Mr. Mohan Kumar Chandol , Dr.E.Mohanraj, Dr. M. Umaselvi , Dr. K. Sankar

Research Article

Enhancement Of Agriculture Based Crop Yield Prediction Using R Tool And Machine Learning

Mr. Mohan Kumar Chandol[1] , Dr.E.Mohanraj[2], Dr. M. Umaselvi[3] , Dr. K. Sankar[4]

Abstract

The Crop yield-based prediction technique is a significant field of research that helps to ensure the food security and safe guard the food all over the nation. The projected system assists and support to predict enhanced and improved yield of crop and as well suggest the preeminent crop in dissimilar types of climatic conditions based on various parameter associated to the climates. This calculation will assist the farmer in the direction of choosing, whether the selected and particular crop is appropriate for that region and during the demanding season. In accumulation this website-based application gives information about details of productivity rate, irrigation, and other crop related details. Therefore, with the help of this website along with the different features help out the farmers to increase the crop production rate and crop quality improvement.

KEYWORDS: Prediction, dataset, R tool, SVM algorithm.

[1]Assistant Professor ,Department of Computer Science and Engineering,Koneru Lakshmaiah Education Foundation,Vaddeswaram, Guntur District,Andhra Pradesh – 522303.
[2]Associate Professor, Department of Computer Science and Engineering ,K.S.Rangasamy College of Technology, Tiruchengode, Tamil Nadu-637215
[3]Associate Professor, Department of Computer Science and Engineering, P. A. College of Engineering and Technology, Coimbatore 642002
[4]Assistant Professor, School of Engineering, Department of Computer Science and Engineering, Presidency University, Bengaluru - 560064.

I. INTRODUCTION

Prediction of crop yield helps in managing storage of crop as well as it directs the transportation decision, and risk management issue related to crop. This projected work assist the farmers to envisage the yield of crop and recommend the preeminent crop in various type of climatic condition related with the various parameter associated with the climates. In this work different dataset are collected from the publicly obtainable in the Indian Government based records for prediction of the crop yield. This kinds of parameters comprises in this dataset are minimum temperature level, precipitation, the average temperature level and the maximum level of temperature along with the   orientation crop evapotranspiration,

production, area, and the season yield for the period from Jan to December for 2001 to 2017 years. Engine education systems could be used efficiently to get better crop yield prediction under the

various climatic based scenarios. These kinds of prediction will assist the farmer in the direction of choosing whether the exacting kind of crop is appropriate for that region and in respective season. This kind of prediction is conceded out with the help of Bayesian based network algorithm in which high amount of accuracy is able to be achieved. The Bayesian based Network classification investigation technique is been used for discovering the different dataset. It is used to create a Net based group from a objects set which the group members are more alike. The Bayesian based Network classification symbolize the probability association between the crop and its rate of productivity. This will then constructs the network with the group of different Arecanut, the Rice based crop Coconut, Dry ginger based crop data and the Black pepper on various climatic based conditions and expects its probability value. This Bayesian Network oriented classification is the kind of supervised learning method that makes the temperature and the rainfall prediction of the crop based data that is used for the classification and the accuracy based values of the crops details that was made in precedent of 3 years.

II. BACKGROUND AND RELATED WORKS

2.1 DataSet Collection

In this module, the agricultural based dataset is submitted to the system for the crop prediction and classification process. From the dataset based on the vast initial, just a restricted number of the significant factors that have the maximum impact on the yield of agricultural were chosen for the experimentation purpose (UCI based Repository). The crop based production will be definitely having a crash due to difference in temperature levels. Hence the minimum, maximum and the average temperature levels for the each year in all districts were measured for projected based method. The suggestion of the crop evapotranspiration is intended on the origin of the monthly mean towards that particular year for all district of the TamilNadu. The present system uses the Microsoft Office Excel for visualizing the data.

2.2 Feature Extraction

In this module, from the given dataset, the attributes are selected and processed for the pprocess of SVM classification and clustering. The soil type in the each district is been identified by its hard kind copy of the state that maps from the dataset received. These types of soil are classified in to different eight major kinds of classes; with each class having the further kind of sub classes. The crop sophisticated regions (hectares) in the particular season for the all year in every selected particular district of the Tamilnadu state be considered in this module. Attaining each kind of parameter (minimum, precipitation, maximum, average, reference crop and the temperature level) by monthly mean of the records in each district.

2.3 SVM Classification

The Support Vector based Machines (SVMs) is a supervised based machine learning method. Numerous examples are there in which it is used in the field of agricultural. To diminish the generalization fault bound and in the direction of achieving the generalized performance, the SVM is been used effectively in this module. This SVM based (Support

Vector Machine) method assist in creating functions with a set of the labeled training data. This function is also called as classification function or a common regression based function. This module uses an algorithm is used in the direction of studying the efficiency of this methods with the help of the dataset that is used in the projected method. The experimental results are generated in R tool by the usage of the algorithm.

Algorithm:

• Step1: Study the crop based dataset.

• Step2: Make the data based frame and the extraction feature of the Crop production, mean rainfall, the year temperature and mean temperature based dataset.

• Step 3: Generate the Support Vector based Machine class using the e107 package and non-linear, linear and the kernel based sequences model.

• Step 4: The first phase is the data pre-processing, in this phase omit the inconsistent data and missing data. After the dataset will be partitioned as 70:30 ratios for achieve high accuracy.

• Step 4: The Second phase is the SVM applied on the trained dataset.

• Step 5: Trained dataset will be classified on the basis of nearest data point of the

Support vectors.

• Step 6: Predict the suitable crop for temperature and rainfall while using predict

function to predict the crop which is comparable of test dataset.

• Step 7: Build the confusion matrix to find the accuracy of the model.

• Step 8: Tuning of a Support Vector Machine will be improve the accuracy of the

model.

• Step 9: Predict the suitable crop for temperature and rainfall while using predict

function to predict the crop which is comparable of test dataset.

• Step 10: Build the confusion matrix of tuned model and find the accuracy of the

model.

2.4 K-Nearest Neighbors

It is the supervised machine learning algorithm. Nearest Neighbors classification is used to find the nearest data points in the datasets is the integer value that must be greater than 0. Algorithm is applied on the trained datasets. Using metrics package to get accuracy of results K-Nearest based Neighbors is the effective and essential classification based algorithms in the Machine Learning. This comes under the supervised based learning domain which finds the intense related application in the data mining, pattern recognition technique, and the intrusion detection. This is widely not reusable in the real-life scenarios because it is usually non-parametric, means it will not create any fundamental assumptions concerning the data allocation. Here it uses some prior 26 kinds of data (also said as training data), that classifies the coordinates into groups recognized by an attribute.

Algorithm:

• Step1: import the crop based dataset.

• Step2: Generate the data frame with its extraction feature of the Crop yield production, temperature, year, mean temperature rainfall and mean rainfall dataset usage.

• Step 3: Install the caret package whose contain K-Nearest Neighbor function.

• Step 4: Normalize the Tamilnadu crop dataset.

• Step 5: Crop dataset where partitioned under 70:30 ratios. 70 percentages of data

comes under the training dataset and remaining 30 percentages of data comes under testing dataset.

• Step 6: Create the model for the crop prediction where the factor variable is the crop name. Select the dependent and independent attributes for the yield prediction.

• Step 7: K-Nearest Neighbor technique applied on the trained dataset where the

dependent variables. Independent variables are year and season wise crop area,

production, temperature and rainfall.

• Step 8: Predict the suitable crop for temperature and rainfall while using predict

function to predict the crop which is comparable of test dataset.

• Step 9: Build the confusion matrix to find the accuracy of the model.

2.5 Random Forest

The random forest based algorithm is a kind of supervised based classification algorithm. With its name, it suggests that the algorithm generates the forest with a numerous number of the trees. Many more trees inside the forest, more vigorous the forest will look like.

Also the random based forest classifier, advanced number of forest trees gives the eminent accuracy results.

Algorithm:

• Step1: Examine the crop based dataset.

• Step2: Generate the data frame and the extraction feature of Crop yield production, temperature, year, mean temperature and mean rainfall related dataset is used.

• Step 3: Install the caret package whose contain Random Forest function.

• Step 4: Crop dataset where partitioned under 70:30 ratios. 70 percentages of data

comes under the training dataset and remaining 30 percentages of data comes under testing dataset.

• Step 5: Create the model for the crop prediction where the factor variable is the crop name. Select the dependent and independent attributes for the yield prediction.

• Step 6: Random Forest technique applied on the trained dataset where the dependent

variables. Independent variables are year and season wise crop area, production,

temperature and rainfall.

• Step 7: Predict the suitable crop for temperature and rainfall while using predict

function to predict the crop which is comparable of test dataset.

• Step 8: Build the confusion matrix to find the accuracy of the model.

• Step 9: Tuning of an Algorithm will be improve the accuracy of the model.

• Step 10: Predict the suitable crop for temperature and rainfall while using predict

function to predict the crop which is comparable of test dataset.

• Step 11: Build the confusion matrix of tuned model and find the accuracy of the

model.

2.6 Performance Evaluation

Every instance is categorized into two different classes in a binary based classification model. This kind of two different classes is false and true classes. It gives ascend to four different possible classifications strategy for each and every instance which is denoted as,

• True - Positive (TP): Number of accurate predictions so as to a positive instance.

• False - Positive (FP): Number of incorrect based predictions that an instance which is positive.

• False - Negative (FN): Quantity of incorrect based predictions with a negative instance.

• True - Negative (TN): Number of correct predictions that is an instance is negative.

III. PROPOSED SYSTEM

The proposed system also implemented with the existing approach. In addition to that Bayesian network is used explore the cluster analytical process of the provided agricultural dataset. Bayesian network analysis which is been used to pre-crop the massive agricultural dataset that follows a most expensive classification based analysis on the sub-set dataset. Bayesian network can be used to quickly play with the network and to explore whether they are ignored patterns or the relationships in the given agricultural dataset. In addition to that, plot each set of data item as a point in the n-dimensional based space (where the n is the number of description) that with the assessment of the each and other feature which is being the value of its coordinate exacting. Next, it performs the classification with finding its hyper-plane so as to discriminate the both classes very closely. With the help of the proposed model the crop based data set is been analyzed and the optimal based parameters of the crop based production are determined. This proposed method is been considered with some factors such as the soil with the some other factors related to the crop which boost the yield of crop production beneath the various climatic conditions. The Barota based package is been applied on the pre-processed based dataset. That package may give the confirmed type of attributes, such as removed attributes and tentative attributes. This system establishes the attributes and the tentative attributes which helps to predict the extra accuracy. The classification is been performed by knowing the hyper-plane that they distinguish the both classes extremely close. With the help of the datasets, the accuracy is at variance under the support based vector machine. The K-Nearest based Neighbor categories the dataset with the foundation of the K-Nearest data points. The data point symbolizes the production, rainfall, temperature and area of a crop. The random based Forest classifier mechanism on the foundation of the decision tree. They create numerous trees in a random order, consequently that this algorithm provides more accuracy.

3.2 Advantages of Proposed System

• This proposed system suitable for all crops.

• Comparisons of algorithms are used to find the best prediction model.

• Temperature, area, production and rainfall to enlarge the crop production under the different climatic conditions.

• In the proposed system implement with different attributes the productivity of the

crop is increased.
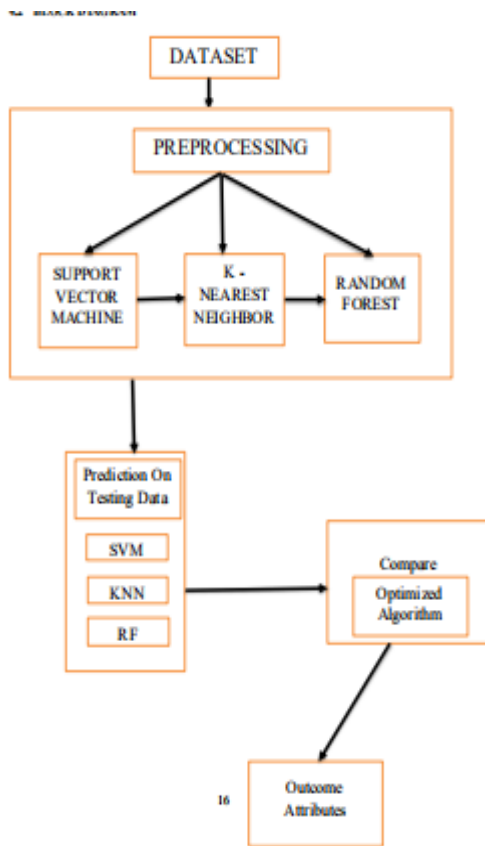
IV. EXPERIMENTAL RESULTS

This section confer the results attained after applying the supervised based Machine learning methods on crop yield production based dataset of the Tamilnadu state, India. R tool is used to build the algorithm.
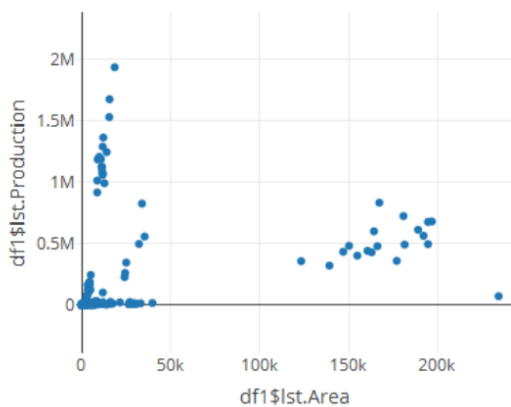
Fig 4.1.1 OVERALL AREA PRODUCTION RATIO

The total production of the crop is plotted form the trained dataset to show the

production rate which helps the farmers to get the clear view of the crop production level. Y axis shows the production and x-axis show the area of the crop. Support vector machine is applied to the dataset; it predicts the 63% of accuracy. In order to increase the accuracy of the model algorithm is tuned.
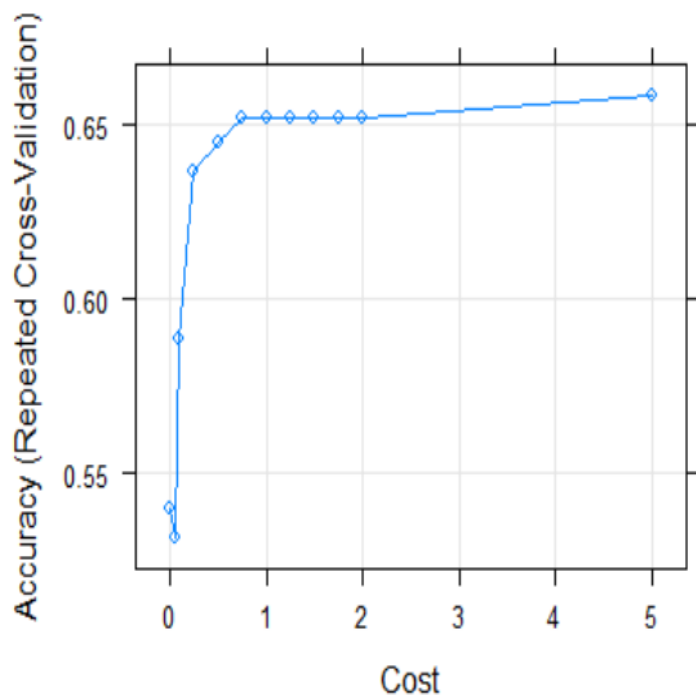

Fig 4.1.2 SVM TUNED ACCURACY

The algorithms are applied to pre-processed trained dataset it shows some accuracy but in order to increase the

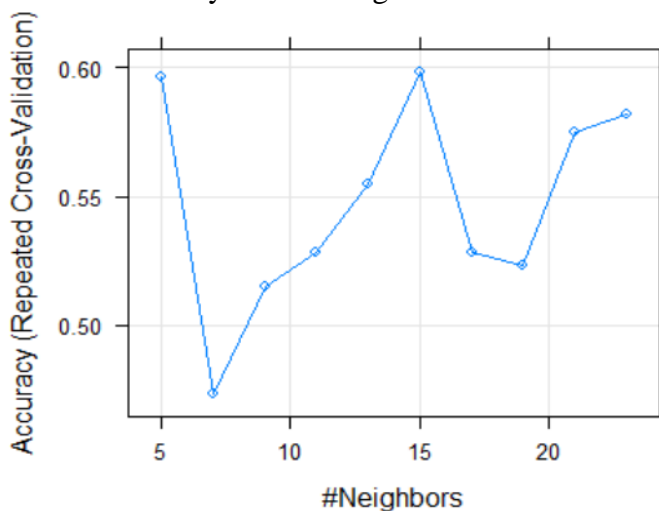accuracy level the algorithm are tuned .

Fig 4.1.3 KNN ACCURACY

The image shows the accuracy level of the tuned KNN algorithm from the Pre
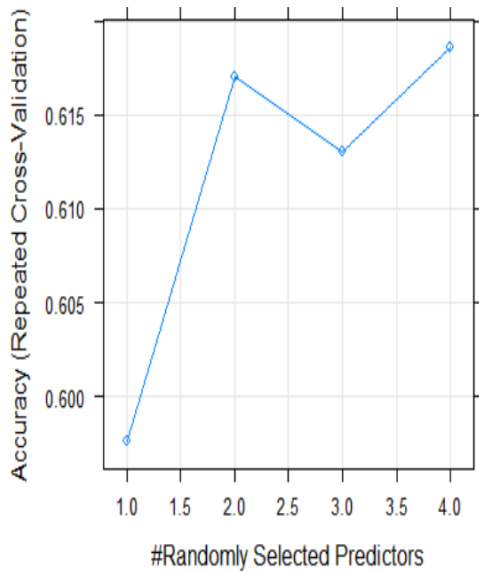
processed Dataset.

Fig 4.1.4 RF ACCURACY

The image shows the accuracy level of the tuned Random Forest algorithm from the Pre-processed Dataset.
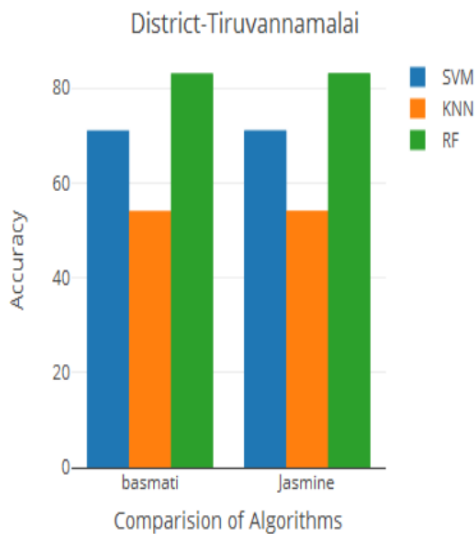

Fig 4.1.6 COMPARISION OF SVM, KNN & RF BASED ON

ACCURACY

The above diagram shows the comparison of all three algorithms. Each bar shows the accuracy of the algorithm. From the above figure Random forest achieve the maximum accuracy then other algorithms. Each shows the show corps production, rainfall and It's shows the crop accurate rainfall, production climatic conditions for the district where it give the maximum productivity of the crop Comparison of Support Vector Machine and K-Nearest Neighbors, K-NN gives more accuracy then SVM. The SVM algorithm achieved the accuracy of 63%, K-NN achieved 50% and RF achieved 94% accuracy.

V. CONCLUSION

To begin with the statistical way of model with the support vector model is been applied on the existing data. The experimental results so found be verified and examined using the Data Mining methods namely Bayesian based Network classification technique. This work uses the R-Tool based implementation with the support vector based machine (SVM) and the bayesian based Network classification oriented algorithm. This project is mean to predict the coconut, rice, Black pepper, Dry ginger and the Areca nut, and the crops yield that is very precise models for crop yield based estimation using information and communication technologies could help the farmers in increasing the yield.

VI. REFERENCES

[1] S.Veenadhari[1] , Dr. Bharat Misra[2] , Dr. CD Singh[3] "Data mining Techniques for Predicting Crop Productivity – A review article" 1,2-Mahatma Gandhi Gramodaya Vishwavidyalaya, Chitrakoot, Satna, India, 3-Central Institute of Agricultural Engineering, Bhopal, India, March-2011.

[2] Akbar Batcha, Syed Musthafa, Securing data in transit using data-in-transit defender architecture for cloudcommunication, Soft Computing, springer, june 2021

[3] Chi-Chung LAU, Kuo-Hsin HSIAO, 2005. "Bayesian Classification For Rice Paddy interpretation". Paper presented in Conference on data mining held at China Tapei. December, 2013.

[4] A Syed Musthafa , M Praveenkumar ," E-agricultural system based intelligent predictive analysis and smart farming with digitalized demand and supply utilization to maximize the yield rate of crops using machine learning algorithm, Turkish Journal of Computer and Mathematics Education, Vol.12 No.10 (2021), pp 2036-2041

[5] Cunningham S.J., G. Holmes. 2005. "Developing innovative applications in agriculture using data mining". Proc. of 3rd International Symposium on Intelligent Information Technology in Agriculture. Beijing, China. Oct, 2013.

[6] Dr. A. Syed Musthafa , C. Prashanth, "Kidney Disease Detection Using Machine Learning:, Turkish Journal of Physiotherapy and Rehabilitation; 32(2), ISSN 2651-4451, may 2021, pp 3359-3365

[7] D Ramesh1 , B Vishnu Vardhan2 "Data Mining Techniques and Applications to Agricultural Yield Data",September 2013.

[8] Syed Musthafa.A, "A Fuzzy based High-Resolution Multi-View Deep CNN for Breast Cancer Diagnosis through SVM Classifier on Visual Analysis", Journal of Intelligent & Fuzzy Systems, IOS Press, 10.3233/JIFS-189174, Page 1-14, September 2020

[9] Verheyen K, Adriaens D, Hermy M, Deckers.S,2014,"High-resolution continuous soil classification using morphological soil profile descriptions". Geoderma Vol.

[10] Syed Musthafa.A, Mohanraj.B, Priyanga.S, Krishnan.C, "High Security Distributed MANETs using Channel De-noiser and Multi-Mobile-Rate Synthesizer", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No. 2, ISSN 2278-3091, Page 1346-1351, April 2020

Mr. Mohan Kumar Chandol , Dr.E.Mohanraj, Dr. M. Umaselvi , Dr. K. Sankar

[11] Syed Musthafa A, Abdul Jaleel D, "Tourist Spot Proposal System Using Text Mining", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No. 2, ISSN 2278-3091, Page 1358-1364, April 2020