IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF CROP YIELD PRODUCTION USING MACHINE LEARNING ALGORITHM

1J. VIJAYALAKSHMI, 2Dr.S.ARIFFA BEGUM

1ASSISTANT PROFESSOR, 2ASSOCIATE PROFESSOR

1VAIGAI COLLEGE OF ENGINEERING, MADURAI, TAMILNADU, INDIA,

2VAIGAI COLLEGE OF ENGINEERING

ABSTRACT

In today's scenario agriculture is an essential for day to day life for the survival of living. The farming land are covered in all over the region in the world, each area differs in their crop yield production. It's varied in different parameters such as water, ultra-violet (UV), pesticides, fertilizer and the land covered region. In this paper, the Machine Learning (ML) algorithm, the Linear Regression (LR) method is used to analyze and predicts the production of crop based on region and its weather report. Using the Linear regression method the performance analysis report is generated. Using that report the production can be enhanced.

Keywords—Machine Learning, Linear Regression Algorithm.

I. INTRODUCTION

The development of country's economy is dependent on the primary occupation that is agriculture for rural survival. Farming occupies major role in our country roughly its 70% for primary and secondary sectors. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance. In India, wheat and rice are the major grown crops along with sugarcane, potatoes, oil seeds etc. Farmers also grow non-food items like rubber, cotton, jute etc. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers. With this result, many farmers have begun to employ new technology and methods to improve their farming operations.

Most of the people are unaware of seasonal crops and its yield production. Crop development is a challenging factor that agriculture input parameters recommend. Many forecasting methods were proposed to identify trends, such as moving average models, exponential smoothing models, trend analysis, time series models, artificial neural networks, and grey forecasting models. Historical information regarding crop yield provides major input for companies engaged in this domain. Farmers experience was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc.

In data mining, unsupervised and supervised methods are being used. In unsupervised learning, clusters are formed using large data sets and in supervised learning classification are done based on the data sets. In clustering technique, 'data points' are examined to group them into 'clusters' according to specific parameter. The data points in same cluster have less distance compared to data points of different clusters. The analysis of the cluster divides data into well organized groups. With these statistical predication methods may not be used for predicting real agriculture output data. To solve this problem, the Linear regression has been used to improve the efficiency and performance of agricultural output forecasting. To enhance efficiency Linear Regression algorithm shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis).

II. RELATED WORK

F F Haque et al., (1) evaluated and implements two different Machine Learning (ML) algorithms are proposed to analyze the crops' yield. These two algorithms, Support Vector Regression (SVR) and Linear Regression (LR), are quite suitable for validating the variable parameters in the predicting the continuous variable estimation with 140 data points that were acquired. The parameters mentioned above are key factors affecting the yield of crops. The error rate was measured with the help of Mean Square Error (MSE) and Coefficient of Determination (R2), where MSE gave out approximately 0.005 and R2 gave around 0.85. The same dataset has been used for quick comparison between the algorithms' performances.

In Ting-Zuo Jheng et al.,(2) the experiment results show that the performance of hybrid SVR models are better than the traditional SVR model in terms of the root mean square error (RMSE) and the correlation coefficient (CC). The average RMSE and CC of hybrid SVR models are 60 and 0.996, respectively. Accordingly, hybrid SVR models are suitable to predict agricultural output because it provides high reliability and high stability prediction results.

In Pallavi Kamath S et al., (3) the crop production prediction involves a huge amount of data, making it a perfect candidate for data mining methods. Data mining is method of accumulating previously unseen anticipated information from vast database. Data mining assists in the analysis of future patterns and character, enabling companies to make informed decisions. For a specific region, this research provides a fast inspection of agricultural yield forecast using the Random Forest approach.

III. METHODOLOGY

After going through a series of learning data, the data pre-processing takes place, and every data gets sorted according to the selected category (based on year & temperature or year & rain). There were around 140 data that was taken manually from data.gov.in with the correct yield information to make the accuracy more valid to real data. The validation steps check whether 80% of all the data are being correctly synchronized with the set of rows of information being called out for training simulation. In addition, with the help of Linear regression algorithm that tests the data is changed after each interval for better prediction of the outcome. And finally, the output comes out as the prediction result of the model that has been deployed and the visual analytical report has been generated based on LR algorithm. The system flow has been shown with training and testing of the process is shown in Fig 1.

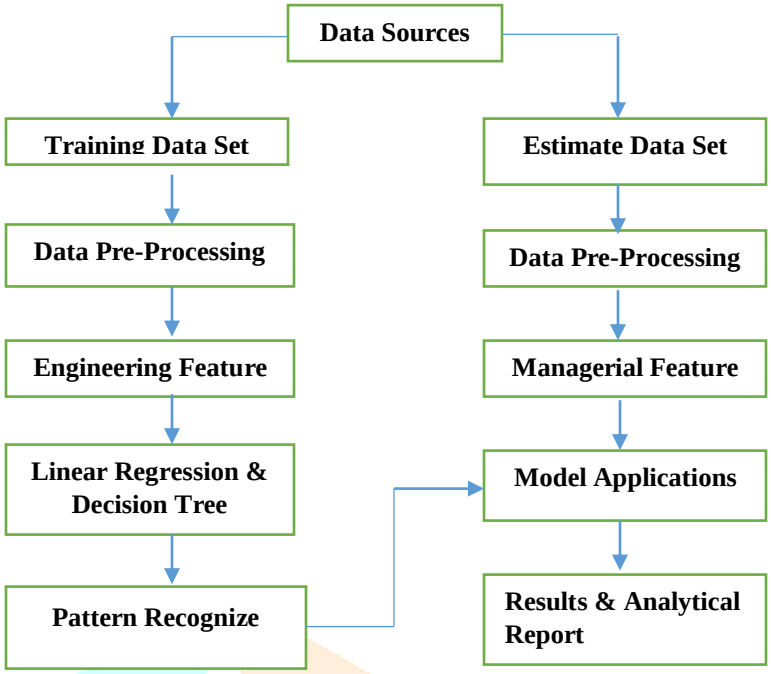

Fig.1 Workflow Model

A. Exploring Data – Linear Regression Model

To enhance efficiency Linear Regression algorithm shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), (Fig.2) from this to find out if there is any correlation between these two variables, find the best fit line for the dataset and to predict the dependent variable is changing by changing the independent variable.

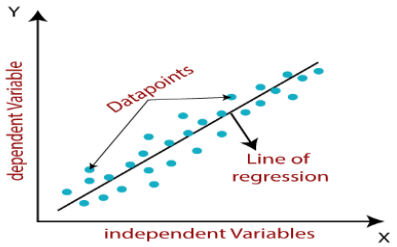

Fig.2

Temperature prediction – Linear Regression

```
def _predict_temp(model, year):
……..
        for j in range(0, model.shape[0]):
            eq_sum += (model[j] * (x**(model.shape[0]-j-1)))          (1)

        if eq_sum < 0:
            eq_sum = 0
        print(eq_sum)
        total += eq_sum
    avrg = total/12
    return avrg
```

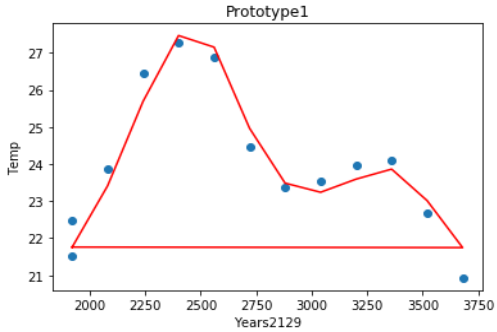

Fig 3 Tempature prediction

Rainfall prediction – Linear Regression

```
def _predict_rain(model, year):

    for j in range(0, model.shape[0]):
        eq_sum += (model[j] * (x**(model.shape[0]-j-1)))
    if eq_sum < 0:
        eq_sum = 0
    print(eq_sum)
    return eq_sum
```

(2)

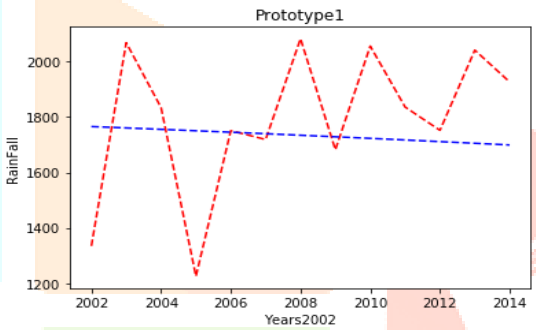

Fig 4 Rainfall prediction

IV. RESULTS

After the data has been made ready to place in the models for the breakdown, it is scattered into different training and testing data. While evaluating the performance with the same ratio of train: test, the result won't be evaluated as a better conclusion for the yield prediction. Hence, there are a few actions that can provide a better model for the error analysis.

The following table (Table1) shows the sample data from the Learning dataset.

Table 1 Learning Data Set (Year, Month and Temperature)

```
['Year', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov',
'Dec']
['1901', '22.729', '23.788', '25.391', '27.646', '25.986', '24.102', '23.22', '23.297',
'24.478', '23.985', '22.746', '20.736']
['1902', '21.589', '22.981', '26.246', '27.356', '26.855', '24.879', '23.762', '24.247',
'23.417', '23.625', '22.406', '21.75']
['1903', '22.07', '23.639', '26.217', '27.584', '26.357', '24.625', '23.275', '23.177',
'23.434', '23.229', '21.37', '20.335']
['1904', '20.955', '22.63', '25.454', '27.349', '25.686', '23.185', '22.658', '23.307',
'23.819', '23.873', '21.8', '20.71']
['1905', '21.529', '23.807', '25.695', '26.638', '26.599', '24.776', '23.677', '23.623',
'23.83', '23.839', '22.752', '20.9']
```

['1906', '22.59', '24.262', '25.404', '28.566', '27.821', '24.36', '23.257', '23.226', '22.797', '23.568', '22.829', '21.38']

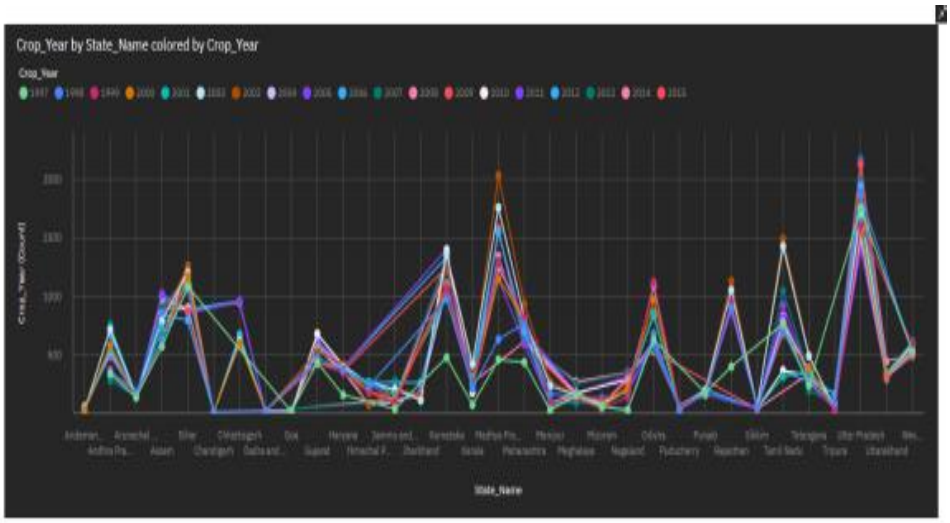

Fig 5 Analysis report for Crop – State-wise

Fig 6 Analysis report for Crop – Year-wise

The above results show the performance of the yield estimation of different categories. From the result it is easy to identify the time and place to get highest yield. The implementation of this liner regression method really identifies exact results of crop yield production increase and decline.

V. CONCLUSION

Thus the Machine Learning (ML) algorithm Linear Regression (LR) method has been used to analyze and predicts the production of crop based on region and its weather report. The results clearly represented using graphs which can be easily readable and understandable. By this way of representation the farmers can easily select the exact time to start their seeding process. So using this crop yield production definitely increases and everyday need on the yield measurement had also be done successfully.

REFERENCES

[1] F. F. Haque, A. Abdelgawad, V. P. Yanambaka and K. Yelamarthi, "Crop Yield Analysis Using Machine Learning Algorithms," 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2020, pp. 1-2, doi: 10.1109/WF-IoT48130.2020.9221459.

[2] T. -Z. Jheng, T. -H. Li and C. -P. Lee, "Using hybrid support vector regression to predict agricultural output," *2018 27th Wireless and Optical Communication Conference (WOCC)*, Hualien, Taiwan, 2018, pp. 1-3, doi: 10.1109/WOCC.2018.8372729.

[3] Pallavi Kamath, Pallavi Patil, Shrilatha S, Sushma, Sowmya S," Crop yield forecasting using data mining", Global Transitions Proceedings, Volume 2, Issue 2,2021,Pages 402-407,ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.08.008.

[4] Anjana, Aishwarya Kedlaya K, Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat, "An efficient algorithm for predicting crop using historical data and pattern matching technique", Global Transitions Proceedings, Volume 2, Issue 2, 2021, Pages 294-298, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.08.060.

[5] "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", B M Sagar, Cauvery N K Department of Information Science & Engg, R V College of Engineering, Bengaluru, Karnataka, India

[6] Bm, Sagar & Cauvery, N.K.. (2018), "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science. 12. 1087-1093. 10.11591/ijeecs.v12.i3.pp1087-1093.