Supporting of Crop Yield prediction using Machine Learning Algorithm Techniques

# Dr.S.Vasundhara

**Asst Professor of Mathematics**

# G. Narayanamma Institute of Technology & Science Shaikpet

**Hyderabad**

# Abstract:

In India Agriculture plays a vital role and also plays an important in Indian economy growth. Crop production was carried out based on knowledge passes from generation to generation. Due to huge growth in population human based cultivation is not enough to meet the demanding need. Machine learning is an important decision support tool for crop yield prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Several machine learning algorithms have been applied to support crop yield prediction research.

This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning techniques, one of the advanced technologies in crop prediction. Naive Bayes, a supervised learning algorithm puts forth in the way to achieve it. The seed data of the crops are collected here, with the appropriate parameters like temperature, humidity rainfall N, P,K and PH levels, which helps the crops to achieve a successful growth.

Key words: Agriculture, Crop Suggestion system, MachineLearningAlgorithms, Farming, Prediction. Linear Regression, Soil, Crop yield prediction.

Introduction:

The History of Agriculture in In India dates back to the Indus Valley Civilization .In India ranks second worldwide in farm outputs. As per 2018 regulations agriculture employed more than 50% of the Indian work force and contributed 17-18% to country’s GDP. The crop yield is the significant factor contributing in agricultural monetary. The crop yield depends on multiple factors such as climate geographic, organic, and financial elements, it is difficult for farmers to decide when and which crops to plant . To feed the rapidly growing global population, modern-day agriculture faces the demand for rising production of food. Hence, the latest technologies are transpiring in the agricultural sector to enhance net productivity by gathering and processing information. Besides, the distressing climate changes have also been hinted at the inevitable demand for modernizing the agriculture domain with the latest tools and technologies. Therefore, in the modern era, agricultural and farming domains are adapting and applying state-of-the-art

technologies, namely, machine learning. On the other hand, agriculture remains the single most important avenue for mankind, and therefore in most countries, the largest part of the workforce is in some way involved in this sector being one of the most densely populated countries and one of the fastest-growing economies in the world, smart agriculture can have significant importance to grow rapidly. Hence, the objective of this research is to explore the possibilities of using a modern technology-driven prediction model to assist the farmers in efficiently selecting crops and maximizing the yield potentialities.

**Precision agriculture** (**PA**) is a farming management concept based on observing, measuring and responding to inter and intra-field variability in crops. The goal of precision agriculture research is to define a decision support system (DSS) for whole farm management with the goal of optimizing returns on inputs while preserving resources.

The practice of precision agriculture has been enabled by the advent of GPS and GNSS. The farmer's and/or researcher's ability to locate their precise position in a field allows for the creation of maps of the spatial variability of as many variables as can be measured (e.g. crop yield, terrain features/topography, organic matter content, moisture levels, nitrogen levels, pH, EC, Mg, K, and others). Similar data is collected by sensor arrays mounted on GPS-equipped combine harvesters. These arrays consist of real-time sensors that measure everything from chlorophyll levels to plant water status, along with multispectral imagery. This data is used in conjunction with satellite imagery by variable rate technology (VRT) including seeders, sprayers, etc. to optimally distribute resources. However, recent technological advances have enabled the use of real-time sensors directly in soil, which can wirelessly transmit data without the need of human presence.

Precision agriculture has also been enabled by unmanned aerial vehicle that are relatively inexpensive and can be operated by novice pilots. These agriculture drones can be equipped with multispectral or RGB cameras to capture many images of a field that can be stitched together using photogrammetric methods to create orthophotos. These multi spectral images contain multiple values per pixel in addition to the traditional red, green blue values such as near infrared and red-edge spectrum values used to process and analyze vegetative indexes such as NDVI maps. These drones are capable of capturing imagery and providing additional geographical references such as elevation, which allows software to perform map algebra functions to build precise topography maps. These topographic maps can be used to correlate crop health with topography, the results of which can be used to optimize crop inputs such as water, fertilizer or chemicals such as herbicides and growth regulators through variable rate applications.

# MACHINE LEARNING TECHNIQUES:

Machine learning is commonly used in conjunction with drones, robots, and internet of things devices. It allows for the input of data from each of these sources. The computer then processes this information and sends the appropriate actions back to these devices. This allows for robots to deliver the perfect amount of fertilizer or for IoT devices to provide the perfect quantity of water directly to the soil. Machine learning may also provide predictions to farmers at the point of need,

such as the contents of plant-available nitrogen in soil, to guide fertilization planning. As more agriculture becomes ever more digital, machine learning will underpin efficient and precise farming with less manual labour.

Machine Learning involves problems in which the input and output relationship is not known. Learning specifies the automatic acquirement of structural descriptions. In contrast to traditional statistical methods, machine learning does not make assumptions about the exact construct of the data model, which describes the data. This feature is very helpful to describe complex non-linear behaviors such as a crop yield prediction. Machine learning is a part of artificial intelligence employed to build an intelligent system . By utilizing the training samples, the test samples can be identified. The accuracy of the system can be measured using metrics such as mean square error, root mean square error, precision, recall, sensitivity specificity etc. Further, machine learning can be employed to address a variety of applications including crop yield prediction through supervised, unsupervised and reinforcement learning methods. Classification, clustering, regression, prediction are some of the techniques involved to attain the intelligent system.

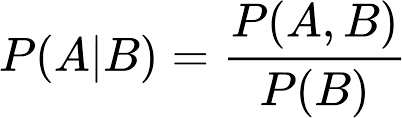
# Linear Regression:

Prediction based on linear regression is discussed in many works. It is a statistical method applied over linear systems. Using this, the relationship between dependent and independent variables can be measured. If independent variable is having more than one input attribute, multiple regressions can be applied. Regression based models are used in prediction as this technique shows consistent results during standard tests. Even though regression based models work fine for linear data, they do not fit for complex and non-linear data. Also, these models may not be able to perform better due to its limitation with regression assumptions for multiple co-linearity among the dependent and independent variables. Regression Models were used to Predict Crop Yield. The results revealed that most of the observed events were correctly predicted. An equation for predicting the yield of label from the climate variables is presented. Analysis of crop yield prediction using Multiple Linear Regression (MLR) technique and Density based clustering technique used.

**Naïve Bayes**: Naïve Bayes algorithm is a type of supervised learning algorithm. It is based on Bayes theorem. It is used for solving classification problems. Text classification involves using Naïve Bayes which includes high dimensional training dataset. It is one of the simple and most effective algorithm for classification. It also helps in building fast machine learning models that can make quick predictions. It is used to predict on the basis of probability of an object. Hence it is known as probabilistic classifier. It is called so because it assumes the occurrence of a certain feature is independent of the occurrence of other features. For example if a fruit is identified on the bases of colour, shape and taste, then red, spherical and sweet fruit is recognised as apple. So here each feature is individually used to identify that it is an apple independently. Since it depends on the principle of Bayes theorem it is called as Bayes. Naïve Bayes can also be used for binary and Multi-class Classifications.

**Bayes' Theorem**: Bayes' theorem is used to determine the probability of hypothesis with prior knowledge and it is also known as Bayes’ Rule or Bayes’ law. It depends on the conditional probability.

The formula foe Bayes’ theorem is,



Where,

P(A|B) is the Posterior probability. P(B|A) is the Likelihood probability. P(A) is the Prior Probability.

P(B) is the Marginal Probability.



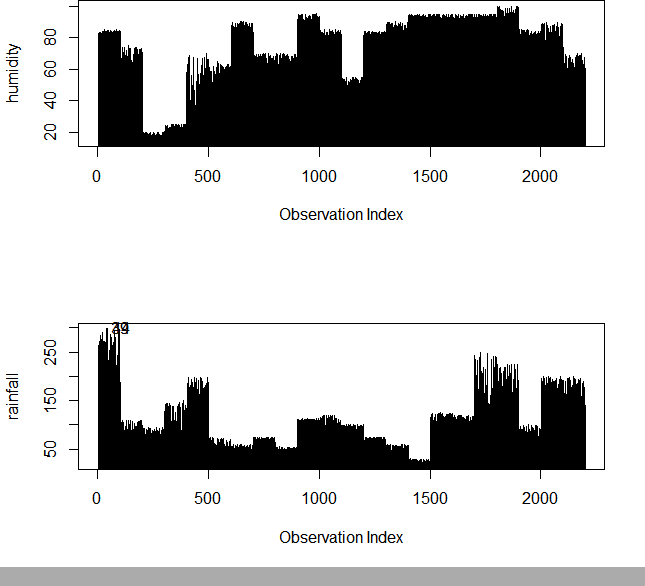
* N - ratio of Nitrogen content in soil
* P - ratio of Phosphorous content in soil
* K - ratio of Potassium content in soil
* Temperature- temperature in degree Celsius
* Humidity- relative humidity in %
* ph - ph value of the soil
* rain fall - rainfall in mm

# Literature Review:

Research proposed by Dahikar and Rode in used parameters related to soil and atmosphere to predict essential crops using an artificial neural network. Areal image detection-based agriculture using machine learning was proposed by Treboux and Genoud in. It showed that the decision tree ensemble outperforms other image identification-based solutions. A paper by Shakoor et al. proposed a crop rank that suggests a list of cost-effective crops for cultivation .

This paper used data collected from kaggle web site supervised machine learning algorithms to generate the crop list. Similar predictive approaches in agriculture domain have been demonstrated by Wang et al. in. The paper proposed the usage of statistical models for yield prediction in response to the change in temperature and perception. Three different statistical models, namely, time series, panel, and cross-sectional models, were used to predict climate change based on the given data. Similar proposals have been presented in.

Shastry and Sanjay presented that the customized artificial neural network (C-ANN) model performs better with a higher R2 statistic and the least percentage prediction error than the MLR and D-ANN models on the test dataset. Furthermore, the prediction of crop yield is very essential in the domain of agriculture. In this study, the wheat yield was predicted by considering its different parameters, and better wheat yield was predicted by applying the C- ANN model.



lm(formula = rainfall ~ humidity + ph + temperature, data = Dataset) Residuals:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |  |
| -89.995 -  Coefficients: | 39.892 | -7.649 |  | 22.395 | 196.649 |

|  |
| --- |
|  |
| Estimate Std. Error t value Pr(>|t|) |
| (Intercept) 150.03334 11.76903 12.748 < 2e-16 \*\*\* |
| humidity 0.25790 0.05319 4.849 0.00000133 \*\*\* |
| ph -7.74982 1.49778 -5.174 0.00000025 \*\*\* |
| temperature -0.58040 0.23389 -2.481 0.0132 \* |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 54.35 on 2196 degrees of freedom

Multiple R-squared: 0.02338, Adjusted R-squared: 0.02204

F-statistic: 17.52 on 3 and 2196 DF, p-value: 3.05-11

# multinom(formula = label ~ N + P + rainfall + K, data = Dataset,

trace = FALSE) Coefficients:

pomegranate 2.0723499 0.03858916 0.07262034 0.01863421 0.04135575

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| rice | 2.6621265 0.03762160 0.06695340 0.01683411 0.06892945 (Intercept) | | N | P |
| rainfall | K | |  |  |
| banana | | -37.382581 0.34901118 0.526269606 -0.056086579 -0.245675144 | | |
| blackgram | | 1.369916 0.06177616 0.404651431 -0.099508009 -0.353511407 | | |
| chickpea | | 15.837260 0.03337300 -0.007931271 -0.096764057 -0.043988937 | | |
| coconut | | 7.504307 -0.09572762 -0.440246244 0.062964264 0.109278844 | | |
| coffee | | -8.596309 0.36075608 -0.238213485 0.030144647 -0.248415108 | | |
| cotton | | -11.117793 0.38258723 0.139570992 -0.102202846 -0.269121256 | | |
| grapes | | 20.437088 0.01248036 -0.012227691 -0.290718715 0.027838242 | | |

jute -7.649953 0.22555312 0.020456416 0.005936672 -0.082576365

|  |  |
| --- | --- |
| maize | 10.834358 0.17311085 0.101104984 -0.075179496 -0.359655975 |
| mango | 27.079695 -0.04793965 -0.169579461 -0.109221551 -0.053617036 |
| mothbeans | 25.985583 -0.03329907 0.046338551 -0.169881187 -0.176752254 |
| mungbean | 26.021031 -0.03685727 0.044432386 -0.176459827 -0.167483005 |
| muskmelon | 15.371433 0.17260809 -0.071587674 -0.251453982 -0.004033961 |

orange 25.170417 0.02243435 -0.172515775 -0.024762801 -0.484942488

papaya 5.080354 0.05319049 0.120644153 -0.021676818 -0.128763360

pigeonpeas -11.544206 -0.03829858 0.499391811 0.009460744 -0.414099287

pomegranate 20.629493 -0.09256941 -0.467597349 -0.061427897 0.202656681

rice -24.048897 0.26958936 -0.058662153 0.105227814 -0.170929089

watermelon 2.849983 0.20812306 -0.196499374 -0.057750651 0.070137811

Std. Errors:

(Intercept) N P rainfall K

|  |  |
| --- | --- |
| banana | 0.3228813 0.05505061 0.11510064 0.04091112 0.07245975 |
| blackgram | 1.8379466 0.03667023 0.06667939 0.01449431 0.04635505 |
| chickpea | 1.4810536 0.03393608 0.05423056 0.01470234 0.03330310 |
| coconut | 2.5367694 0.04588786 0.08116619 0.01786379 0.06982299 |
| coffee | 2.7831264 0.04090030 0.07403263 0.01606197 0.06546043 |
| cotton | 3.7126461 0.04381403 0.06857903 0.02270906 0.05635346 |
| grapes | 4.5477121 0.04395770 0.07196937 0.05914746 0.04185704 |
| jute | 2.1127950 0.03740123 0.06096342 0.01179041 0.04790403 |

kidneybeans 2.0504527 0.03845105 0.06836078 0.01253802 0.04929947

lentil 1.8126204 0.03765383 0.06757916 0.02144053 0.04203887

maize 2.3039993 0.03819382 0.06360826 0.01706200 0.05961658

mango 1.6673948 0.03705293 0.06270755 0.01579605 0.03890605

mothbeans 1.4533674 0.03639309 0.05991961 0.01659790 0.03953105

mungbean 1.4956980 0.03669368 0.06032778 0.01732601 0.03984896

muskmelon 1.6792088 0.04059221 0.06752078 0.02579490 0.06397822

orange 2.6974704 0.04646133 0.07814834 0.02357571 0.08328525

papaya 1.4003668 0.03451394 0.05609211 0.01030303 0.03570124

pigeonpeas 2.4812953 0.04038665 0.07108730 0.01316948 0.05725702

watermelon 2.0017355 0.03957569 0.06838703 0.01582482 0.04718259

Value/SE (Wald statistics):

|  |  |  |  |
| --- | --- | --- | --- |
| (Intercept) N P rainfall | | | K |
| banana -115.7780840 6.3398246 4.5722559 | | | -1.3709374 -3.3905050 |
| blackgram 0.7453513 1.6846407 6.0686132 | | | -6.8653154 -7.6261681 |
| chickpea 10.6932385 0.9834077 -0.1462509 | | | -6.5815435 -1.3208660 |
| coconut 2.9582140 -2.0861207 -5.4240099 | | | 3.5246863 1.5650840 |
| coffee | -3.0887239 | 8.8203768 -3.2176824 1.8767716 -3.7948897 | |
| cotton | -2.9945739 | 8.7320707 2.0351846 -4.5005322 -4.7755938 | |
| grapes | 4.4939274 | 0.2839176 -0.1699013 -4.9151510 0.6650790 | |
| jute | -3.6207741 6.0306336 0.3355523 0.5035170 -1.7237874 | | |

kidneybeans -1.3540573 -1.1180133 6.2164039 -2.7168131 -6.7030894

|  |  |
| --- | --- |
| lentil | 4.8053717 -0.9242691 6.3846412 -11.1754393 -6.4944973 |
| maize | 4.7024140 4.5324310 1.5894946 -4.4062530 -6.0328181 |
| mango | 16.2407219 -1.2938153 -2.7042908 -6.9144859 -1.3781156 |

mothbeans 17.8795689 -0.9149834 0.7733453 -10.2351005 -4.4712254

mungbean 17.3972494 -1.0044584 0.7365161 -10.1846748 -4.2029450

muskmelon 9.1539737 4.2522469 -1.0602317 -9.7482063 -0.0630521

orange 9.3311190 0.4828606 -2.2075425 -1.0503523 -5.8226698

papaya 3.6278738 1.5411305 2.1508220 -2.1039269 -3.6066916

pigeonpeas -4.6524918 -0.9482980 7.0250499 0.7183841 -7.2322881

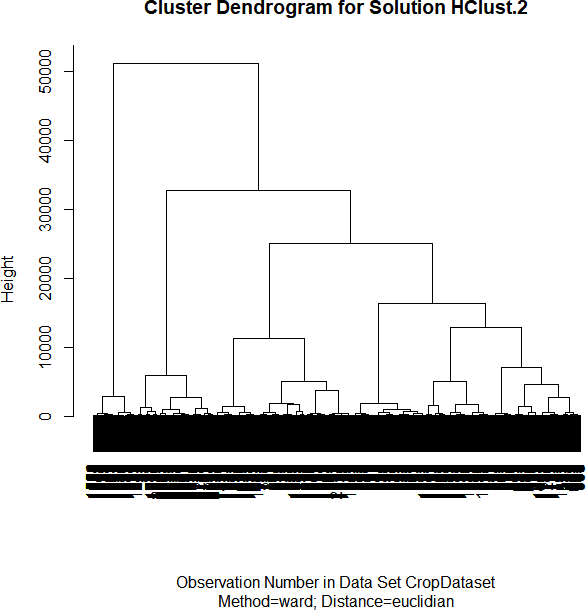
pomegranate 9.9546380 -2.3988448 -6.4389305 -3.2965115 4.9003268

rice -9.0337170 7.1658132 -0.8761639 6.2508681 -2.4797688

watermelon 1.4237561 5.2588616 -2.8733429 -3.6493715 1.4865189

Cluster Analysis:





K-NEAREST NEIGHBOR: K-Nearest Neighbour can be used for both classification and regression. K-Nearest Neighbors is a non-complex algorithm which stores all the available cases and classifiers new cased based on some similarity measure. The sampled set is classified based upon the ”closeness” that is the distance measure such as Euclidean distance or Manhattan distance.

Machine Learning Algorithm for Prediction:

Machine learning predictive algorithms has highly optimized estimation has to be likely outcome based on trained data. Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.

In our system we used supervised machine learning algorithm having subcategories as classification and regression .Classification algorithm will be most suitable for our system.

* Rainfall prediction: - Crop prediction:

Then fitting the classifier into training set. Radial basis function can be written in the mathematical form in equation (1),

(𝑋1, 𝑋2) = 𝑒𝑥𝑝𝑜𝑛𝑒𝑛𝑡 (−𝛾||𝑋1 −𝑋2|| 2 ) …. ..(1)

Where, ||X1-X2|| = Eucliden distance between X1 & X2 𝛾 = Gamma after fitting and test the model, at last this model will predict the appropriate annual rainfall. The predicted rainfall is the one of the input parameter for the crop prediction system.

Crop Prediction: Crop prediction process being with the loading the external crop datasets. Once the dataset read then pre-processing will be done by various stages as discussed in Data Pre- processing section. After the data pre-processing, train the models using Decision tree classifier into training set. For a prediction of the crop, we consider a various factor such as temperature, humidity, soil PH and predicted rainfall. Those are the input parameter for a system that can be entered by manually or taken from the sensors. Predicted rainfall and input parameter values will be appended in a list.

Based on predicted rainfall, soil contents and weather parameters the system will recommend the most suitable crop for cultivation. This system also provides details about required fertilizers like Nitrogen(N), Phosphorus (P) and potassium(K) in Kg per hectare and display the required seed for a cultivation in Kg per acre for recommended crop. This system as contain some other feature such as display the current market price and approximated yield in quintal per acre for recommended crop. Those all details will helps to farmers for choosing the most profitable crop. IV. EXPERIMENTAL OUTCOME The proposed system recommends the best suitable crop for particular land by considering parameters as annual rainfall, temperature, humidity and soil pH.

Among these parameters annual rainfall is predicted by system itself by using previous year data with Supervised machine learning algorithm and other parameters are have to be entered by the user. In the output section the system displays a suitable crop, required seeds/acre, market price and approximate yield of the recommended crop and also the system takes NPK values in the input section to display the required NPK for the recommended crop. The overall output is shown with Graphical user interface as shown in figure (4), Figure 4. Overall output with GUI We tested the system for various data set that has been collected from the different farmers for their lands conditions. Land condition’s just like the land having different pH, humidity and NPK values. The predicted annual rainfall is constant Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income. To reduce those type of loses we have developed a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the farmers to take right decision in selecting the crop for cultivation such that agricultural sector will be developed by innovative idea. FUTURE SCOPE: We have to collect all required data by giving GPS locations of a land and by taking access from Rain forecasting system of by the government, we can predict crops by just giving GPS location. Also, we can develop the model to avoid over and under crisis of the food. International Journal of Engineering Research & Tech

# EXPERIMENTAL OUTCOME:

The proposed system recommends the best suitable crop for particular land by considering parameters as annual rainfall, temperature, humidity and soil pH. Among these parameters annual rainfall is predicted by system itself by using previous year data with SVM algorithm and other parameters are have to be entered by the user. In the output section the system displays a suitable crop, required seeds/acre, market price and approximate yield of the recommended crop and also the system takes NPK values in the input section to display the required NPK for the recommended crop.

# CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENTS:

Agriculture is a major input sector for economic development of our country. Traditional agricultural system suffers from many problems that are caused due to ill soil health. The proposed system provides a very useful Android based application that facilitates an interface for farmers to monitor soil conditions and generate recommendations to improve crop production. The model includes interfaces for crop selection, maintenance of soil moisture and nutrient contents during plant growth. The modules are implemented using main components such as sensors and the Arduino, and the algorithms that are used in this system. The developed system is user friendly and gives proper suggestions. Automated crop recommendation system has tremendous demand and potential for the future too. It is time-saving, leading to the elimination of human error in

adjusting the available levels of soil moisture and optimizing net profits in terms of factors such as market price, product quality and production. It involves the additional sensors and outputs implemented by the system at an enhanced degree of smart interventions.

References:

1. Biradar, H. B., & Shabadi, L. (2017). Review on IOT based multidisciplinary models for smart farming.
2. Mehta, P., Shah, H., Kori, V., Vikani, V., Shukla, S., & Shenoy, M. (2015). Survey of unsupervised machine learning algorithms on precision agricultural data. IEEE Xplore.
3. Sandhiya, D., & Thiyaneswaran, B. (2017). Extraction of dorsal palm basilic and cephalic hand vein features for human authentication system. 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE.
4. Sanja Brdar, Dubravko Culibrk, Branko Marinkovic, Jovan Crnobarac and Vladimir Crnojevic, “Support Vector Machines with Features Contribution Analysis for Agricultural Yield Prediction”, in the Proc. of Second International Workshop on Sensing Technologies in Agriculture, Forestry and Environment, 2011. 5.D.S.Zingade, Omkar Buchade, Bilesh Mehta, Shubhm Ghodekar, Chandan Mehta, “Crop prediction System using Machine Learning”, International Journal of Advance Engineering and Research Development, vol. 4, Special Issue 5, Dec 2017, pp. 1-6

6.Narayanan Balakrishnan, Dr. Govindarajan Muthukumarasamy, “Crop production- ensemble machine learning model for prediction, International Journal of Computer Science and Software Engineering, vol. 5, Issue

7, July 2016, pp. 148-153 [25] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L, “Prediction of Crop Yield using Machine Learning”, International Research Journal of Engineering and Technology, vol. 5, Issue 2, Feb-2018, pp.2237-2239 [26] Athmaja S., Hanumanthappa M, “Applications of Mobile Cloud Computing and Big data Analytics in Agriculture Sector: A survey”, October 2016.

1. Kenya Betty J., Sitienei , Shem G. Juma, and Everline Opere,”:On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County”, MDPI Sensors, 2017
2. D Ramesh and B Vishnu Vardhan,”Analysis Of Crop Yield Prediction Using Data Mining Techniques”, International Journal of Research in Engineering and Technology, 2015, 4(1)

10 K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: a review,” *Sensors*, vol. 18, no. 8, p. 2674, 2018.

1-7682-8347