International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 11th Jan 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 01

DOI: 10.48047/IJIEMR/V12/ISSUE 01/68

Title **Comparative Study of Regression models for Prediction of Crop Yield**

Volume 12, ISSUE 01, Pages: 717-724

Paper Authors
Jayshree Hajgude, Dr. Tanuja Sarode

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Comparative Study of Regression models for Prediction of Crop Yield

Jayshree Hajgude*, Dr. Tanuja Sarode[1]
Research Scholar, Department of Computer Engineering, Thadomal Sahani Engineering College, Mumbai, India.
Email: Jayashree.hajgude@ves.ac.in
[1] Head of Computer Engineering Department, Thadomal Sahani Engineering College, Mumbai, India.

Abstract
Agriculture is the major industrial sector contributing to the economy of India. Predictive crop yield can guide farmers in the crop cultivation and scheduling process. Maharashtra is one of the leading States in the agriculture section. This study uses a regression model to predict Soybean crop yield and study the factors affecting the Soybean crop yield of major districts of the Marathwada region from Maharashtra state. Data about crop yield has been sourced from the Indian Government's records. The weather features like precipitation, minimum maximum and average temperature, term and heavy rainfall, rainy days, an area, and yield for the years 2001 to 2019 has been collected from India Meteorological Department. The dataset was pre-processed and feature engineering is being done on the data for determining features contributing to crop yield. Various types of regression Models have been implemented before and after feature engineering and performance evaluation has been done. Performance metrics like mean absolute error, root mean squared error, and mean squared error has been calculated for the present study and different regression model performance metrics have been analyzed.

Keywords—regression, preprocess, ridge, crop

Introduction

Agriculture majorly contributes to the Indian economy since Agriculture majorly contributes to the Indian economy since it is one of the most important fields practiced in India. It is one of the substantial economic sections and plays a huge role in the development of the country. More than 60 % of the land in India is used for agricultural purposes [1]. Agriculture has played an important role in improving the Indian economy and will continue to be the same in the future [2]. Introducing new techniques in agriculture is helping to increase productivity. Machine learning is one of the important fields that can be applied in this area for far further improvement. Crop growth and crop yield are sensitive to different parameters like weather parameters, Soil parameters, etc. Maharashtra has a major contribution to soybean production. Marathwada Region is one of the major regions which contributes to Soybean crop production. Marathwada has a variety of temporal and spatial features. So, it is important to study different temporal and spatial features of the Marathwada region which are responsible for crop growth and helps in crop yield prediction. The contents of the paper are structured as below. We have discussed existing methods of crop yield prediction in section II. In section III we discussed the implementation of the proposed system using various Regression models In Section IV we evaluated different regression models by considering different performance evaluation parameters and conclusions in Section V.

Literature Survey

Crop yield prediction is a principal section of agriculture that contributes majorly to the Indian economy. In this section, we have discussed existing methods used for crop yield prediction. S.Veenadhari[3] developed a Crop Advisor model for crop yield using climatic parameters and the C4.5 algorithm for selected districts of Madhya Pradesh. Sellam and E. Poovammal[4], applied a Regression model on environmental parameters and Food Price Index data. Niketa Gandhi[5][6][7], applied neural networks, SMO classifier, Bayes Net, and naive Bayes algorithms for predicting crop yield. Prof. D.S. Zingade [8], applied a Multiple Linear Regression on the integrated data obtained from the weather department, and Soil parameters to predict the most appropriate crops according to the current

environment. Igor Oliveira [9], applied a recurrent neural network on satellite-derived precipitation, soil properties datasets, and seasonal climate forecasting data. The system provides significantly useful results without using high-resolution remote-sensing data and provides farmers information about climate influence on the crop cycle. S. Dharmaraja [10],used regression and time-series techniques to predict crop yield using historical data and improved accuracy in crop yield prediction when auxiliary variables are chosen based on the plant growth phase information. Narayanan Balakrishnan [11], developed AdaSVM and AdaNaive ensemble models to predict crop production over time. In literature, different Machine Learning methods have been applied in the area of crop production improvement. Machine Learning provides several effective algorithms which are used to find the contributing features and their effect on yield. This research work analyses the various regression model for the prediction of crop yield of the Soybean crop for the Marathwada region.

Research Methods

This section includes the area used in this study, data set details, and methodology used for implementing crop yield prediction.

Study Area:

The area used in this study is the Marathwada region of Maharashtra state in India. The principal food products of Marathwada are Soybean, Wheat, Jawar, Bajra, and pulses [21]. For the present study, 8districts Aurangabad, Beed, Hingoli,Jalana, Latur, Osmanabad, Parbhani, and Nanded from the Marathwada region are considered.

Dataset Details:

The crop production Dataset used in this research has been collected from the publicly available records of the Indian Government for the year 2001 to 2019. Feature dataset about whether has been collected from Indian Meteorological Department Pune. The features considered for this study are rainfall, precipitation, minimum, average, and maximum temperature, evapotranspiration, area, and yield.

Methodology

The proposed work includes building a regression model by considering various weather-influencing factors for Crops and analyzing the performance of different regression models.

The working of proposed model includes the following modules as shown in figure 1:

a. Collection of data: This module collects data from different sources like crop yield dataand climatic data.

b. Data Fusion Model: This module performs pre-processing of data and then integrates it to build fusion data set for crop yield prediction.

c. Feature Engineering: Recursive feature elimination method is used for feature selection.

d. Prediction model: different types of the Regression model are used to predict crop yield.

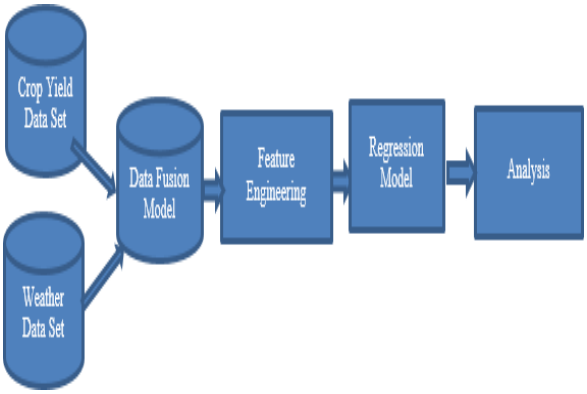

Figure 1. Regression Model for Crop YieldPrediction.

Proposed Framework:

In the proposed work, different types of Regression algorithms are implemented to predict crop yield. The weather parameters like rainfall, precipitation, minimum, average, and maximum temperature, area, etc.

Algorithmic details:

Following regression models have been implemented to study crop yield predictionand their performance.

The linear regression model is used to determine the linear relationship between dependent and independent variables. If data consist of more than one independent variable is called a multiple linear regression model. Ridge Regression is a variation of the regression model which is used when there is a high correlation between the independent variables. The Bayesian Ridge is a variation of Ridge regression where coefficients are determined using a posteriori estimation under the Gaussian distribution. Lasso Regression performs regularization whileselecting features and allows the selection of a set of features from the dataset to build the model. Principle Component Analysis generates principal components from the predictor variables, with several principal components matching the number of original features p. Support Vector Machine regression uses the hyperplane

which has the maximum number of points.

b) Architecture of Proposed Model: The architecture of the proposed work is shown in figure 2. Different Regression models have been applied for the predictionof crop yield.

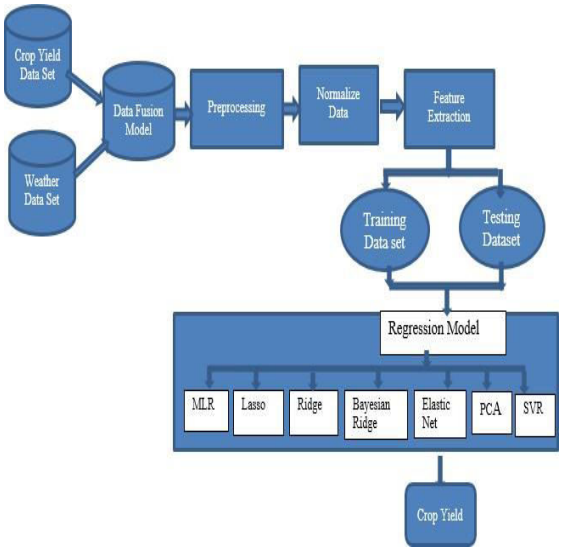

Figure 2. The architecture of the proposed Regression Model

a) Implementation Details:

The steps involved in the implementation ofthe proposed model are listed below:

Step 1: Extract data from the crop production data set and weather data set.

Step 2: Integrate both data sets by standardizing data.

Step 3: Perform Data pre-processing on integrated data like removing missing values, and outliers.

Step 4: Normalize all features using the Min-Max normalization technique

Step 5: Divide data in the training set and testing set. Step 6: Build models using various regression techniques.

Step 7: Analyze the performance of the different models.

Performance Evaluation

After integrating crop production and weather data set multiple regression model has been implemented. The performance ofthe different models has been evaluated by using performance metrics like mean absolute error, mean squared error, and root mean squared error.

Regression Model	Split Ratio	MAE	MSE	RMSE
Multiple Linear Regression	**70:30**	0.18	0.047	0.217
	80:20	0.13	0.027	0.167
Ridge Regression	**70:30**	0.13	0.029	0.170
	80:20	0.29	0.128	0.357
Lasso Regression	**70:30**	0.24	0.113	0.337
	80:20	0.34	0.225	0.474
Bayesian Ridge Regression	**70:30**	0.14	0.024	0.157
	80:20	0.14	0.031	0.176
Elastic Net Regression	**70:30**	0.24	0.114	0.337
	80:20	0.34	0.225	0.475
Principle Component Analysis Regression	**70:30**	0.15	0.065	0.256
	80:20	0.28	0.117	0.342
Support Vector Regression	**70:30**	0.25	0.084	0.290
	80:20	0.28	0.102	0.320

Table 1. Crop Yield Prediction Regression Model Result

Table 1 depicts the result of different regression models by considering weather parameters. The result has been calculated for the 80:20 and 70:30 split ratio of the training and testing dataset respectively. It

Ridge Regression algorithms performed well for Soybean crop yield prediction based on weather parameters as compared to the remaining algorithm.

Feature Engineering:

The recursive elimination method is used for feature extraction. Different regression algorithm has been implemented for n=3 and n=4 and the performance of the model has been analyzed where n is several features. By using the recursive elimination method different domination features have been extracted. The performance of the different models has been evaluated by using different performance metrics.

Regression Model	**Feature Set**	**MAE**	**MSE**	**RMSE**
Multiple	**AREA,HVYRF,RD**	0.06	0.006	0.080
	LMIN,AREA,HVYRF, RD	0.12	0.06	0.179
Ridge	**AREA,HVYRF,RD**	0.14	0.031	0.178
	LMIN,AREA,HVYRF, RD	0.13	0.034	0.185
Lasso	**AREA,HVYRF,RD**	0.26	0.120	0.347
	LMIN,AREA,HVYRF, RD	0.23	0.134	0.367
Bayesian Ridge	**AREA,HVYRF,RD**	0.06	0.006	0.079
	LMIN,AREA,HVYRF, RD	0.13	0.034	0.185
Elastic Net	**AREA,HVYRF,RD**	0.26	0.120	0.347
	LMIN,AREA,HVYRF, RD	0.23	0.134	0.367
PCA	**AREA,HVYRF,RD**	0.11	0.024	0.156
	LMIN,AREA,HVYRF, RD	0.12	0.032	0.179
SVM	**AREA,HVYRF,RD**	0.16	0.046	0.214
	LMIN,AREA,HVYRF, RD	0.12	0.032	0.179

Table 2. Crop Yield Prediction Regression with feature engineering Model Result

After performing feature engineering, it is found that the performacnce of the model with feature set AREA, HVYRF, RD is better than LMIN, AREA, HVYRF, RD. The performance of models has been improved after performing feature engineering.

Conclusion

The proposed system takes into consideration weather data collected from the meteorological department about the Marathwada Region and crop production data of the past years and predicts crop yield. Feature engineering has been done and dominating features have been chosen from the integrated data. Various regressiontechniques have been applied to the data. After applying different algorithms it is found that the Bayesian Ridge Regression model is performing better as compared to the remaining algorithm. After performing Feature engineering performance of the model has been improved.

References

[1] Ghodsi, R.; Yani, R.M.; Jalali, R.; Ruzbahman, M, "Predicting wheat production in Iran using an artificial neural networks approach", Int. J. Acad. Res. Bus. Soc. Sci. 2012, 2, 34– 47

[2] Gonzalez-Sanchez, A ., Frausto-Solis, J., Ojeda- Bustamante, W.," Attribute selection impact on linear and nonlinear regression models for crop yield prediction", Sci. World Journal, 2014.

[3] S. Veenadhari , Bharat Misra , CD Singh," Machine learning approach for forecasting crop yield based on climatic parameters ",International Conference on Computer Communication and Informatics, 2014, Coimbatore, India .

[4] V. Sellam and E. Poovammal ,"Prediction of Crop Yield using Regression Analysis",Indian Journal of Science and Technology, October 2016.

[5] Niketa Gandhi, Owaiz Petkar and Leisa J. Armstrong "Rice Crop Yield Prediction Using Artificial Neural Networks", IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2016, 105-110.

[6] Niketa Gandhi, Leisa J Armstrong, RiPtkar, Owaiz Petkar, Amiya Kumar, "Rice crop yield prediction in India using support vector machines", International Joint Conference on Computer Science and Software Engineering, 2016, IEEE xplore.

[7] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong,"Predicting Rice Crop Yield Using Bayesian Networks", 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India,Publisher:IEEE

[8] 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

[9] Gor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto ,"A Scalable Machine Learning System for PreSeason Agriculture Yield Forecast",2018 IEEE 14th International Conference on e-Science (e-science).

[10] S. Dharmaraja Vidyottama Jain Priyanka Anjoy Hukum Chandra , "Empirical Analysis for Crop Yield Forecasting in India", research

article,Publisher:Springer.

[11] 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