International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206616

Application of Machine learning in Crop Yield Prediction of Finger Millet using Multiple Linear Regression

Tarun Mourya. S*1, S. Nagini2, Anne Slagha3

*1,3Department of Computer science, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

2Department of Computer science, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

Corresponding author : tarun.nanoscience@gmail.com

Article Info
Volume 6, Issue 6
Page Number: 280-284
Publication Issue :
November-December-2020

Article History
Accepted : 29 Nov 2020
Published : 16 Dec 2020

ABSTRACT

Agriculture is the mainstay of the Indian economy and it is important to enhance the production with the help of technology. Crop production is a complex phenomenon that is influenced by various parameters like climatic conditions, fertilizers, production, rainfall, etc. In the present study Machine learning is used to predict the crop yield of finger millet using multiple linear regression analysis. Multiple linear regression is considered as a model for prediction and the accuracy of the model for the given data is significantly high when compared to other models. Object oriented Python is used and packages like NumPy and Pandas are utilized for performing operations and data analysis respectively. The main advantage of the research is the prediction of the approximate crop yield well ahead of its harvest, which would help the farmers in taking appropriate measures in crop cultivation, marketing and storage. Such predictions will also help the associated industries for planning the logistics of their business.

Keywords : Machine Learning, Multiple Linear Regression, Python, Numpy, Pandas

I. INTRODUCTION

Millets are coarse grains and a repository of protein, fiber, vitamins and minerals. Millets have high nutrition and are gluten free. Millets are a must have on the list of healthy diet. There are different varieties of millets like jowar, ragi, foxtail millet, bajra, etc. Finger Millet is one such variety which is cultivated in areas with rainfall of 700-1200mm and temperature of nearly 27 degree Celsius. It does not tolerate heavy rainfall. It is grown on red soil, yellow soil and laterite soils. Over the years several long – term fertilizer experiments conducted in different parts of India on crops like finger millet, rice, wheat, and maize showed wide variations in crop productivity as per the studies conducted by R. H. Wanjari etal.,[1]. Traditional agricultural practices to obtain higher crop yields include harrowing and ploughing using fertilizers, insecticides and herbicides Shakil Ahamed etal.,[2].

With the advancement in technology several methods are used to predict the crop yields with more precision. Several researchers used Data mining techniques to predict the crop yields Ramesh etal.,[3] Data mining techniques are used with agriculture data, the term is known as precision agriculture. The main aim of the work is to improve and substantiate the validity of yield prediction, which is useful for the farmers.

In the present study of the prediction of the yield of finger millet, Object oriented Python is used and packages like NumPy and Pandas are used for performing operations and data analysis respectively. Multi linear Regression is used for prediction since yield has to be predicted considering multiple independent variables. Similar studies were carried out in predicting the crop yields using different tools. Decision Support System for Agro-Technology Transfer (DSSAT) model under different climate scenarios was used in analyzing the performance of two varieties of millet yields Halimatou A etal, [4]. Regression technique carried out by B.Bapuji Rao etal, [5] resulted in identifying minimum temperature Tmin and moisture adequacy index MAI as the critical weather parameters influencing the millet yield in Mali. Moreover ML techniques will provide cost-effective and comprehensive solutions for better crop and environment state estimation as stated by Anna Chlingaryan etal.,[6] Using data on corn yield from the US Midwest, Andrew Crane-Droesch[7] showed that Machine learning in crop prediction outperforms both classical statistical methods and fully nonparametric neural networks in predicting yields of years. Machine learning is a class of computational methods for deriving insightful knowledge (including heuristics, strategies, or structure) from data, observations, or past solutions Shaw etal,.[8]. Adaptive modeling techniques derived from machine learning are increasingly being used in areas where there is little or incomplete understanding of the problem to be solved but where training data are available Park et al., [9].

II. METHODS AND MATERIAL

1. System architecture

The system architecture of the project which best describes the processes involved in analyzing the data and predicting the output is depicted in Figure 1.

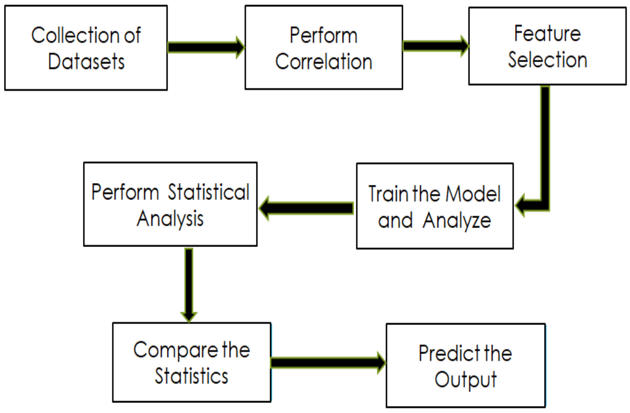

Fig.1 System Architecture for Finger Millet Yield Prediction

The initial step is to collect the data sets that are to be analyzed. The required attributes are to be selected and these are correlated with the dependent variable, which is the yield. Based on the correlation, the appropriate features are selected that are useful for analyzing and predicting the yield. Multiple linear regression is used as the machine learning model to make the yield predictions. Almost 70% of the original data is used for training the model. The coefficients for each input variable are calculated and this is done for all the 3 spatial locations that have different data of rainfall and temperature. The accuracy of the model is determined finally.

2. Selection of Input Variables:

The initial step for the analysis is to select the suitable agricultural factors that affect the growth of the crop. Based on the data availability, the factors that are considered are production, area under irrigation, rainfall, temperature, fertilizers used per hectare[10].

3. Data Collection:

The input variables are selected based on the availability of data. The production is taken in million tons pa, annual rainfall in mm, temperature in degree Celsius, area under irrigation in hectare, fertilizers (total NPK) in kgs. Rainfall and temperature data are collected for 3 different spatial locations like southern peninsular region, central India and the North-West regions of India where there is rich source of red, yellow and laterite soils which is best suited for cultivation of Finger Millet [11].

4. Training the Model:

Since the yield is a dependent variable which depends on various independent variables, multiple linear regression is used a regression model for predicting the yield, in Python. The factors are initially correlated so as to consider the best input variables for the model. Correlation value lies between -1 and 1. Then, the regression model is trained with the available data.

2.5 Testing and Predicting the Yield:

Simple Linear Regression faces the challenge of "outlier-sensitivity" and "Heterogeneous Datasets Management". Hence, multiple linear regression is considered as a model for a better prediction.

The accuracy of the model is determined and the coefficients for each input variable are calculated. The model is tested using the testing data and the yield is calculated using the multiple linear regression equation given below.

$Y = b_0+b_1X_1+b_2X_2+\ldots b_nX_n$ for 'n' observations

Where b_0 = y- intercept

X_i = independent variables [i = 1,2,3,…n]

Y = dependent variable

b_i = partial regression coefficients [i = 1,2,3…n]

6. System Requirements: It includes functional, hardware and non functional requirements.

i) Functional requirements:

- Operating System: Windows 7 and above
- Technology: Python 3.7

ii) Hardware requirements:

- Processor : GPU Processor
- RAM : Minimum 4 GB
- Hard Disk : Minimum 100 GB

iii) Non-functional requirements

- Performance requirements:As for the developed version of the project, the performance is limited to comparing utmost 3 geospatial locations.
- Safety Requirements: The safety requirements are nothing but an operation of backup of data on which we are performing analysis. The data is stored in csv file.

III. RESULTS AND DISCUSSION

Testing and predicting the yield

The accuracy of the model is determined and the coefficients for each input variable are calculated. The model is tested using the testing data and the yield is calculated using the multiple linear regression. The accuracy of the model for the given data is significantly high due to the use of regression model.

The implementation screen shots are given as follows.

```
In [2]: def final_yield(a,b,c,d,e):
            k=(a*85.14)+(b*39.06)+(c*0.0011)+(d*180.27)+(e*1.53)
            print(k)

In [3]: final_yield(3.19,3.8,622.38,23.48,113)

        4826.338818
```

```
train['Production'].fillna((train['Production'].mean()), inplace=True)
train['Area under irrigation'].fillna((train['Area under irrigation'].mean()), inplace=True)
train['Annual Rainfall in mm'].fillna((train['Annual Rainfall in mm'].mean()), inplace=True)
train['Average Temperature'].fillna((train['Average Temperature'].mean()), inplace=True)
train['Fertilizers'].fillna((train['Fertilizers'].mean()), inplace=True)
train['Yield'].fillna((train['Yield'].mean()), inplace=True)
```

```
x_train
```

	Production	Area under irrigation	Annual Rainfall in mm	Average Temperature	Fertilizers
73	3.65	2.6	721.4	24.00	224.59
161	12.20	4.9	784.9	24.20	135.27
61	15.86	NaN	784.4	24.67	230.15
53	14.98	19.1	788.7	24.72	235.35
82	8.03	4.0	910.4	24.26	217.58
51	13.16	20.5	528.4	24.73	237.05
123	9.15	NaN	767.7	24.67	165.60

```
lreg.score(x_cv, y_cv)

0.8363812885054043
```

	0	Coefficient Estimate
0	Production	85.144628
1	Area under irrigation	39.061940
2	Annual Rainfall in mm	0.001142
3	Average Temperature	180.279576
4	Fertilizers	1.533521

A 7 X 7 correlation matrix is created by using area, production, yield, area of irrigation ,annual rainfall, average temperature and fertilizers. Correlation values falls between -1 and 1.The closer the value of correlation (r) to 1 the data points fall to a straight line indicating a stronger linear association and when r is closer to 0 linear association becomes weaker [3].

	Area	Production	Yield	Area under irrigation	Annual Rainfall in mm	Average Temperature	Fertilizers
Yield	-0.531027	0.665348	1	0.874039	0.152793	0.396895	0.234878

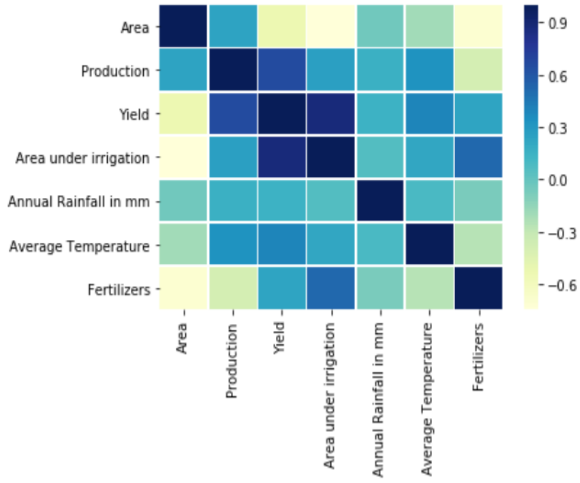

Fig. 2 Visual representation of correlation matrix

IV. CONCLUSION

Yield prediction is one of the most critical issues faced in the agricultural sector. Farmer's lack of knowledge about harvest glut, uncertainties in the weather conditions and seasonal rainfall policies, depletion of nutrition level of soils, fertilizer availability and cost, pest control, post–harvest loss and other factors leads to decrease in the production of the crops. Through the analysis and results obtained from this project, it would help the farmer to have an approximate idea about the yield and aid in emphasizing the need to cultivate millet crops in immense numbers to enhance the health and nutrition of people in our country. Also, the agricultural costs for millets is quite less when compared to others crops. This would not only provide a source of cultivation for farmers at reasonable price but also aid in sustaining the country's health. A comparison of three geospatial locations in India would provide significant analysis of area of cultivation for finger millets. There is also a great level of accuracy and the yield obtained by considering various factors together is high.

V. FUTURE SCOPE

Using price as a factor, considering the tension of supply-demand relationship, promotion of production cost and circulation cost for building long-effect mechanism for production and sales of agricultural products. Considering more factors for crop yield prediction by building high level models using polynomial regression, this could be achieved with large datasets and consistent data and to prevent from overfitting.

VI. REFERENCES

[1]. R. H. Wanjari,M. V. Singh &P. K. Ghosh Sustainable Yield Index: An Approach to Evaluate the Sustainability of Long-Term Intensive Cropping Systems in India Journal of Sustainable Agriculture ,Volume 24,2004(4).

[2]. A.T.M Shakil Ahamed, NavidTanzeem Mahmood, Nazmul Hossain, Mohammad TanzirKabir, Kallal Das, Faridur Rahman, Rashedur M Rahman,"Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh",IEEE,2015. DOI: 10.1109/SNPD.2015.7176185

[3]. D Ramesh, B Vishnu Vardhan,"Analysis of Crop Yield Prediction Using Data Mining Techniques", IJRET, 2015.

[4]. Halimatou A. Toure Idriss Serme, Kalifa Traore,Nicholas Kyei-Baffour, International Journal of Biological and chemical sciences, Issue Vol. 12 No. 1 (2018) DOI: 10.4314/ijbcs.v12i1.29

[5]. Bapuji Rao, B. and Rao, V.U.M. and Praveen Kumar, I. and Khandagonda, I.R. Finger millet production in Southern Karnataka - An agroclimatic analysis. Journal Agrometeorology, (2013) 15 (1-5). pp. 6-12.

[6]. Anna Chlingaryan ,Salah Sukkarieh , BrettWhelan Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review Computers and Electronics in Agriculture, Volume 151, August 2018, Pages 61-66. https://doi.org/10.1016/j.compag.2018.05.012.

[7]. Andrew Crane-Droesch, Machine learning methods for crop yield prediction and climate change impact assessment in agriculture,Environmental research letters, 2018,Volume 13, Number 11 https://doi.org/10.1088/1748-9326/aae159

[8]. Shaw, M.J. Machine learning methods for intelligent decision support: An introduction Decis. Support Syst. 1993, Volume 10, pp 79 -83.

[9]. Park, S.J., Hwang, C.S. & Vlek, P.L.G. 2005 Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions Agr. Syst. 85 59 81.

[10]. E. Manjula, S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

[11]. Prof. D.S. Zingade ,Omkar Buchade ,Nilesh Mehta ,Shubham Ghodekar ,Chandan , Crop Prediction System using Machine Learning ,Mehta,International Journal of Advance Engineering and Research Development Special Issue on Recent Trends in Data Engineering Volume 4, Special Issue 5, Dec.-2017.

Cite this article as :

Tarun Mourya. S, S. Nagini, Anne Slagha, "Application of Machine learning in Crop Yield Prediction of Finger Millet using Multiple Linear Regression", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 6, pp. 280-284, November-December 2020. Available at doi : https://doi.org/10.32628/CSEIT206616

Journal URL : http://ijsrcseit.com/CSEIT206616