

Smart Yield Accuracy Prediction using Linear Regression and Collaborative Filtration



V Prithvi Ram, Rajeshwari S.B, Jagadish S Kalliman

Abstract: India is mainly based on farming. Agriculture is the main source of economy in India, but the farmers are suffering with many problems such as lack of crops yield, lack of water, soil fertility etc. To address those issues this recommendation system is proposed, and it significantly influences the crops yields. The need for the accessible data on the accomplishment for getting crops in good yields are investigated. To accomplish that, real-time data are collected from the farmers from different places of Karnataka. In this paper linear regression and collaborative filtering are used, and results are compared to draw an inference for more accurate recommendation system.

Keywords: crop yields, collaborative filtering, linear regression.

I. INTRODUCTION

In the progressing years, the gigantic volume of firm data in the agrarian party and its prerequisite for a beneficial and amazing data getting ready, animate the use of novel headways and stage to pick up, store, process, analyze and imagine tremendous enlightening lists for future conjectures. Enormous information is a propelling term given to a wide scope of data gathered headways in which the datasets are inconceivably broad that overseeing them become more testing than how it was [1]. Previously, in light of the essential challenges going up against the agribusiness division farmers feel continuously constrained to get heightened developing practices and doable rustic ones, in order to increase both money related and natural costs.

The utilization of Data and Correspondence Innovation (ICT) in agribusiness is continuously growing. E-Agribusiness incorporates the conceptualization, plan, progression, evaluation likewise, utilization of imaginative ways to deal with useful information and correspondence developments in provincial space, with the fundamental focus on cultivating. ICT can play a basic occupation in keeping up properties of information as it contains three guideline headways. These advances are associated for taking care of, exchanging and administering data information and learning [2]. The speedy movement in Data and Correspondences Advances (ICTs)

has offered various climb to new application that were incomprehensible just couple of years back. Cultivating is a basic section with the greater part of the nation crowds in making countries depending upon it. The part faces huge troubles of improving age in a condition of diminishing customary resources significant for age. ICT expect a crucial activity in testing and lifting the occupations of the natural masses using an Argo PC based information structure [3]. There are numerous Agro-Data Framework that will enables farmer to have noteworthy details about that harvest, for instance, thus groupings and various necessities like soil type, temperature, proportion of the manure, time of planting, time of progress, planting segment, sicknesses, bug, unsettling influence and Disease some confinement, precipitation, daylight, and so forth of that the yield. The various component for usage of this data picks the volume and modicum of the harvest yield. Agro-data is composing PC projects that orchestrated and finished, which engages the farmers to accomplish the ahead of time of referenced objectives.

The new ICT advances are hasty made just as, are delivering increasingly current systems and instruments. The Web orchestrate have ended up being principal specific instruments in business frames starting late. Using the Web by techniques for convenient devices constructs the potential results. The farming has some distinguishing strength in data progression. The ICT assembling in the development and essential drivers has been bankrupt down. For the profitable application key exercises need to get it. The new gadgets like tablets and new associations [4]. Conveyed processing have momentous potential in development. Passed on figuring gives better asset, he boards and reasonable cost control. By and by, the business assessment of these degrees of progress must be done essentially reliant on the advancement and expelled from its condition indiscriminately since the whole area is extraordinarily astounding.

II. LITERATURE SURVEY

In one of the approach, break down the segments related with the element off connection of Rural Expansion workers Data and Correspondence Innovation (ICT). It furthermore recognized the prerequisites to the usage of ICT to source cultivating knowledge and the data to the examination have been accumulated from the Topic Authorities and Extension Agents (EAs) of the Kwara State Cultivating Advancement Venture (KWADP) through the usage of a sorted-out study. The quantities of years at work and the age of the respondents had negative yet colossal relationship with access. Apart from the general prerequisites to the usage of ICT.

Manuscript published on 30 September 2019

* Correspondence Author

V Prithvi Ram*, Department of Information Science and Engineering, MSRIT, Bangalore, Karnataka, India.

Rajeshwari S B, Assistant Professor, Department of Information Science and Engineering, MSRIT, Bangalore, Karnataka, India.

Dr Jagadish S Kallimani, Associate Professor, Department of Computer Science and Engineering, MSRIT, Bangalore, Karnataka, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Smart Yield Accuracy Prediction using Linear Regression and Collaborative Filtration

For instance, stunning cost of PC, lacking force supply and poor web get to, poor getting ready and mechanical ability were moreover recognized as express objectives looked by the EAs. The examination along these lines endorsed the necessity for more straightforward accessible by all cultivating expansion workers in the ICT [5]. And also preparing workshop should be made with dealt with for Expansion officers in the zone of ICT and PC appreciation.

Bounteous research with driven on ICT motorization and applies autonomy in agribusiness and related bio linear issues. ICT and mechanical autonomy improvements are rapidly rising and can disturb future developing through their huge consequences for effectiveness and profitability. Tragically human cash related resources and attempts are isolated and confined. this incited the Period NET that gives the central sorted out structure. That is essential objective is to fortify and encourage European research concerning ICT and mechanical innovation in cultivating. Other than the generation of the meta Learning base, an average Europe research plan will be made and typical analyze calls are moved. That meet Information Base (<http://db-ictagri.eu>) is attempt to the outline appropriate creative work inside the picked research territory. To mastermind the postings, a three-dimensional assignment development arranged structure was arranged. The results demonstrated that the three tomahawks: errand, advancement and expansion gave off an impression of being missing to depict the whole research territory [6]. Thusly, an improved structure was made. By widening the errand advancement organized frame with a procedure control information structure, an accommodating framework was organized.

Mobile phones are the certain natural and support widespread learning the condition. In this paper arranging and the working up an adaptable courseware for ICT understudies using issue based on learning approach is discussed. the courseware is planned to survey the common sense of grasping the issue based getting the hang of training strategies in an adaptable learning condition for ICT understudies. A linear examination is worked for the java programming work is realized on the machine learning framework arranged. the meaning structure is made using organization arranged plan. That structure and the transport of the learning object for that flexible learning will being depicted in the PBL condition [7]. Agriculture is the main occupation of the all over the world includes a basic position in the improvement of that countries. Regardless of criticalness agribusiness, overhauls around there have been uneven and, when all is said in done, baffling. In any developing structure that is basic to see the various employments of women. Various women experience a real presence is that stunning trap of multiple employments and plays out different errands which requires the ordinary woman to guide one of a kind thing in an idea to fulfill the family needs the women in commonplace systems are generally connected with strenuous farm errands and country works out, from planting to procuring and other post gathering undertakings. So the Nigerian women are in a basic position to add to sustenance supply this portion faces huge troubles for improving age in a condition of diminishing normal re fundamental for creation. ICT plays a basic employment in keeping an eye on these challenges. The paper separates the issues going up against women in the cultivating section and prescribes ways to deal with deal with these issues [8]. Furthermore, the paper audits information needs to nation

women and how ICT can be used to meet their information needs.

Information related to corresponding advances headway is well advancing moors law desire for geometric development for the PC execution records. Truth be told, these advances are quick made just as, in addition, are considering a posterity to newer ones wonderfully extending existing "old style" instruments. These improvements of ICT are not simply recouping customary sciences, similar to Agribusiness, for example, developing, yet what's more, attentive all around disregarded human affectability and lack of concern for dejection, natural security, climatic disintegrating issues and the destiny of our planet with everything taken into account. Parallel Figuring passes on exponentially extended focus taking care of two low-end PCs empowering the usage of gigantic PC control by minimal agrarian research units [9]. Never the less country and developing systems, in their lion's offer, don't get new gadgets and the structures for the degree required for extensive provincial improvement. In this article expanded the quite a while is used to survey for the poor reason execution on the zone to the propriety to improvements and the instruments by the wide margin larger number of farmers all through the that the world.

In one of the approaches, expressed that the ARMA models demonstrated to be a more powerful time-arrangement models then the smoothing procedures specifically dampened-pattern direct exponential for Twofold exponential and Basic exponential for exponential to foreseeing harvest crop yield [10].

In another approach, its expressed as yield reproduction models are basic to bolster field inquire about and there are valuable especially yield expectation, horticulture arranging, ranch the board what's more, agriculture. The impact of the climatic changes on wheat improvement has additionally considered [11].

In this method, connected choice tree enlistment strategy on the dissecting the soybean profitability dependent because of different climatic parameters [12]. The end clients can be profited from the principles inferred throughout the discoveries of the Choice tree.

This method expresses that Utilizing field information a com yield reaction technique for llois yield announcing areas is anticipated. weather factors are arranged according to edit development organizes; this permits the utilization of model if environmental change swings dates of the yield arising seasons [13]. Results demonstrates that this model is better than typically indicated model in explaining yield deviation in llois com.

In this method its expressed that the coming about continuously every year precipitation profiles for the explicit examination territory in south west Agrarian zone of Australia was utilized to discover zones of most noteworthy rate of yield creation. the examined territories are spatially scaled to unmistakable shires the examples prescribed that harvest yield was firmly identified with the yearly precipitation for certain shires where area of the shire progresses toward becoming an and no more significant when contrasted and other shires [14].

In this work its expressed that a basic space state model aide in breaking down yield related time arrangement information for a few significant yields on the earth (example, wheat, maize, rice). The proposed model is a dynamic straight relapse design anticipating future yield patterns and their related validity interims utilizing the Kalman channel calculation .the accuracy of the high yield forecasts fit with their technique is esteemed utilizing wheat yield information given by Sustenance and Agribusiness Association of the Joined countries [15] .

it was demonstrated that, with these information the dynamic straight relapse model is increasingly adaptable and performs superior to the vast majority of the factual strategies presently used to assess harvest yield time arrangement information.

III. PROPOSED WORK

In this paper we have made some accuracy prediction of crop yields. To overcome the agriculture segment, farmers feel increasingly compelled to embrace escalated cultivating more yields in crop and supportable farming ones, so as to increment both monetary and natural expenses. Formers anticipating physically the enthusiasm of grains and vegetable so its effect to formers monetarily to beat this issue we are proposed man-made intelligence auto solicitation and yield figure process. The underneath figure demonstrates the framework engineering, which shows two significant modules, administrator and rancher. The administrator module fundamentally deals with the primary server which stores the dataset, examines utilizing legitimate data mining method which aides in giving the rancher right data according to the solicitation and subtleties refreshed by rancher. The rancher needs to enlist through the gateway giving the Aadhar subtleties and land subtleties and can demand for the contribution for developing the harvests.

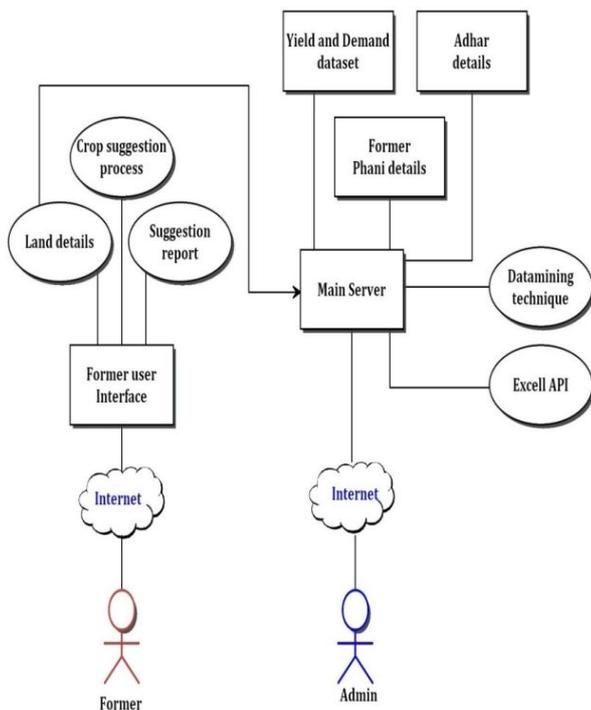


Figure 1: Proposed Model

Implementation work done based on the linear regression and collaborative filtration algorithms. The work flow is as shown in Figure 2.

Info: the expectation of the harvest dependent on various factors, for example, soil supplement, climate and past harvest generation to anticipate the yield precisely. These whole factors are area dependent and along these lines the area of client is taken as info framework.

Information securing: depending upon the present area the framework primary the dirt properties in the individual are from the dirt vault in a comparative methodology climate parameter are separated from the climate informational index.

Information alter: A yield can be obtained just it is pertinent situation met in these incorporate broad attributes partnered to soil and weather these imperatives are analyzed and the well-suited harvests are found out. Linear regression and collaborative filtering are utilized by the framework to the anticipate the yield. the expectation depends on past created information of harvests i.e.: distinguishing the substantial weather and soil attributes and comparing it to current conditions which will foresee the yield more precisely and in a reasonable way.

Yield: The most benefit of harvest is an anticipated on the framework utilizing linear regression (LR) calculation and the client is furnished with various recommendations of yield presenting to the length of harvest and to obtain the good yield in further prediction

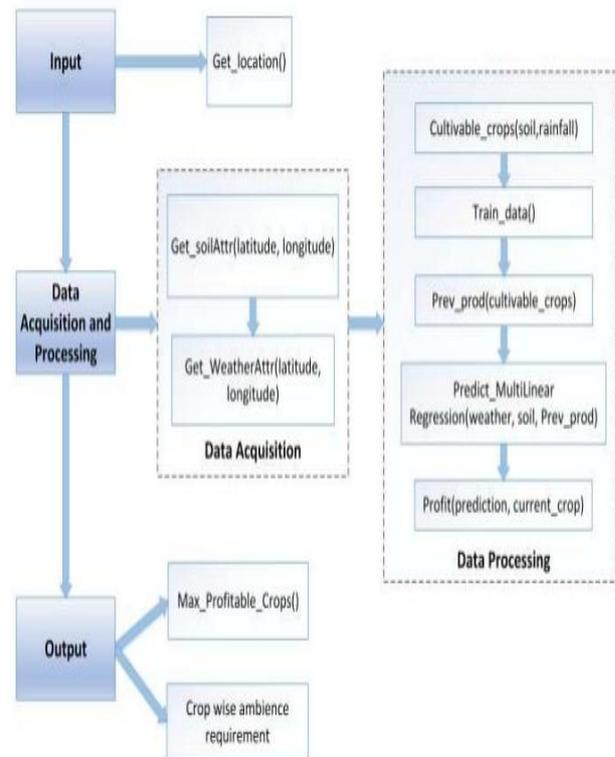


Figure 2: Proposed Work Flow

IV. IMPLEMENTATION OF PROPOSED WORKFLOW

Linear Regression Algorithm

```

Input: yield predict, crop yield;
Output: yield predict by linear regression
Initialization: MAXN = 1000, n = 0;
Predicted value=0.0;
x={2008,2009,2010,2011,2012,2013,2014,2015,2016,2017};
y=crop yield;
sum x = 0.0, sum y = 0.0, sumx2 = 0.0;
while (n<x.length)
{
    sum x += x[n];
    sum x2 += x[n] * x[n];
    sum y += y[n];
    n++;
}
if{
    x bar = sum x / n;
    y bar = sum y / n;
    // second pass: compute summary statistics
    xx bar = 0.0, yy bar = 0.0, xy bar = 0.0;
    for (int i = 0; i < n; i++) {
        xx bar += (x[i] - x bar) * (x[i] - x bar);
        yy bar += (y[i] - y bar) * (y[i] - y bar);
        xy bar += (x[i] - x bar) * (y[i] - y bar);
    }
    beta1 = xy bar / xx bar;
    beta0 = y bar - beta1 * x bar;
    // print results
    system.out.println("y = " + beta1 + " * x + " + beta0);
    year= Calendar.getInstance().get(Calendar.YEAR);
    predicted value=(beta1*(year))+beta0;
    // analyze results
    intdf = n - 2;
    rss = 0.0; // residual sum of squares
    sssr = 0.0; // regression sum of squares
    for (int i = 0; i < n; i++) {
        fit = beta1*x[i] + beta0;
        rss += (fit - y[i]) * (fit - y[i]);
        sssr += (fit - ybar) * (fit - ybar);
    }
    R2 = sssr / yybar;
    svar = rss / df;
    svar1 = svar / xxbar;
    svar0 = svar/n + xbar*xbar*svar1;
    svar0 = svar * sumx2 / (n * xxbar);
    return predicted value;
}
End if
    
```

V. RESULT AND DISCUSSION

The main role in this paper played by the linear regression model is generation in tonnage = $2354.6675 * \text{crop} = \text{coffe, custard, dry grapes, banana, maize, banana, berry ('000nuts),mesta ,nuts ,rice, surgarcane} + 365423765 * \text{crop} = \text{groundnut, banana ,coconut ('000 nuts),Mesta, cotton('000 nuts),ok ,radish (limt),rice, sunflower} + -52635423 * \text{crop-cotton(lint),rice, sunflower} + 115388.8649 * \text{harvest} = \text{coffee,surgarcane} + 4466464 * \text{yield} = \text{sugarcane} + 1.5454 * \text{zone in hectares.} + -55345565. \text{the accompanying condition speaks to just when the two properties territory total land area and yield they took for foreseeing for creation in tonnage.to fit the straight model,lm is utilized .further this can the utilized for completing examination of the difference and two indicator recipe that is demonstrated as follows}$

$lm(\text{recipe} = \text{Generation in -tons} \sim \text{Zone inHectares} + \text{Harvest, information} = \text{datavar})$
 at the point when these qualities are appointed for Region in Hectares and Harvest it will foresee the Creation in Tons
 The three indicator recipe is brought beneath 2mcrop production – in tonnage – total raea of land +crop grown+total span of data summed up straight model 'glm is utilized, what's more, this is point by point by giving an emblematic portrayal of the blunder dispersion and the depiction off the straight indicator.
 $Glm(\text{production-intonnage} \sim \text{Harvest} + \text{Territory in Hectares, family} = \text{gaussian, datavar})$

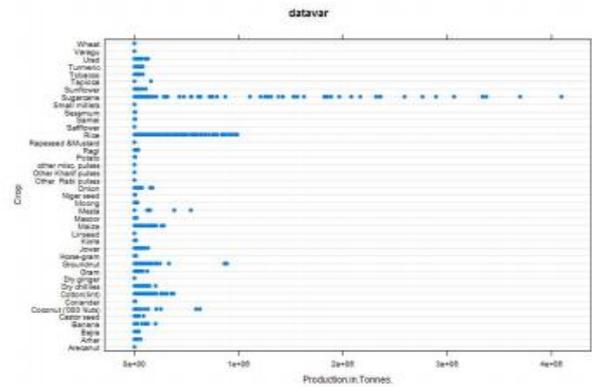


Figure 3: Chart of Yield versus Creation in tonnage.
 Generation in tonnage brought to the down part of the graph other more, crop is brought right part of the graph. Coffee has the most elevated yield and creation prize though ragi as the 2nd most noteworthy yield most security prize contrasted with every single other yield. The following most astounding yield and creation of harvests are to be specific grapes, mangoes, tea, onions, garlic and carrot cucumber and so forth.

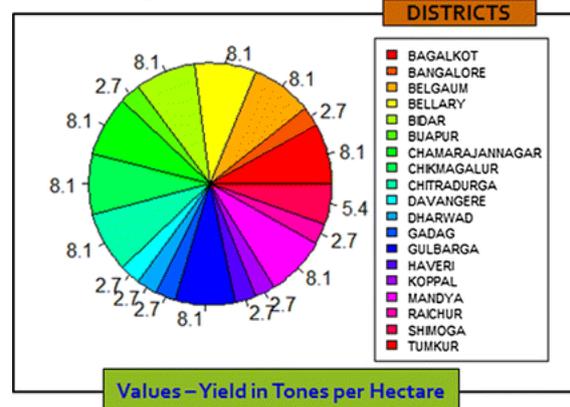


Figure 4: Diagram represent the overall crop yield in Karnataka district.

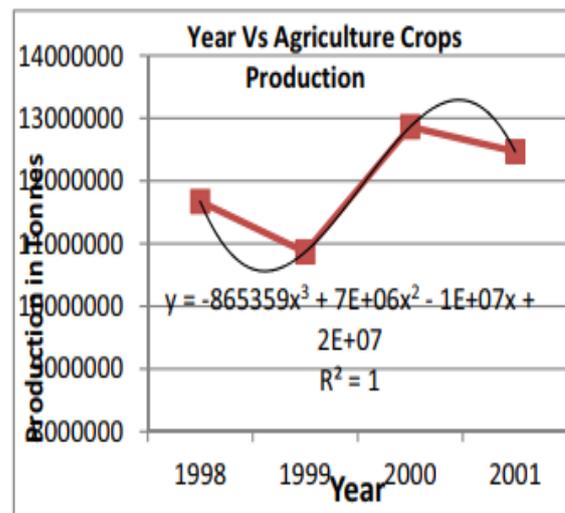


Figure 5: Karnataka state– Year versus Agricultural Harvests Generation in tonnage.

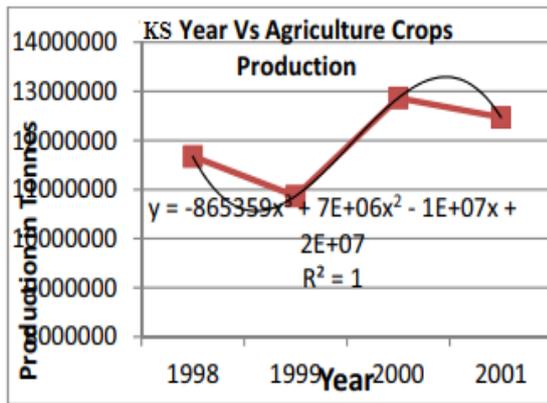


Figure 6: Karnataka state – Year versus Farming

Yields Generation in tonnage.

Here in the discussion we predict and analysis the crop yields as per the data available in the previews year. Figure 3 says the yield of maize is 10.18, maize total predicted yield is 31.0, demand prediction per year is 549.272 and finally based on the approaches we followed and predicted is as followed: Maize linear regression accuracy is :51% Maize collaborative filtering is:454.0 Maize collaborative accuracy is 13% These processes take place for the figure 4 also.



Figure 7: Report of Maize Crop



Figure 8: Report of Carrot Crop



Figure 9: Report of Ragi Crop

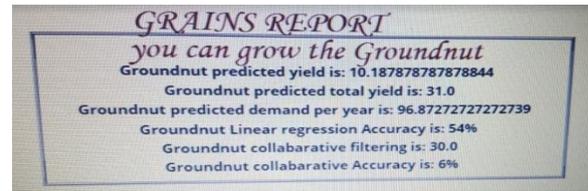


Figure 10: Report of Groundnut Crop

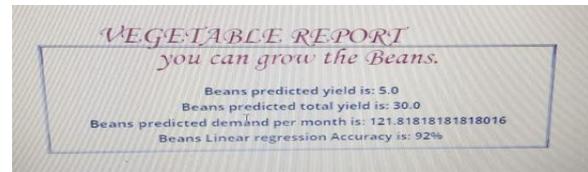


Figure 11: Report of Beans Crop

Experimental Performance Analysis

In this session we deal with the accuracy prediction by comparing the linear regression and collaborative filtering algorithm. crop yield forecast is basic to useful comment of the greatly collected organic successions, which prompts a basic requirement for the improvement of high return crops, as a first and key advance in growing and predication became an increasingly challenging task.

Table:1 Comparison Accuracy In 2015

Crops	Prediction Accuracy In %	
	Linear Regression	Collaborative Filtering
Maize	30	15
Carrot	40	20
Ragi	50	25
Ground Nuts	60	30
Beans	78	34

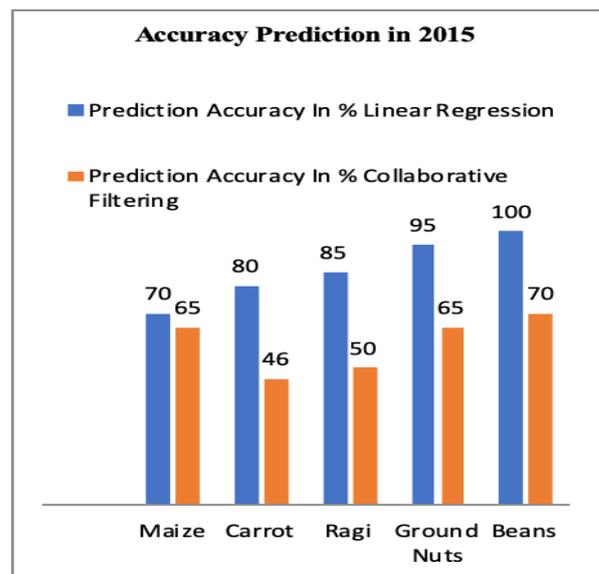


Figure 12: Prediction Process in 2015

Table 2: Comparison Accuracy in 2016

Crops	Prediction Accuracy In %	
	Linear Regression	Collaborative Filtering
Maize	35	20
Carrot	45	22
Ragi	55	35
Ground Nuts	65	38
Beans	75	40

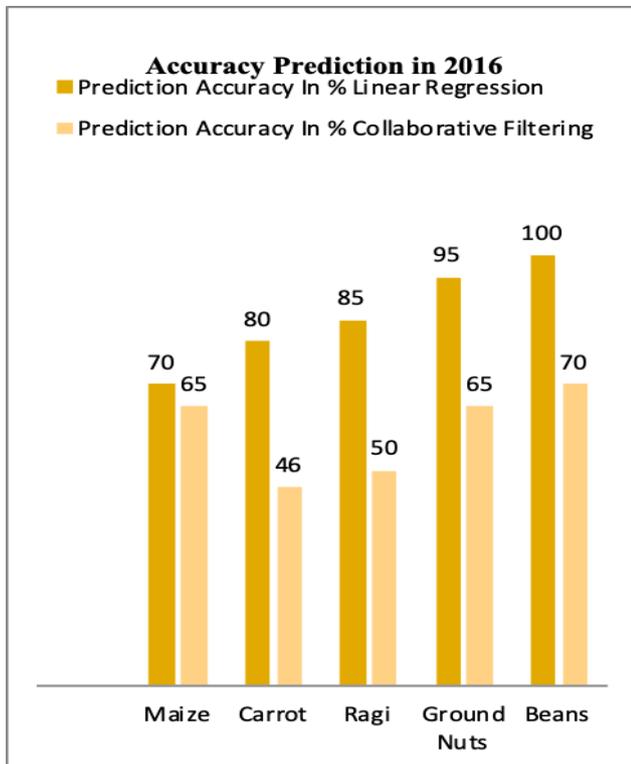


Figure 13: Prediction Process in 2016

Table 3: Comparison Accuracy in 2017

Crops	Prediction Accuracy In %	
	Linear Regression	Collaborative Filtering
Maize	40	30
Carrot	45	40
Ragi	65	45
Ground Nuts	75	50
Beans	80	60

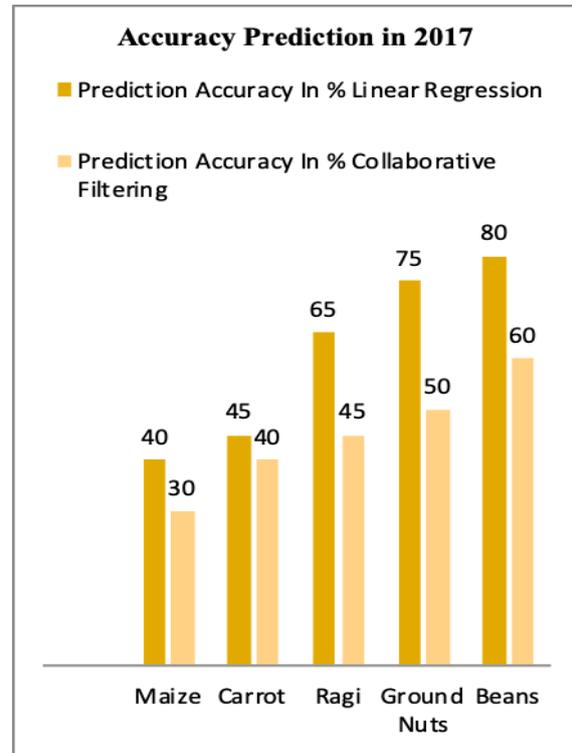


Figure 14: Prediction Process in 2017

Table 4 Comparison Accuracy in 2018

Crops	Prediction Accuracy In %	
	Linear Regression	Collaborative Filtering
Maize	60	35
Carrot	75	46
Ragi	85	50
Ground Nuts	95	55
Beans	90	60

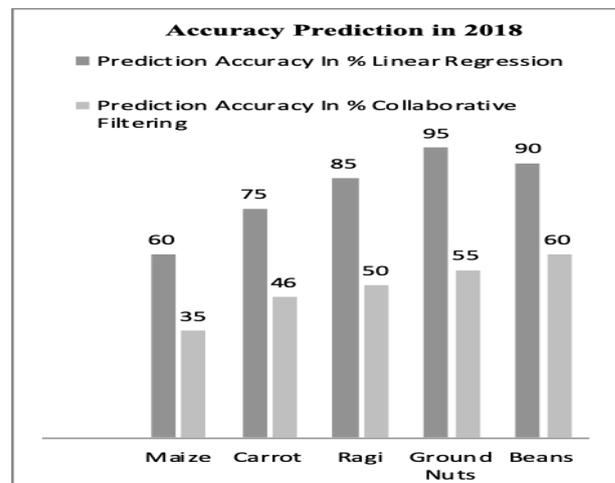


Figure 15: Prediction Process in 2018

Table 4: Comparison Accuracy in 2019

Crops	Prediction Accuracy In %	
	Linear Regression	Collaborative Filtering
Maize	70	65
Carrot	80	46
Ragi	85	50
Ground Nuts	95	65
Beans	100	70

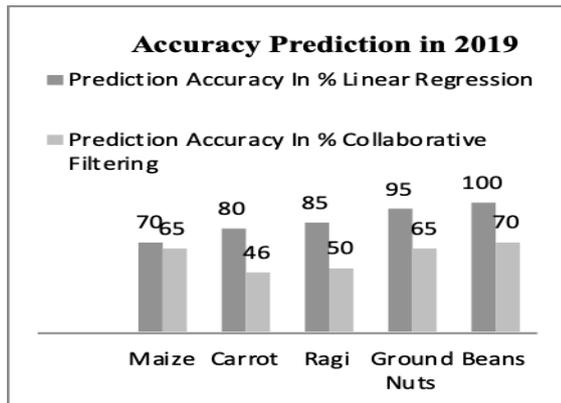


Figure 16: Prediction Process in 2019

VI. CONCLUSION

The proposed framework considers over the information identified with area climate and past few span generations and recommends which are in more gainful harvests which can be developed in the agriculture land natural condition. As the framework records out every conceivable harvest, it helps the agriculturists in basic leadership of which yield to develop. Likewise, this framework takes in to think for few generations of information which will enable the agriculturists to get understanding into the interest and the expense of different crops in market. As greatest kinds of harvests will be secured under this framework, agriculturist may become more acquainted about the yield which may never been developed.

VII. FURTHER ENHANCEMENT

In our further enhancement we can process that all farming equipment’s can be associated over the web utilizing IOT. The sensors can be utilized in land which will gather the data about the present land conditions and can improve crop yield by future prediction of linear regression which will helps in getting with good crop yield. In upcoming days, we can get more yield crop.

VIII. DECLARATION

I would like to declare my thanks to our beloved guides Ms. Rajeshwari S.B and Dr. Jagadish S Kallimani for being a great source of inspiration and extending facilities for the successful completion of this project.

REFERENCE

1. JeetendraShenoy, YogeshPingle, “IOT in agriculture”, 2016 IEEE.
2. M.R. Bendre, R.C. Thool, V.R.Thool, “Big Data in Precision agriculture”,Sept,2015 NGCT.
3. Monali Paul, Santosh K. Vishwakarma, Ashok Verma,“Analysis of Soil Behavior and Prediction of Crop Yield using Data Miningapproach”, 2015 International Conference on Computational Intelligence and Communication Networks.
4. Abdullah Na, William Isaac, ShashankVarshney, EkramKhan, “An IoT Based System for Remote Monitoring of Soil Characteristics”, 2016International Conference of Information Technology.
5. Dr.N.Suma, Sandra Rhea Samson, S.Saranya,G.Shanmugapriya, R.Subhashri, “IOT Based Smart Agriculture MonitoringSystem”, Feb 2017 IJRITCC.
6. S. Veenadhari Dr. Bharat Mishra Dr.CD Singh, Soybean Productivity Modelling using Decision Tree Algorithms, Vol.27– No.7, August 2011, International Journal of Computer Applications (0975 – 8887).
7. Bruce L. Dixon, Steven E. Hollinger, Philip Garcia, and Viswanath Tirupattur, Estimating Corn Yield Response Models to Predict Impacts of Climate Change, Journal of Agricultural and Resource Economics, 19(1): 58-68.
8. Yunous Vagh, An Investigation into the Effect Of Stochastic Annual Rainfall on Crop Yields in South Western Australia, International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012.
9. David Makowski, Lucie Michel, Use of dynamic linear model for predicting crop yield trends in foresight studies on food Security, ICAS VI Sixth International conference on Agriculture Statistics, 2013.
10. N.Heemageetha, “A survey on Application of Data Mining Techniques to Analyze the soil for agricultural purpose”,2016IEEE.
11. DhivyaB ,Manjula , Siva Bharathi, Madhumathi, “A Survey on Crop Yield Prediction based on Agricultural Data”, International Conference in Modern Science and Engineering, March 2017.
12. Giritharan Ravichandran, Koteeshwari R S “Agricultural Crop Predictorand Advisor using ANN for Smartphones”,2016 IEEE,
14. R.Nagini, Dr. T.V. Rajnikanth, B.V. Kiranmayee, “Agriculture Yield Prediction Using Predictive Analytic Techniques, 2nd International Conference on Contemporary Computing and Informatics (ic3i),2016
15. Awanit Kumar, Shiv Kumar, “Prediction of production of crops using K-Means and Fuzzy Logic”, IJCSMC, 2015