IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

Comparative Analysis of Machine Learning Techniques for Crop Yield Prediction

Sivanandhini P[1], Prakash J[2]

PG Scholar, Computer Science and Engineering, PSG College of Technology, Coimbatore, India[1]

Assistant Professor, Computer Science and Engineering, PSG College of Technology, Coimbatore, India[2]

Abstract: In agricultural field the crop yield prediction is significant and also a challenging task. Earlier, yield prediction was performed by considering farmer's experience on particular field and crop. This always requires involvement of farmer in prediction of crop yield which is not possible always. To overcome this challenge automated way to predict the yield of crop is proposed. In this work comparative analysis of crop yield prediction model using Machine learning techniques for the selected region i.e. district of Tamil Nadu in India. The machine learning algorithms like K-Nearest Neighbour, Decision Tree (Regression), Support Vector Regression were implemented and the performance of crop prediction model was analysed. The experimental analysis suggested that the performance for Support Vector Regression is better than K-Nearest Neighbour, Decision Tree, Support Vector Regression models.

Keywords: Machine Learning, K-Nearest Neighbour, Decision Tree, Support Vector Regression, Crop Yield Prediction.

I. INTRODUCTION

Agriculture is vital in any country growth and their economy. India is also highly dependent on agriculture and farmers. The farmers are which will be Crop yield prediction is very important issue in the field of agriculture. The prediction of crop yield may be very much useful for the farmers in the selection of crop for a particular place [1]. Farmers play a dynamic role in the cultivation of the crop. With the changing climate the farmers are facing difficulty in choosing the suitable crops for the particular place. The economy of the country will go down if there is no proper agriculture and have to rely on other countries for their foods.

The selection of suitable crop to cultivate depends on different factors like and geographical conditions of the particular place like rainfall, humidity, temperature, etc.,[2] It is the responsibility of the farmers to look into the suitable crops for their particular places with the help of the government. The selection of suitable crop by the farmers were earlier done by their previous experiences, with the unpredictability and changing climate it becomes a challenge for the farmers. The farmers have to determine the favourable and unfavourable conditions for a particular crop before they start cultivating them, using their prior knowledge or with the support and suggestions from the government [3][9].

With the development of the technology, machine learning techniques could be used in crop prediction in suggesting the suitable crop for cultivation. These machine learning models can use the attributes like Crop name, Crop season, Crop year, Area and production to predict the suitable crop for the particular place. The accurate information about history of crop yield predictions is an important thing for making decisions for all agricultural adapted risk management.

The crop prediction is need to support the farmers in selecting suitable crops which may help the yield of crop and also improve the economy of the country. It is also found that there are many researches are working on to develop an efficient crop prediction model [12][13][14], which may suggest crop prediction with higher accuracy.

The main objective of the paper is to predict the yield of crop using the machine learning models like: K-Nearest Neighbour, Decision Tree, Support Vector Regression and compare those models to suggest the efficient model based on the performance metrics.

The remainder of the paper is organized as section 2 discuss about the related work and the methodology deployed in crop yield prediction in section 3, followed by the performance evaluation in section 4 and conclusion in section 5.

II. RELATED WORK

This related work provides the better understanding of various Machine Learning techniques: Support vector Regression, Decision Tree Regression, k-Nearest Neighbour (Regression).
Nari Kim [4] discussed on statistical analysis based on linear relationships between crop yields and vegetation indices obtained from optical satellite sensors. It compares the Artificial intelligence models was made in order to develop the

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

best crop yield prediction model. The Artificial Intelligence can help to provide more accurate estimations of crop yields. In addition to vegetation indices, various land surface variables, such as weather elements, soil moisture, hydrological conditions, soil properties, and fertilizer application, have also been used in crop yield estimation. Author also suggested artificial intelligence could be an alternative to statistical modelling, because it can be efficiently handle problems of non-linearity and complexity. Since crop yield is a nonlinear process. AI can be regarded as a suitable prediction approach.

Adityashastry [5] discussed about the crop yield prediction used for the data mining techniques. In crop yield prediction of grain yield Multi Linear Regression (MLR) and Artificial Neural Network (ANN) were used. Regression analysis is used for prediction crop yield between the dependent entities. The prediction comparison made for best property subset used in validation purpose and gives better performances. Regression provided the better crop yield model. The errors found between the actual value and predicted value in crop yield model prediction made using Root Mean Squared Error.

S.Veenadhari [6] aimed to develop a web site for findings out the influences of climate parameters on crop production in selected districts. The methodology suggested the set some thresholds value for predicting the crop yield value. The crop yield field in maximum district are selected for crop yield prediction and missing attributes were handled and the model prediction was used on the climate parameters [10].

E.Manjula [7] discussed about the agricultural management and need of a simple and estimation techniques to predict the crop yield properly. The main focus is to identify whether Artificial Neural Network model could effectively predict rice yield for typical climate conditions. The crop model prediction model framework was developed and concluded the climate related variables were not the main determinants of crop yield, rather yield was greatly affected by the planning of crop model prediction [8].

III. METHODOLOGY

In the machine learning methods, the supervised learning techniques are which trains the given data based on input and output. The techniques have the capability to learn the past experience or historical data. The machine learning algorithms used in this paper for better crop model prediction namely: Support Vector Regression, K-Nearest Neighbour (Regression), Decision Tree Regression.

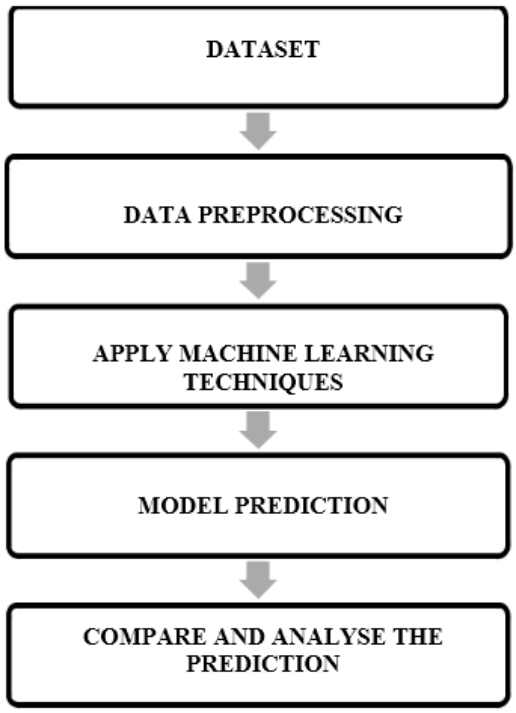

Fig 1: Flow Chart of Crop Prediction

A. Dataset Description: The dataset used for this work "Tamil Nadu Agriculture Dataset" is obtained for Kaggle repository. It consists of data of various districts of Tamilnadu from 2008 to 2012. The dataset consists of information of the crops like: variables in the dataset includes: State Name, District Name, Crop Year, Season, Crop, Area, Production.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

B. Data Preprocessing: Pre-processing refers to the transformation applied to our data before feeding into the algorithm. Some machine learning model needs a information in a specified format. The main pre-processing carried out are handling missing data, detecting noisy data, a removal of outliers and minimizing duplication.

C. Support Vector Regression: Support Vector Regression (SVR) is a regression algorithm, so we can use SVR for working with the continuous values instead of classification kernel. The SVR can be used to predict the continuous values such as integer, floating points. The Support Vector Regression uses principles as the SVM for classification, with only a few minor differences. Support Vector Regression model try to minimize the error between the predicted and the actual values and tries to fit the best line within a predefined or threshold error value. SVR acknowledges that the presence of non- linearity in the data and provides a proficient prediction model. In, SVR model error prediction difference between actual value (i.e., accurate value from the dataset) and prediction value (Output obtained by training the given algorithm based on the historical data) are determined to evaluate the model [11].

Pseudo code for Support Vector Regression

i. **Split** data into two parts (i.e, Training & Test data).
ii. **Apply** cross validation techniques (i.e., 5- fold cross validation)
iii. **Select** support vector machine (i.e., $\sum$ - SVR model)
iv. **Select** the type of core kernel (i.e., Gaussian Kernel)
v. **Optimize** model parameters using Pattern Search (PS) algorithm.
vi. **Validate** model and prediction of the SVR. D. **K-Nearest Neighbor (Regression)**

K-nearest neighbour is simple algorithms that stores all available cases and predict the numerical target based on a similarity measure. K-Nearest Neighbour (k-NN) algorithm is simple can be used solve both classification and regression problems. It is used for estimating continuous variables and unsupervised learning techniques. K-NN analysis consists of a set of machine learning methods allow us to predict a continuous outcome variable based on the value of one or more multiple predictor variable. K-nearest neighbour (Regression) model tries to minimize the error between actual value (i.e., accurate value from the dataset) and prediction value (Output obtained by training the given algorithm based on the historical data).

E. Decision Tree Regression: Decision tree regression or classification models form a tree structure. Decision tree can handle both categorical and numerical data. Decision trees can be used to visually and explicitly represent the decisions and decision making. Decision tree regression observes features of an object and trains a model in the structure of the tree to predict the data in the future to produce meaningful continuous output. Continuous output means that the output/results is not discrete, it is just known as numbers. The Decision tree regression model try to minimize the error between actual value (i.e., accurate value from the dataset) and prediction value (Output obtained by training the given algorithm based on the historical data).

Pseudo code for K-Nearest Neighbor

i. **Import** dataset
ii. **Handling** the missing data
iii. **Split** data into two parts (i.e, Training & Test data).
iv. **Prediction** model (i.e., Decision tree regression)
v. **Evaluate** the accuracy

IV. PERFORMANCE EVALUATION

The performance of the crop yield prediction model using K-Nearest Neighbour, Decision Tree (Regression), Support Vector Regression was evaluated using Root mean square error (RMSE) and Mean Squared Error (MSE) and the performance was compared.

A. Root Mean Square Error: Root mean square Error (RMSE) measures the error between actual and predicted. Lower values of RMSE indicate better fit and higher values indicate worst fit. RMSE is a suitable measure of how accurately the model predicts the response and it is most important criterion for fit if the main purpose of the model is prediction. RMSE is a non-negative value. It should be interpreted cautiously because positive and negative error. RMSE quadratic scoring rule that also measures the average magnitude of the error.

B. Mean Squared Error: Mean Squared Error (MSE) or mean squared deviation (MSD) is the measures of average of the squares of the errors. Average squared difference between the estimated values and the actual value.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

Table 1: Performance of Prediction Models

MODELS	RMSE	MSE
Support Vector Regression	0.5	0.3
K-Nearest Neighbour	2.22	4.90
Decision Tree Regression	2.50	6.29

The performance of the K-Nearest Neighbour, Decision Tree (Regression), Support Vector Regression model was determined and the obtained result was tabulated in Table 1. The results infer that the performance of Support Vector Regression is better than the other models in terms for both RMSE and MSE. The graphical comparison of the model in RMSE and MSE is shown in Fig 2 and Fig 3 respectively.

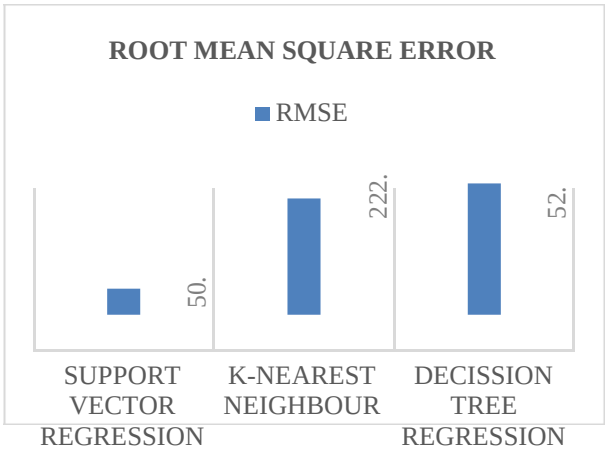

Fig 2: RMSE Comparison of the Predicted Model

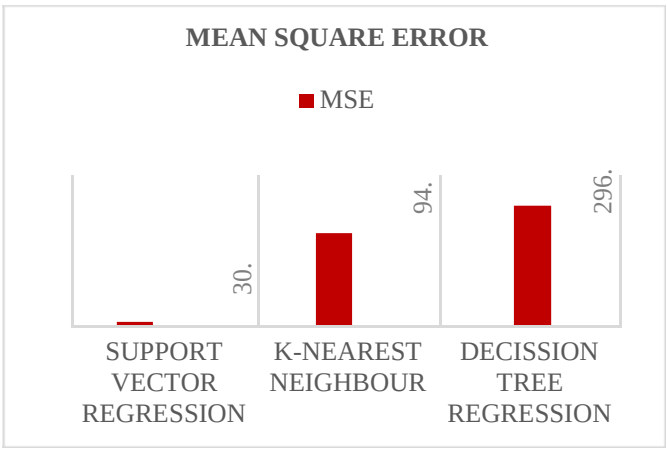

Fig 3: MSE Comparison of the Predicted Model

V. CONCLUSION

This paper discusses about the crop yield prediction model using Support Vector Regression, K – Nearest Neighbour and Decision Tree Regression. For the performance analysis of model, it is evaluated using Root Mean Square Error (RMSE), Mean Squared Error (MSE) between actual and predicted value of three models. The experimental analysis suggested that the Support Vector Regression provides better prediction comparatives to other two models used in this paper. The Future research focus can be done using Artificial Neural Network (ANN) for crop yield Prediction model by considering more features such as temperature, soil, humidity, etc., to provide more stable and accurate prediction models.

Copyright to IJARCCE

REFERENCES

[1] Priyanka, T., Soni, P., Malathy, C. (2018). Agricultural Crop Yield Prediction Using Artificial Intelligence and Satellite Imagery. Eurasian Journal of Analytical Chemistry, 13(SP), emEJAC181194.
[2] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-5, doi: 10.1109/I2CT45611.2019.9033611.
[3] Sivanandhini, P., & Prakash, J. (2020). Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network. International Journal of Innovative Science and Research Technology, 5(5), 1092-1096. https://doi.org/10.38124/volume5issue5
[4] Nari Kim , Kyung-Ja Ha , No-Wook Park , Jaeil Cho , Sung wook Hong and Yang-Won Lee "A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction" International Journal of Geo – informatics (2016) vol pp : 145 -150.
[5] Adityashastry ,H.A.Sanjay , E.Banushree " Prediction of crop yield using Regression"- International Journal of Soft Computing (2018), vol pp : 1908-1910
[6] S.Veenadhari , Dr.BharatMistra , Dr.C.D Singh "Machine Learning approach for forecasting crop yield based on climatic parameters" – International Conference on Computer Communication and Informatics (2014), vol pp: 239- 246.
[7] E.Manjula , S.Djodiltavhoumy "A Model for prediction of crop yield" – International Conference on Computer Communication and informatics (2014).vol pp: 1777-1785.
[8] C. O. Stockle., S. A. Martin and G. S. Campbell, "CropSyst, a cropping systems simulation model: water/nitrogen budgets and crop yield," Agricultural Systems, 1994, vol. 46(3), pp. 335-359.
[9] X. K. Chen and C.H. Yang, "Characteristic of agricultural complex giant system and national grain output prediction," System Engineering Theory and Practice, 2002, vol. 6(6), pp. 120-125.
[10] Tripathi S, Srinivas VV, Nanjundiah RS Downscaling of precipitation for climate change scenarios: a Support Vector Machine approach. J Hydrol ss330:621–640, 2006.
[11] Fagerlund S Bird species recognition using Support Vector Machines. EURASIP J Adv Signal Processing, Article ID 38637, p 8, 2007.
[12] Sandhya, R., Prakash, J., & Vinoth Kumar, B. (2020). Comparative Analysis of Clustering Techniques in Anomaly Detection Wind Turbine Data. Journal of Xi'an University of Architecture & Technology, 12(3), 5684-5694.
[13] Rajesh, R., &Mathivanan, B. (2017). Predicting Flight Delay using ANN with Multi-core Map Reduce Framework. In Communication and power engineering (p. 280). Walter de Gruyter GmbH & Co KG.
[14] Prakash, J. (2018). Enhanced Mass Vehicle Surveillance System. Journal for Research, 4(3), 5-9.