Rev. Fac. Agron. (LUZ). 2018, 35: 128-138. Enero-Marzo.

Geographical Information Based Crop Yield Prediction Using Machine Learning

Predicción del rendimiento de cultivos basada en la información geográfica mediante el aprendizaje automático

Previsão de rendimento de culturas baseadas em informações geográficas usando o aprendizado de máquina

Punithavathi. R * and Kalaavathi. B

[1] Professor and Head, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, India

[2] Professor and Head, Department of Computer Science and Engineering, KSR Institute for Engineering and Technology, Tiruchengode, Namakkal, India

*Corresponding Author

Abstract

Indian agriculture is facing a huge, challenging task of producing 480 Mt by the year 2050 to meet out the food grain requirement of the population. However the problems are plenty like biotic and abiotic stresses experienced by crops, soil degradation, erratic rainfall, climate change scenarios, emergence of new pest and diseases etc. Yield prediction is one of the most critical issues faced in the agricultural sector, which will help the government to device suitable planning strategies to improve the productivity and to create a good market price for the poor farmers. If application of soft computing on geographical information to improve the life of farmers, it will be useful to our nation. In this study we attempted the modern science & technology tool of machine learning and developed a model as Supervised Learning Algorithm for Crop Yield Prediction (SLACYP) to predict crop yield rate. The variables considered in this model are area under cultivation (per Hectares), Rainfall (mm), Consumption of fertilizers (MT) viz., Nitrogen(N), Phosphorous(P), Potassium(K). The model was designed and developed the algorithm using R Programming to predict the popularly cultivated crops in the selected region of Tamil Nadu, India for crops like Paddy, Maize, Turmeric, Groundnut, and Sugarcane. The R2 value for all the crops are greater than 80% indicating the best fit model for the crops. Among the crops, the multiple regression equation of paddy is found to be highly significant with the R2 value of 0.95 and the lowest value was found to be with the sugarcane with the R2 value of 0.81.

Keywords: Geographical Information, Machine learning, Crop yield, Prediction, Regression, RainfallandFertilizers

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

Resumen

La agricultura india se enfrenta a una enorme y desafiante tarea de producir 480 Mt para el año 2050 para cumplir con los requisitos de grano alimentario de la población. Sin embargo, los problemas son similares a los estreses bióticos y abióticos experimentados por los cultivos, la degradación del suelo, las precipitaciones irregulares, los escenarios de cambio climático, la aparición de nuevas plagas y enfermedades, etc. La predicción del rendimiento es uno de los problemas más críticos del sector agrícola, que ayudará el gobierno debe diseñar estrategias de planificación adecuadas para mejorar la productividad y crear un buen precio de mercado para los agricultores pobres. Si aplicamos soft computing a la información geográfica para mejorar la vida de los agricultores, será útil para nuestra nación. En este estudio, intentamos la herramienta moderna de ciencia y tecnología de aprendizaje automático y desarrollamos un modelo como Algoritmo de Aprendizaje Supervisado para la Predicción del Rendimiento de Cultivos (SLACYP) para predecir la tasa de rendimiento de los cultivos. Las variables consideradas en este modelo son área bajo cultivo (por Hectáreas), Precipitación (mm), Consumo de fertilizantes (MT), Nitrógeno (N), Fósforo (P), Potasio (K). El modelo se diseñó y desarrolló el algoritmo utilizando la Programación R para predecir los cultivos popularmente cultivados en la región seleccionada de Tamil Nadu, India, para cultivos como arroz, maíz, cúrcuma, cacahuete y caña de azúcar. El valor de R2 para todos los cultivos es mayor que 80%, lo que indica el mejor modelo para los cultivos. Entre los cultivos, la ecuación de regresión múltiple del arroz con cáscara es altamente significativa con el valor R2 de 0.95 y el valor más bajo se encontró con la caña de azúcar con el valor R2 de 0.81.

Palabras clave: Información geográfica, Aprendizaje automático, Rendimiento de los cultivos, Predicción, Regresión, RainfallandFertilizers

Abstrato

A agricultura indiana está enfrentando uma enorme e desafiadora tarefa de produzir 480 Mt até o ano de 2050 para atender às necessidades de alimentos da população. No entanto, os problemas são muito parecidos com as tensões bióticas e abióticas experimentadas pelas plantações, degradação do solo, chuvas irregulares, cenários de mudança climática, surgimento de novas pragas e doenças etc. A previsão de rendimento é uma das questões mais críticas enfrentadas no setor agrícola. o governo para projetar estratégias de planejamento adequadas para melhorar a produtividade e criar um bom preço de mercado para os agricultores pobres. Se a aplicação de soft computing em informações geográficas para melhorar a vida dos agricultores, será útil para a nossa nação. Neste estudo, tentamos a moderna ferramenta de ciência e tecnologia de aprendizado de máquina e desenvolvemos um modelo como o Algoritmo de Aprendizado Supervisionado para Previsão de Produção de Cultivos (SLACYP) para prever a taxa de produtividade das culturas. As variáveis consideradas neste modelo são a área cultivada (por Hectares), Precipitação (mm), Consumo de fertilizantes (MT), Nitrogênio (N), Fósforo (P), Potássio (K). O modelo foi projetado e desenvolvido o algoritmo usando programação de R para prever as culturas cultivadas popularmente na região selecionada de Tamil Nadu, na Índia, para culturas como arroz, milho, cúrcuma, amendoim e cana de açúcar. O valor de R2 para todas as culturas é

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

superior a 80%, indicando o modelo de melhor ajuste para as culturas. Entre as culturas, a equação de regressão múltipla de arroz é considerada altamente significativa com o valor de R2 de 0,95 e o valor mais baixo foi encontrado com a cana-de-açúcar com o valor de R2 de 0,81.

Palavras-chave: Informação Geográfica, Aprendizagem de máquina, Rendimento de culturas, Previsão, Regressão, Chuvas e Fertilizantes

Introduction

Agricultural systems science has been evolving over a century or more with contributions from a wide range of inter disciplinary subjects. In recent decades, huge number of inter disciplinary scientists, engineers, statisticians, and economists with training in systems modeling, analytical approaches, and information technology (IT) tools are using Agricultural Systems science mainly the demand for this system based approach is increasing.

India as an agricultural country and more than 80% of its population is dependent on agricultural income for their livelihood. It also contributes a significant figure to the Gross Domestic Product (GDP). Currently the farmers are shifting from the tradition of agriculture and diverting from it, since they are facing heavy losses, due to erratic and untimely / poor rain fall and fluctuating market price. Indian agriculture is facing a huge, challenging task of producing 480 Mt by the year 2050 to meet out the food grain requirement of the population. However the problems are plenty like biotic and abiotic stresses experienced by crops, soil degradation, erratic rainfall, climate change scenarios, emergence of new pest and diseases etc (Surendran et al., 2016).

At this juncture, the introduction and adoption of modern technology is inevitable to meet the challenges. One such technology is adoption of Decision support Systems or tools (DSS/ DST)

Introducción

La ciencia de sistemas agrícolas ha estado evolucionando durante más de un siglo con contribuciones de una amplia gama de temas interdisciplinarios. En las últimas décadas, un gran número de científicos interdisciplinarios, ingenieros, estadísticos y economistas con capacitación en modelado de sistemas, enfoques analíticos y herramientas de tecnología de la información (TI) están utilizando la ciencia de Sistemas Agrícolas, principalmente la demanda de este enfoque basado en sistemas está aumentando.

La India como país agrícola y más del 80% de su población depende de los ingresos agrícolas para su sustento. También contribuye con una cifra significativa al Producto Interno Bruto (PIB). En la actualidad, los agricultores están abandonando la tradición de la agricultura y desviándose de ella, ya que se enfrentan a grandes pérdidas, debido a una caída de la lluvia errática e intempestiva / deficiente y al precio fluctuante del mercado. La agricultura india se enfrenta a una enorme y desafiante tarea de producir 480 Mt para el año 2050 para cumplir con los requisitos de grano alimentario de la población. Sin embargo, los problemas son similares a los estreses bióticos y abióticos experimentados por los cultivos, la degradación del suelo, las precipitaciones erráticas, los escenarios de cambio climático, la aparición de nuevas plagas y enfermedades, etc. (Surendran et al., 2016).

En esta coyuntura, la introducción y

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

for improving the crop productivity by overcoming the stresses. DSS/ DST are intended to help different stakeholders to make more efficient / meaningful decisions through comprehensible decision stages and indicating the likelihood of various outcomes resulting from different strategies (Dicks et al., 2014). These can be dynamic software tools, whose recommendations vary according to the user's inputs, and they may suggest an optimal decision path. For farmers, and their advisers, software tools can facilitate effective farm management by recording data efficiently, analyzing it, and generating a series of evidence based recommendations (Rossi et al., 2014). Data Science is the practice of identifying the problem, working with existing datasets to provide solution that in turn business value (Drew Conway, 2010). Another new approach is Machine learning (ML), which may be also used in agriculture field for taking effective decisions, planning purposes and also for improving the crop productivity. Machine Learning is part of data science. Machine learning is about designing algorithms and developing automated systems that constantly train the datasets (Vincent Granville, 2014). ML is an interdisciplinary approach to data analysis that portrays the inspiration, from statistics, theory of probability and decision making, visualization, and optimization. Successful application of ML in various domains ranging from speech processing (Google voice) computer vision (face recognition), and natural language processing (IBM Watson), consumer predictive analytics to bioinformatics (drug design) etc are evident. ML approaches are classically helpful in situations where huge amount of datasets are available, in the case of agriculture it may be relating inputs from geographical information (fertilizers, pesticides, weather parameters) to output quantities of interest (crop yield).

para mejorar la productividad del cultivo al superar las tensiones. DSS / DST están destinados a ayudar a diferentes partes interesadas a tomar decisiones más eficaces / significativas a través de etapas de decisión comprensibles e indicando la probabilidad de varios resultados como resultado de diferentes estrategias (Dicks et al., 2014). Estas pueden ser herramientas dinámicas de software, cuyas recomendaciones varían según las entradas del usuario, y pueden sugerir una ruta de decisión óptima. Para los agricultores y sus asesores, las herramientas de software pueden facilitar la gestión efectiva de las granjas registrando los datos de manera eficiente, analizándolos y generando una serie de recomendaciones basadas en la evidencia (Rossi et al., 2014). La ciencia de datos es la práctica de identificar el problema, trabajando con conjuntos de datos existentes para proporcionar una solución que a su vez sea de valor para el negocio (Drew Conway, 2010). Otro enfoque nuevo es el aprendizaje automático (Machine Learning, ML), que también se puede usar en el campo de la agricultura para tomar decisiones efectivas, planificar y también para mejorar la productividad de los cultivos. Machine Learning es parte de la ciencia de datos. El aprendizaje automático se trata de diseñar algoritmos y desarrollar sistemas automatizados que entrenan constantemente los conjuntos de datos (Vincent Granville, 2014). ML es un enfoque interdisciplinario para el análisis de datos que retrata la inspiración, a partir de estadísticas, teoría de la probabilidad y toma de decisiones, visualización y optimización. La aplicación exitosa de ML en varios dominios que van desde el procesamiento de voz (voz de Google) visión por computadora (reconocimiento facial) y procesamiento de lenguaje natural (IBM Watson),

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

One of the major advantages of using ML approaches for agricultural and other inter disciplinary scientists are the opportunity to search large datasets to discover patterns and govern discovery by simultaneously looking at a combination of factors instead of analyzing each feature individually. Recently ML has become more user friendly because of the robust, and flexible software tools (R packages and Matlab toolboxes).

Yield prediction is one of the most critical issues faced in the agricultural sector (Sellam and Poovammal, 2016), which will help the government to device suitable planning strategies to improve the productivity and to create a good market price. Currently there are several methods are available for crop yield forecasting. The traditional method of yield forecasting is the evaluation of crop status by experts, subject matter specialists and agriculture department officials. Observations and measurements are made throughout the crop growing season and also crop cutting experiments will be carried out to predict the productivity. Apart from that crop yield is also predicted by using remote sensing and crop simulation models. The objective of the yield forecast is to give a precise, scientific sound and independent forecasts of crops' yield as early as possible during the crops' growing season by considering the effect of the weather and climate. Enhancing the accuracy and reliability of forecasting such crop yield variables has always been an important research topic for researchers. To date, there has been no single universal approach that provides the most appropriate yield prediction model, which forecasting results under all circumstances. This may be due to the fact that this natural processes of plant growth evolved through time, while modeling approaches (which are based on finite-length datasets) are synthetic by construction and are controlled by parametric forms that differ from one modeling appro

analítica predictiva del consumidor hasta bioinformática (diseño de fármacos), etc. Los enfoques ML son clásicamente útiles en situaciones donde hay gran cantidad de conjuntos de datos disponibles; en el caso de la agricultura, puede estar relacionando los aportes de la información geográfica (fertilizantes, pesticidas, parámetros climáticos) con las cantidades de interés de salida (rendimiento del cultivo).

Una de las principales ventajas de utilizar enfoques ML para científicos agrícolas y otros científicos interdisciplinarios es la oportunidad de buscar grandes conjuntos de datos para descubrir patrones y controlar el descubrimiento al observar simultáneamente una combinación de factores en lugar de analizar cada característica individualmente. Recientemente, ML se ha vuelto más fácil de usar debido a las herramientas de software robustas y flexibles (paquetes R y cajas de herramientas Matlab).

La predicción del rendimiento es uno de los problemas más críticos que enfrenta el sector agrícola (Sellam y Poovammal, 2016), que ayudará al gobierno a diseñar estrategias de planificación adecuadas para mejorar la productividad y crear un buen precio de mercado. Actualmente hay varios métodos disponibles para la predicción del rendimiento de los cultivos. El método tradicional de pronóstico de rendimiento es la evaluación del estado del cultivo por expertos, especialistas en la materia y funcionarios del departamento de agricultura. Las observaciones y mediciones se realizan a lo largo de la temporada de cultivo y también se llevarán a cabo experimentos de corte para predecir la productividad. Además de ese rendimiento de cultivo también se predice mediante el uso de sensores remotos y modelos de simulación de

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

ach to the next. However, despite their apparent value the uptake of DSS/ DST by stake holders (farmers, Agriculture Department, Policy makers and planners in India, has been limited.

There has been relatively little investigation into decision support tool use, but studies elsewhere (e.g. Australia, Belgium, Italy) have developed a number of important characteristics that determine use (Rossi et al., 2014; Kerselaers et al., 2015). ML Research and data science has also been undertaken in different disciplines, especially medicine / health, industries etc to a larger extent, however the takers in agricultural science is only a few (Shibl et al., 2013; Venkatesh et al., 2012). Yet despite sustained interest from inter-disciplinary researchers, uptake from agricultural specialists is still low, may be this is because of lack of coordination between agricultural scientists and other inter disciplinary subject experts. By considering these points, the present study was carried out to predict the yield of selected crops grown in Tamil Nadu, India using these techniques.

Materials and Methods

The figure 1 shows the machine learning supervision process (Taiwo Oladipupo Ayodele, 2010). Supervised machine learning is used for analyzing the past data and predicting new data (Karandeep Kaur, 2016; Kumar et al, 2016). We collected Area under cultivation (per Hectares), Rainfall (mm), Consumption of fertilizers (MT), Nitrogen(N), Phosphorous(P), Potassium(K) for a specific region of Erode District in Tamil Nadu State from Govt of Tamil Nadu website (www.tn.gov.in), Directorate of Economics and statistics (eands.dacnet.nic.in) Tamil Nadu Agricultural University Agritech Portal (agritech.tnau.ac.in). We have considered for Paddy,

cultivos. El objetivo del pronóstico de rendimiento es proporcionar un pronóstico preciso, científico y pronósticos independientes del rendimiento de los cultivos lo más temprano posible durante la temporada de crecimiento de los cultivos, teniendo en cuenta el efecto del clima y el clima. Mejorar la precisión y fiabilidad de las predicciones de tales variables de rendimiento de los cultivos siempre ha sido un tema de investigación importante para los investigadores. Hasta la fecha, no ha habido un único enfoque universal que proporcione el modelo de predicción de rendimiento más apropiado, que pronostica los resultados en todas las circunstancias. Ha habido relativamente poca investigación sobre el uso de herramientas de apoyo a la decisión, pero estudios en otros lugares (por ejemplo, Australia, Bélgica, Italia) han desarrollado una serie de características importantes que determinan el uso (Rossi et al., 2014; Kerselaers et al., 2015). ML Investigación y ciencia de datos también se ha llevado a cabo en diferentes disciplinas, especialmente medicina / salud, industrias, etc. en mayor medida, sin embargo, los tomadores en ciencias agrícolas son solo unos pocos (Shibl et al., 2013; Venkatesh et al., 2012) . Sin embargo, a pesar del interés sostenido de los investigadores interdisciplinarios, la aceptación de los especialistas en agricultura sigue siendo baja, puede deberse a la falta de coordinación entre los científicos agrícolas y otros expertos interdisciplinarios. Al considerar estos puntos, el presente estudio se llevó a cabo para predecir el rendimiento de los cultivos seleccionados cultivados en Tamil Nadu, India, utilizando estas técnicas.

Maize, Turmeric, Groundnut, and Sugarcane crops for predicting yield based on the five parameters (Snehal el al, 2014; Suraparaju V, 2014).

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

Design Methodology

The collected data in the real world can be in numerous types and formats (Taiwo Oladipupo Ayodele, 2010). It could be structured or unstructured, readable and small or big. We have designed the Supervised Learning Algorithm for Crop Yield Prediction (SLACYP) to predict crop yield rate.

This was done by analyzing the relationship between the factors considered and yield of the crop in our case rainfall, quantity of N, P and K fertilizers applied using regression equations. Regression analysis helps one to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Multiple regression is the best for this model as it considers more than one independent variable and regression analysis estimates the quantity of the yield. The regression model is fitted to the dataset. The model is fitted by using the lm () or "linear model" function. According to the relation observed the type of regression has been chosen. Later, the model designed were subjected to test by the training set.

The predict() function is used to test the model with the testing dataset. The accuracy of the model is evaluated with the R^2 value. R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determinations for multiple regression. R-squared is always between 0 and 100%. 0% indicates that the model explains none of the variability of the response data around its mean.100% indicates that the model explains all the variability of the response data around its mean. In general, the higher the R-squared, the better the model fits the data.

For easy understanding the stepwise approach has been explained below

The steps are as follows:

Step1: Convert Dataset into CSV (Comma separated Variables) format.

Step 2: Read the dataset and store in a variable.

Step3: Recognize the relation between the factors and the crop yield.

Step 4: Based on relation, decide regression type.

Step 5: Do regression analysis to estimate the quantity of the yield.

Step 6: Fit the regression model with dataset using the lm () function.

The general mathematical equation for multiple regression is

$$y = a + b1x1 + b2x2 + ...bnxn \quad ----- \quad (1)$$

Where y is the response variable.

a, b1, b2...bn are the coefficients.

x1, x2, ...xn are the predictor variables

The entire analysis was done with the help of R studio R/ Programming (Karthik Ramasubramanian and Abhishek Singh, 2016).

Results and Discussion

The model designed for paddy is given in Fig 2.

1) In the figure 2, taking only the last row, we can infer that the Nitrogen(N), Phosphorous(P) and Potassium(K) have a very low relationship whereas the Rainfall value seems to be in constant and Area produce a sharp decline in values.

2) Based on that we took, N, P and K as N*P*K for having both addicting and interactive effects on the yield factors.

3) As both, Rainfall and Area factors are seems to be positive, we have taken as Rainfall+Area.

Based on this, the regression relation for paddy is given in eqn. 2

$$Yield \sim Rainfall + Area / (N * P * K) \quad ---- \quad (2)$$

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

The model design for maize is given in Fig 3.

1) From the figure 3, we can infer that the all the factors have strong relationship.
2) Therefore, the factors are taken as Rainfall + Area + N + P+ K as all the factors contribute independently to the yield.
The regression relation for maize is as follows

Yield ~ Rainfall + Area + N + P + K ------- (3)

The model design for turmeric is given in Fig 4

1) From the figure 4, we conclude that Rainfall, Phosphorous(P) and Potassium(K) and Area graph are positive and increasing. The N value is decreasing and going in negative.
2) The N value is taken in a quadratic term as I(N^2) as the graph produce a curve structure.
3) The I(x) is known as a 'as is' function which produce a vector of squares for fitting quadratic terms.
4) The Area value is also taken as a quadratic term as I(Area^2).
The regression relation is as follows:
Yield ~ Rainfall + Area + N +P + K+ I(Area^2) + I(N^2) ------- (4)

The model design for sugarcane is given in Fig. 5
1) From the figure, We are taking the addictive effect of all the factors and quadratic terms for the factor Rainfall, Area and K.
2) From the figure 5 taking the last row, the factors Rainfall, Area and K have a slow rise. The P and N are declining.
3) Therefore the quadratic terms are taken for Rainfall, Area and K.
The regression relation is as follows:

Yield ~ Rainfall + Area + N +P + K+ I(Rainfall^2) + I(Area^2) + I(K^2)

The model design for groundnut is given in Fig 6.

1) From the figure 6 taking the last row, the Rainfall is looking positive. The factors Area, P and K produce a V-shaped structure.
2) As Area, P and K value produce a curved shape, the interactive function of Area : P + Area : K is taken.
3) As both Rainfall and N are quite similar in values, the interactive term of both the factors are taken.
4) The Rainfall factor is taken in quadratic form as it has a positive correlation with the yield.
The regression relation is as follows:

Yield ~ Rainfall + Area + N +P + K+I(Rainfall^2)+ Rainfall: N +Area : P + Area : K
Based on the developed model of, Supervised Learning Algorithm for Crop Yield Prediction (SLACYP) crop yield rate was predicted and presented in Table 1 and fig 7 to 11. A common way of accessing the designed model is by the value of R-squared or the coefficient of determination. A designed model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased. R-squared value is a statistical measure of how close the data are to the corresponding regression line. The value of R-squared is the percentage of the response variable variation that is given by a linear model. R-squared lies between 0 and 100 percentage. If R – squared is 0%, then the model explains none of the variability of the response data around its mean. If R – squared is 100%, then the model explains all the variability of the response data around its mean. In general, if the R-squared value is close to 100%, the model fits best with the data (Jeong et al, 2016).

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

For all the crops the R2 value was above 0.80 which indicated the model is fitting well to the observed values. (Fig.12. and Table 2) Among the crops, the multiple regression equation of paddy is found to be highly significant with the R2 value of 0.95 and the lowest value was found to be with the sugarcane with the R2 value of 0.81.

Although the developed model is not a crop physiology simulation model such as CERES -DSSAT, APSIM, EPIC etc the proposed Supervised Learning Algorithm for Crop Yield Prediction (SLACYP) using R Programming may be used for simulation or prediction of crop yield. This model is convenient for data acquisition, compilation, regionally applicable, parametrically simple, and effective for multi-scale factor integration. Besides, later on we can upgrade the same in future by integrating various other factors such as pest and diseases, socio economic factors, for improving the prediction accuracy.

Conclusion

The present study was carried out to predict the yield of selected crops grown in Tamil Nadu, India using Supervised Learning Algorithm for Crop Yield Prediction (SLACYP) in R language. With this model, the crop yield for the major crops i.e. paddy, maize, turmeric, sugarcane and groundnut which are grown in Erode district is predicted. The multiple regression equations showed that there is slight difference in values of rainfall, area under cultivation and the consumption of fertilizers, and these factors produced a significant effect on the yield of the crop. The R2 value for all the crops are greater than 80% which gives a very good prediction model. This study paved the way for further extension by adding many other crops like pulses, non- food crops and other food crops. Similarly it can be tried in any of the semi arid climatic condition area, i.e all the other districts of Tamil Nadu can be considered. As discussed, the model may be further elaborated by including the other factors such as temperature, solar radiation, soil moisture and depth, pest and disease infestation, minimum support price of the crop, and disasters. This will help to improve the prediction and also by considering the other factors we can determine whether the yield predicted is profitable or not. Besides, this model can be extended to create a smart agriculture for suggesting recommendations on nutrient, irrigation, which can be useful to farmers for improving the crop productivity.

References

Dicks, L.V.,Walsh, J., Sutherland,W.J., (2014). Organising evidence for environmental management decisions: a '4S' hierarchy. Trends in Ecology and Evolution, 29, 607–613.

for Statistical Computing, Vienna, Austria URL http://www.R-project.org/.

Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, et al. (2016) Random Forests for Global and Regional Crop Yield Predictions. PLoS ONE 11(6): e0156571. https://doi.org/10.1371/journal.pone.0156571

Karandeep Kaur. (2016), Machine Learning: Applications in Indian Agriculture." International Journal of Advanced Research in Computer and Communication Engineering, 2016.

Karthik Ramasubramanian and Abhishek Singh, (2016), Machine Learning Using R A Comprehensive Guide to Machine Learning Published by A press, 2016.

Kerselaers, E., Rogge, E., Lauwers, L., Van Huylenbroeck, G., (2015). Decision support for

Kumar, Rakesh, M.P. Singh, Prabhat Kumar, and J.p. Singh, (2015), Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique. In. International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

(ICSTM) (2015).
Michael J Crawley. (2012), "The R book". Published by Wiley &Sons, 2012.
prioritising of land to be preserved for agriculture: can participatory tool development help? Computers and Electronics in Agriculture. 110, 208-220
problem in agricultural decision support systems: the example of vite.net®. Computers and Electronics in Agriculture. 100, 88–99.
R Core Team, (2013). R: A Language and Environment for Statistical Computing. R Foundation
Rossi, V., Salinari, F., Poni, S., Caffi, T., Bettati, T., (2014). Addressing the implementation Sellam and E. Poovammal (2016), Prediction of Crop Yield using Regression Analysis, Indian Journal of Science and Technology, 2016.
Shibl, R., Lawley, M., Debuse, J., (2013). Factors influencing decision support system acceptance. Decision Support Systems 54, 953–961.
Snehal, S.Dahikar, Dr.Sandeep V.Rode, (2014) "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", International Journal of Innovative Research in Electrical,Electronic, Instrumentation and Control Engineering, 2014.
Suraparaju V, Misra B, Singh CD, "Machine learning approach for forecasting crop yield based on climatic parameters", International Conference on Computer Communication and Informatics (ICCCI–2014); Coimbatore, India, 2014.
Surendran, U., M.Jayakumar and S.Marimuthu. (2016). Low cost drip irrigation: Impact on sugarcane yield, water and energy saving in semiarid tropical agro ecosystem in India. Science of the Total Environment, 570: 1430-1440
Taiwo Oladipupo Ayodele (2010).
"Types of Machine Learning Algorithms, New Advances in Machine Learning " Yagang Zhang (Ed.), InTech, DOI: 10.5772/9385.
technology: extending the unified theory of acceptance and use of technology. MIS Q. 36, 157–178.
Venkatesh, V., Thong, J.Y.L., Xu, X., (2012). Consumer acceptance and use of information
Vincent Granville, (2014), 16 analytic disciplines compared to data science. July 2014 in online.
http://agritech.tnau.ac.in
http://eands.dacnet.nic.in/
http://mathworld.wolfram.com/LeastSquaresFitting.html
https://davetang.org/muse/2013/05/09/on-curve-fitting/
https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/
www.coastal.edu/kingw/statistics/R-tutorials/multregr.html.
www.r-bloggers.com
www.tn.gov.in

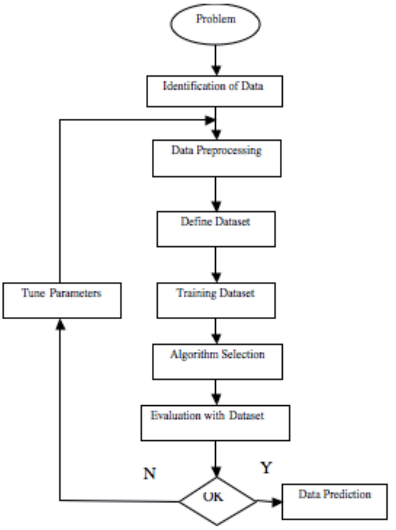

Figure 1 Machine Learning – Supervision Process

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818

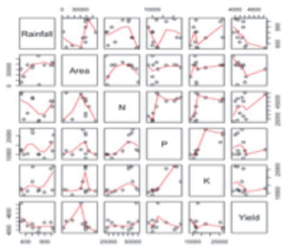

Figure 2 Plot Diagram for Paddy Crop

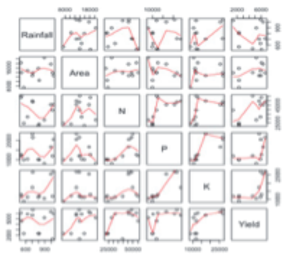

Figure 3 Plot Diagram for Maize Crop

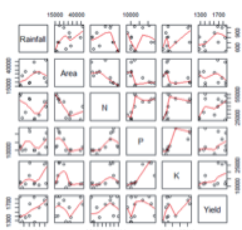

Figure 6 Plot Diagram for Groundnut Crop

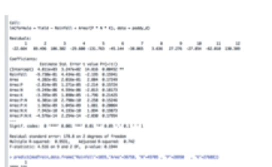

Figure 7 Crop Yield Prediction for Paddy

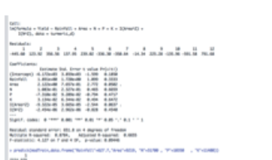

Figure 8 Crop Yield Prediction for Maize

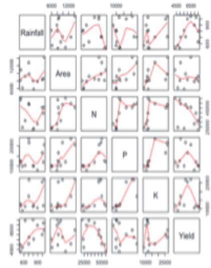

Figure 4 Plot Diagram for Turmeric Crop

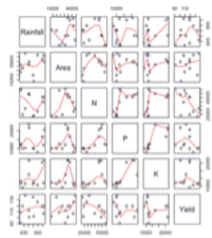

Figure 5 Plot Diagram for Sugarcane Crop

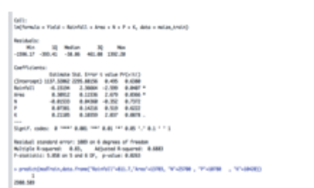

Figure 9 Crop Yield Prediction for Turmeric

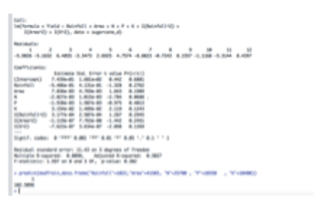

Figure 10 Crop Yield Prediction for Sugarcane

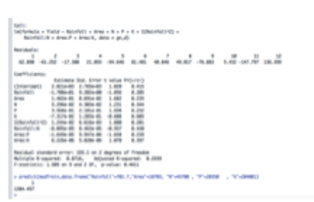

Figure 11 Crop Yield Prediction for Turmeric

Esta publicacion cientifica en formato digital es continuacion de la Revista Impresa: Deposito legal pp194802ZU42, ISSN 0378-7818