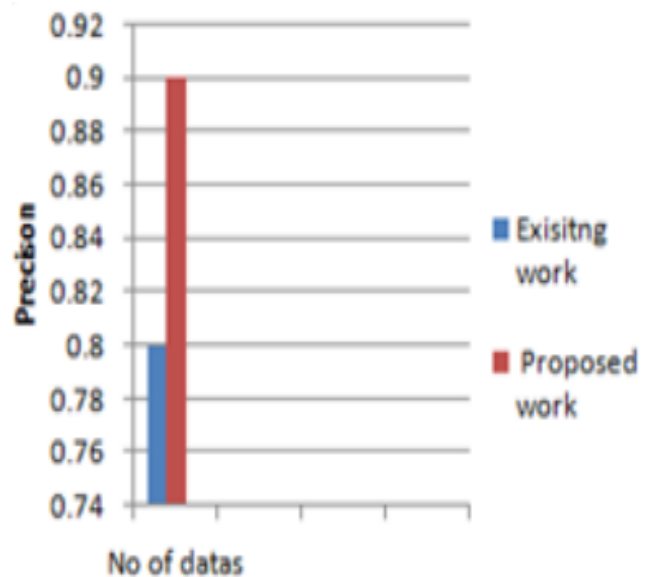


Fig. 3. Precision Comparison for Different Soil Ratios.

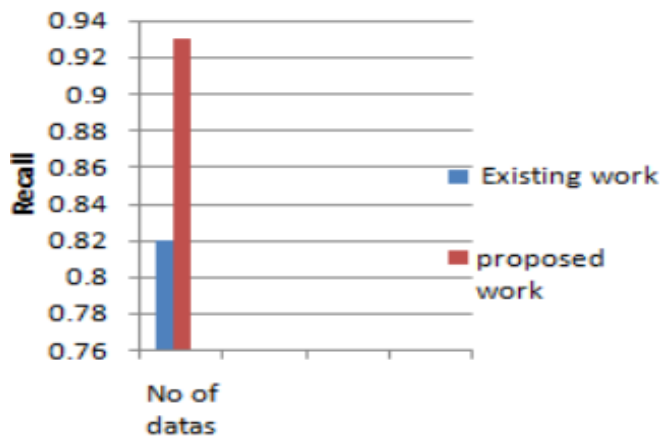


Fig. 4. Recall Comparison for Different Soil Parameters.

The F- Measure computes some average of the information retrieval precision and recall metrics is shown in Fig. 5.

Total comparison results for different soil parameters present in agriculture crop yield areas for real time applications is shown in Table II and Fig. 6.

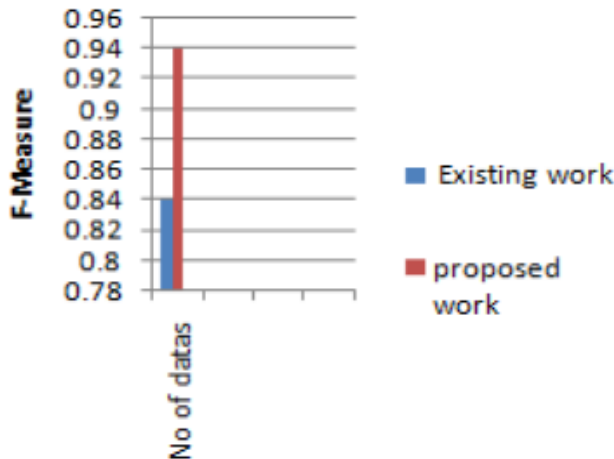


Fig. 5. F-Measure Parameter Sequences for Different Soil Parameters.

TABLE II. DIFFERENT VALUES FOR PROPOSED MEASUREMENTS FOR DIFFERENT SOIL QUALITY PARAMETERS

Soil Quality Parameters						
Performance Measures	PH	TA	TDS	MG	F	CA
Accuracy	80.25	83.42	83.89	82.38	81.29	84.39
Precision	0.83	0.88	0.85	0.83	0.70	0.82
Recall	0.84	0.82	0.81	0.84	0.84	0.85
F Measure	0.84	0.81	0.85	0.84	0.82	0.78

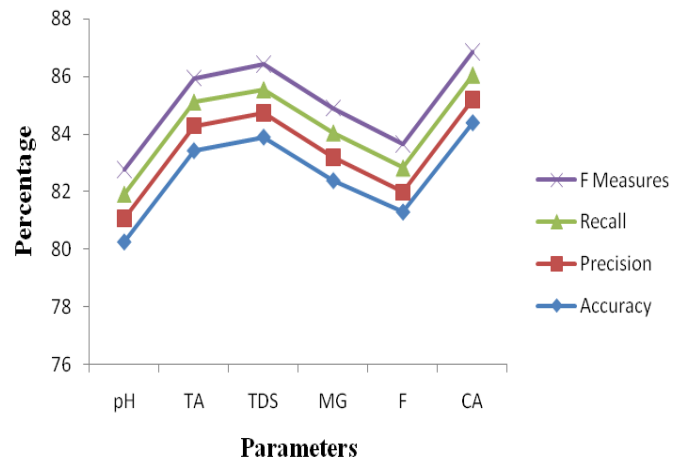


Fig. 6. Proposed Approach Performance at Different Soil and Water Quality Parameters.

Above results shown in Table II was proposed approach performance with different soil and other representations.

VI. CONCLUSION

Majority of farm owners in the state or nation are not aware that real time servers can be used to outperform companies and get details. The government should also perform sensitization to make attention for the farm owners on how best they can use information technological innovation to perform agribusiness. Our perform would help farm owners to increase efficiency in farming, avoid ground deterioration in harvested area, and decrease substance use in plants manufacturing and effective use of water sources. Our upcoming work is targeted at an enhanced information set with large number of features and also utilizes how to generate prediction.

REFERENCES

- [1] Babu, Satish. "A software model for precision agriculture for small and marginal farmers." In Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013 IEEE, pp. 352-355. IEEE, 2013.
- [2] Savla, Anshal, Nivedita Israni, Parul Dhawan, Alisha Mandholia, Himtanaya Bhadada, and Sanya Bhardwaj. "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture." In Innovations in Information, Embedded and Communication Systems (ICIECS), 2015 International Conference on, pp. 1-7. IEEE, 2015.
- [3] Kumar, Rakesh, M. P. Singh, Prabhat Kumar, and J. P. Singh. "Crop Selection Method to maximize crop yield rate using machine learning technique." In Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference on, pp. 138-145. IEEE, 2015.
- [4] Yang, Liying. "Classifiers selection for ensemble learning based on accuracy and diversity." Procedia Engineering 15 (2011): 4266-4270.
- [5] Ahamed, AT M. Shakil, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, and Rashedur M. Rahman. "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh." In 2015 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 1-6. IEEE, 2015.

- [6] Khedr, Ayman E., Mona Kadry, and Ghada Walid. "Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector applied case on food security information center ministry of agriculture, Egypt." *Procedia Computer Science* 65 (2015): 633-642.
- [7] Paul, Monali, Santosh K. Vishwakarma, and Ashok Verma. "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach." In *Computational Intelligence and Communication Networks (CICN), 2015 International Conference on*, pp. 766-771. IEEE, 2015.
- [8] Sharma, Latika, and Nitu Mehta. "Data mining techniques: A tool for knowledge management system in agriculture." *international journal of scientific & technology research* 1, no. 5 (2012): 67-73.
- [9] Mucherino, Antonio, and Georg Ruß. "Recent Developments in Data Mining and Agriculture." In *Industrial Conference on Data Mining-Workshops*, pp. 90-98. 2011.
- [10] Ruß, G., 2009, July. "Data mining of agricultural yield data: A comparison of regression models". In *Industrial Conference on Data Mining* (pp. 24-37). Springer, Berlin, Heidelberg.
- [11] Rajesh, D. "Application of spatial data mining for agriculture." *International Journal of Computer Applications* 15, no. 2 (2011): 7-9.
- [12] Darcy Miller, Jaki McCarthy, Audra Zakzeski, "A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing" National Agricultural Statistics Service 3251 Old Lee Highway, Fairfax, VA 22030 - JSM 2009
- [13] Kant, Srivastava Uma. "Agro-Processing Industries: Potential, Constraints and Task Ahead". No. WP1989-10-01_00902. Indian Institute of Management Ahmedabad, Research and Publication Department, 1989.
- [14] Gandhi, Vasant, Gauri Kumar, and Robin Marsh. "Agroindustry for rural and small farmer development: issues and lessons from India", *The International Food and Agribusiness Management Review* 2, no. 3-4 (1999): 331-344.
- [15] Mollinga, Peter P. "The rational organisation of dissent: Boundary concepts, boundary objects and boundary settings in the interdisciplinary study of natural resources management". No. 33. ZEF working paper series, 2008.