

Prediction of Crop Production through Hybrid Approach using Machine Learning Algorithms¹Sudhendra Pal Singh, ²Dr. Ajay Khuteta, ProfessorE- Mail: sudhendra.jain@gmail.com , khutetaajay@poornima.org

Poornima College of Engineering, Jaipur

Abstract

Ensuring food security is an important task in many countries. With the growth of the population of the Earth, issues of increasing the efficiency of agriculture become the most urgent. Farmers are looking for new ways to increase yields, and governments of different countries are developing new programs to support agriculture. In this paper we have implement a new methodology for prediction of suitable crop production in any particular area/district. In this work Linear regression and Random Forest model are used for prediction and result show that accuracy is not very high so proposed a new model which overcome the limitation of Linear Regression and random forest model and give the good accuracy of prediction model. Result show that Accuracy of Linear Regression model is calculated 63.73% .Ridge Regression model is created using Python library and trained it with same agriculture dataset. Accuracy of the model is calculated 63.73%Random Forest is also trained with the same agriculture dataset and Accuracy of the model is calculated 87.08, Proposed Hybrid Approach give the high accuracy for huge dataset. So in order to achieve the high accuracy we used the hybrid model approach.

Keyword: Food Security, Blockchain, Irrigation Technologies, Splitting Data**1. Introduction**

The agro-food sector faces numerous Problems. If in 2018 the population of the planet was 7.6 billion people, then by 2050 it is, according to preliminary estimates, exceed 9.6 billion, which will lead to a significant increase food needs (UN DESA, 2017). In the same time available natural resources, including fresh water and productive arable land is getting smaller. Production is not the only factor causing concern: agricultural produced today is enough to feed the whole world, and yet 821 million people on the planet still suffer from hunger (FAO, 2018). Other processes for example, high-speed urbanization, also have a strong impact on production patterns and food consumption. The solution of the task can be provided by digital innovation and technology like Blockchain, Internet of things and artificial intelligence etc. now extends access for small farmers to information, production resources, market, finance and training. Digital technology opens up new opportunities for integrating small farms into digital agri-food systems.

1.1 Use of machine learning in agriculture Sector:

Learning technologies are already being introduced to agriculture. Machine learning is a technology that takes input, analyses it, learns and makes decisions without human intervention. A set of machine learning methods that work with different sets of raw data and find solutions is called deep learning.

1.2 Forecast of crop productivity depending on conditions:

Machine learning improves plant growing accuracy. The farmer can evaluate a much wider set of factors. Computer modelling will provide an assessment of how different crops will react to different soil types, weather conditions, and many other factors. Such digital testing does not replace field tests, but allows breeders to more accurately predict crop productivity.

1.2.1 Intellectual irrigation with AI support: Modern irrigation technologies with ML are able to distinguish weeds from crops and spray the first herbicides. As a result, less chemicals are used and the cost of growing crops is reduced.

1.2.2 Satellites with artificial intelligence: a startup Harvesting analyses data from a satellite and predicts corn yield using a patented machine learning algorithm. This algorithm has learned to evaluate the state of plants.

1.3 The role of machine learning in the modern agricultural era

Machine learning is a trend of today's technology and can be used in the modern agricultural industry. ML applications in agriculture contribute to healthy seeds. Artificial machine learning is one of the most dynamic areas in agriculture. Artificial techniques are used in agriculture to increase accuracy and find solutions to problems. Agriculture plays a very important role in the global economy. Due to population growth, there is constant pressure on the agricultural system to improve crop productivity and produce more crops.

1.3.1 Machine Learning Methods

ML gains experience in performing models over time. Different mathematical and statistical models are used in ML model and machine learning algorithm agriculture to determine performance. After you finish the learning process, the model can then be used for data estimation, sorting and testing. This training process is done after gaining experience. Learning machine functions can be divided into two categories, namely supervised and unsupervised learning.

1.3.2 Popular application of machine learning (ML) in agriculture

- **Agricultural Robots:** Most companies are now working on programming and designing robots to handle the necessary tasks related to farming. These include harvesting and working faster than human workers below. This is an excellent example of a learning machine in agriculture.
- **Culture and Floor Monitoring:** Now companies are in-depth using technology and learning algorithms. Data is collected using drones and other software to monitor cultivation and soil. They use software to manage soil fertility. Using new technology in agriculture, farmers can find effective ways to preserve their crops and protect them from weeds.

2. Literature Review

[1] Priya et al., 2018 studied crop yield prediction ability using Random Forest Algorithm. This work took the data from the openly accessible records of Government of India. Proposed work only used single algorithm for the yield prediction based on certain parameters. This work does not show any comparative analysis of accuracy among various machine learning algorithm to choose an efficient algorithm for crop yield prediction. This is basic limitation of this paper and author reports to determine the same in future enhancement.

[2] Arun Kumar et al., 2018 this research work performs efficient crop yield prediction using supervised machine learning algorithms like KNN, SVM, LS-SVM. This research work is based on data from agriculture production, rainfall and soil data and it shows the comparative analysis of these machine learning algorithms. As result shows that algorithm are strong enough to perform the efficient prediction of crop yield, linear version of SVM is most accurate than KNN and SVM. Random Forest regression is also very strong supervised machine learning algorithm, which uses the concept of multiple decision tress and support both type of problems such as classification as well as regression problems. This algorithm may be used for performing the more efficient crop yield prediction with better accuracy.

[3] S.Veenadhari et al., 2014this research work performs forecasting of crop yield based on climatic parameters. The author adopted decision tree classifier for classifying the crops on the basis of climatic parameters. Algorithm is used to find out the most effective climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. For example- for Soyabean crop the most influencing parameter was cloud cover and for wheat crop it was minimum temperature. There are many other parameters can be considered for crop yield prediction like production area, pesticides etc. but in this study only climatic parameters were used for predicting the crop yield.

[4] E. Manjula et al., 2017this research work proposed the data mining technique for crop yield prediction. So for estimating the crop production in the future years, association rule mining approach is applied on agriculture data of 31 districts of Tamil Nadu for the years from 2000 to 2012. In this work only Tamil Nadu state dataset was used which can be further enhanced by using other neighbour state data also and with increasing size of data, achieving accuracy is always an important factor for precise estimation of crop yield so other machine learning algorithms may be used to deal with larger size of data with better accuracy.

3. Proposed Model

We have implemented all the processing code algorithms and programmable data using the Python language. The IDE that we used is Anaconda and Jupiter Notebook, which is a very effective prime Python IDE for programming related to data science.

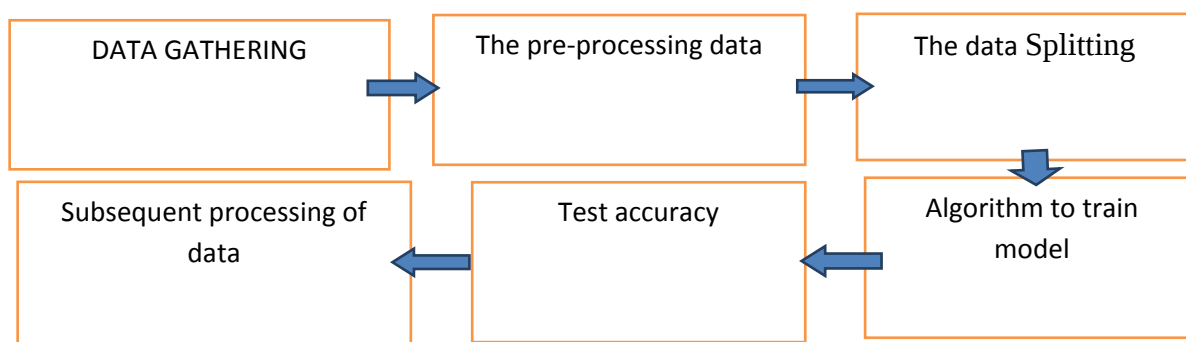


Figure1: Block Diagram of proposed model

3.1 Data gathering

Data collection for this kind of investigation, it is essential to have a set of data available for work. It is very difficult to find readable and reliable sets of data. It took a lot of time and effort to find a suitable for us. The data set containing

exact features that really wanted. Following columns were chosen for our research work. The columns were "District Name", "Crop Year", "Type of Crop", "Crop Production" and "Area of Production". The "District Name" column were the names of the districts in states. The column "Crop Year" was the year of harvest for the past few years.

3.2 The pre-processing data

One of the main tasks was to convert our entire data into desired format which is essential to perform the data modelling. To do this, we filtered our entire data set district wise then crop wise. We created District as a data folder then crop as a sub folder into district folder. We chose to use the data of Rajasthan state and its all districts as our data set. In each of the subfolder we store the .csv file which contains Year of Production, Area of Production and Total Production.

3.3 Data Splitting

Splitting data is the division process the data set into training and test data. This process is very useful for any learning process of the machine as the main idea of learning the machine depends on test data and training and research of the accuracy of the result given by machine. The algorithms were trained and we will apply the trained algorithm to predict target data and thus precision of the machine is measured. There are many built in Python toolkits to divide the data, we used "scikit-learn" for machine learning methods in this research because of its built-in libraries.

3.4 Algorithm used to train the model

- Training: Proposed model is trained on training data sets stored in form of .csv file in a particular district data folder and crop sub folder using the combination of both linear and random forest regression as hybrid approach
- Trained model is then pickled on to persistent media for further predicting the production.
- Regression: On the basis of user input following steps will take place for predicting data.
 - Corresponding data model will be unpickled or loaded for predicting the crop production on the basis of name of district and crop name.
 - Crop Year and crop area will be used as predictors by linear regression and random forest regression methods to predict the amount of production in a particular district and for particular crop.
 - Output of both the method will be used to predict the final crop production.
 - Here, Linear Regression and Random Forest Regression both are used to form the proposed hybrid technique.
 - Random forest (RF) is a combination of decision trees.

3.5 Methodology

Following are the steps were followed to prepare the data for processing.

Step 1: Acquisition of each parameter (District Name, Crop Type, Production Area, and Total Production) for each district of Rajasthan from the openly accessible Government of India data for year 1997-2010

Step 2: Data set is filtered out on the basis of Districts Name and Type of Crop and then Data folders are created on the basis of District Name and Crop Type.

Step 3: Then the raw data set is collected on the basis of District and Crop type which contains the following columns: Year, Production Area and Total Production.

Step 4: Data which was not available for some of the given parameters, that data is not used for the actual study.

Step 5: Once the data set is prepared, we trained the proposed hybrid model for each of the data sets.

Step 6: Then this hybrid model predicted the total production for a particular crop in a particular district.

Step 7: When proposed model is asked for predicting production of a crop, it asks the user for the District Name, Crop Type with area of production.

Step 8: On the basis of District Name and Crop type appropriate model will be loaded for predicting the production.

3.6 Evaluation of performances

Following is the performance evaluation of various machine learning methods with their score of accuracy.

```
In [34]: fr = RandomForestRegressor()
fr.fit(X_train,y_train)
fr.score(X_train,y_train)
```

```
Out[34]: 0.8708855045722896
```

```
In [35]: rl = LinearRegression()
rl.fit(X_train,y_train)
rl.score(X_train,y_train)
```

```
Out[35]: 0.6373943953621659
```

```
In [36]: rd = Ridge()
rd.fit(X_train,y_train)
rd.score(X_train,y_train)
```

```
Out[36]: 0.6373935734861627
```

```
In [37]: stregr.score(X_train,y_train)
```

```
Out[37]: 0.9372416190009722
```

S.No.	Machine Learning Methods	Accuracy Score
1	Linear Regression	0.63739
2	Random Forest Regression	0.87088
3	Proposed Hybrid Model(Linear Regression +Random Forest Regression)	0.93724

4. Results

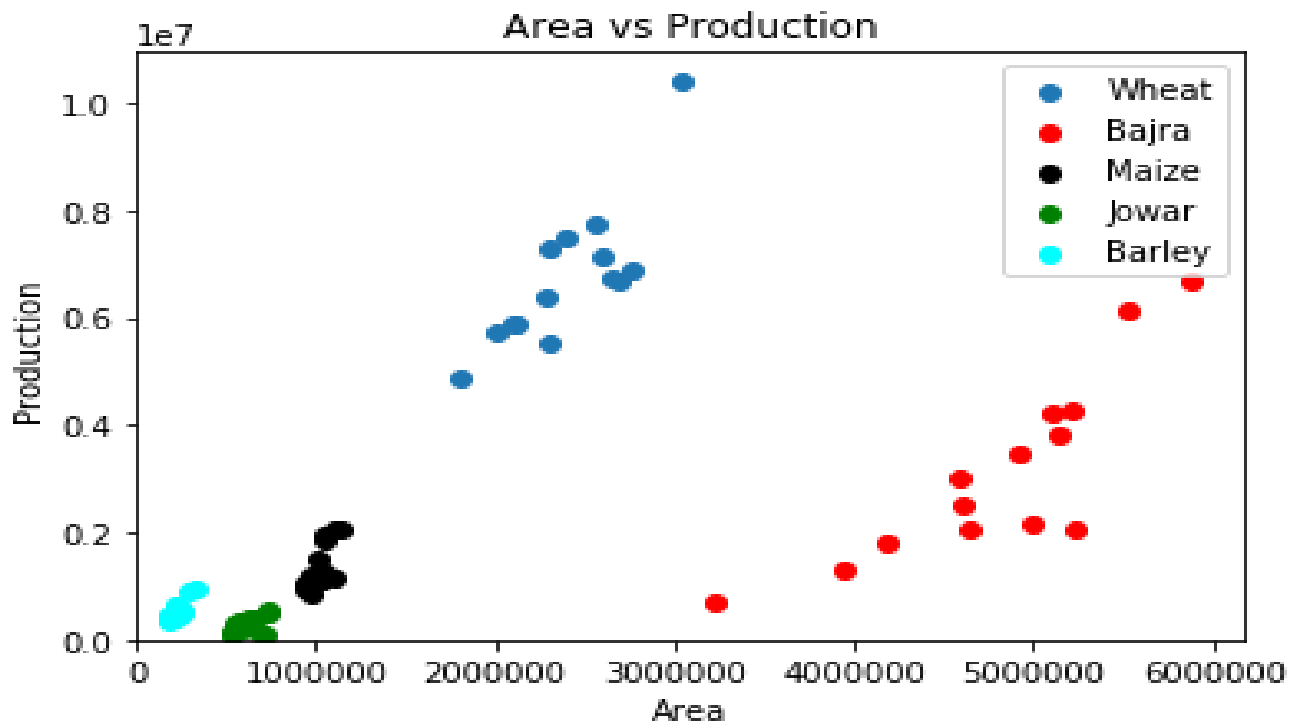


Figure 2: Crop production distribution (Area vs Production)

The above chart displays the observation points between Production Area and Total Production for various crops like Wheat, Bajra, Maize, Rice and Barley from the previous year's Statistical dataset. Different colours are used to show the observation points for different crops. In the Statistical Analysis of existing data we can see that Wheat and Bajra needs larger production area as compared to other crops.

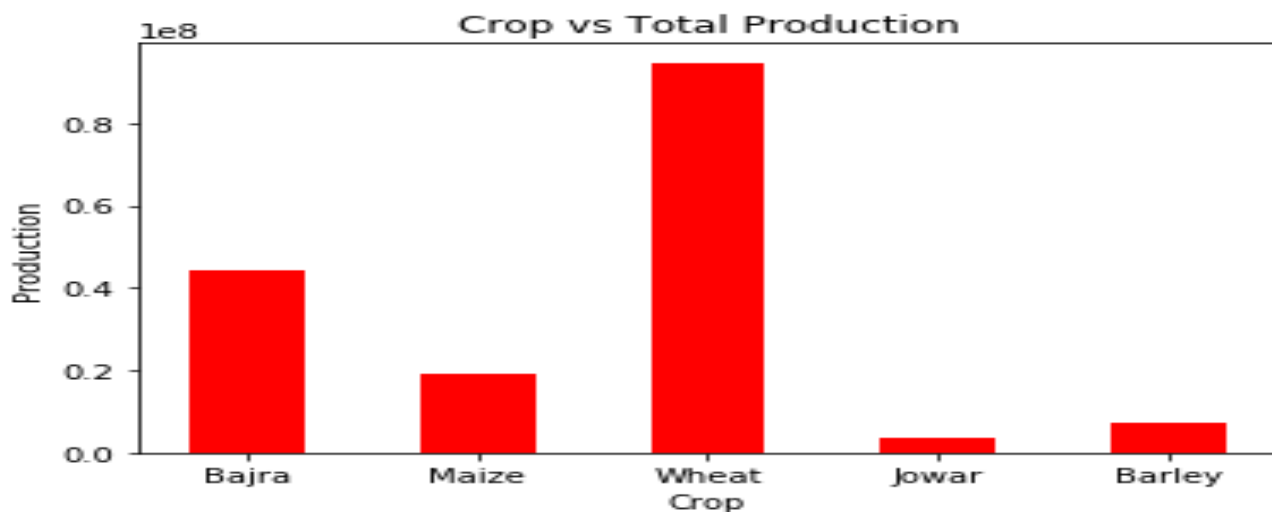


Figure 3: Different Crop Production Distribution

The above chart displays the relation between crop and their production for various crops like Wheat, Bajra, Maize and Barley etc. from the previous year's dataset. The Statistical Analysis of existing data in above chart shows that the

production of wheat, Bajra and Maize is highest than other crops that's way we can say that Wheat, Bajra and Maize crops have the more suitable environment in Rajasthan.

Different models are trained with the common dataset and then these models are used for crop yield prediction and their accuracy of prediction is calculated as per below details:

- a) Linear Regression – Linear model is created using Python library and trained it with real time agriculture dataset. Accuracy of the model is calculated 63.73%
- b) Ridge Regression - Ridge Regression model is created using Python library and trained it with same agriculture dataset. Accuracy of the model is calculated 63.73%
- c) Random Forest Regression – This model is based on the concept of decision trees.

It creates the multiple decision tree for predicting the output and finally combine the result of all decision trees to get the final prediction. This model is also trained with the same agriculture dataset and Accuracy of the model is calculated 87.08

- d) Proposed Hybrid Approach – Above mentioned model are good for machine learning purpose but may not give the high accuracy for huge dataset. So in order to achieve the high accuracy we used the hybrid model approach.

5. Conclusion

Proposed approach merges the concept of linear, Random Forest model and creates a hybrid model with the improved accuracy calculated 93.72. In the proposed work accuracy is very important factor, model with the high accuracy can precisely predict the production of specific crop in the specific location and the same can be informed to farmer that what amount of crop can be produced in his/her production area. Balanced production of the crops will not affect the production of other suitable crops in the same location and this way it will not lead the increment in the cost of crop due to shortage of production and will fulfil the need of every people within the cost.

6. References

1. P.Priya, U.Muthaiah, &M.Balamurugan (2018). Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Sciences & Research Technology, ISSN: 2277-9655
2. Arun Kumar, Naveen Kumar, & Vishal Vats (2018). Efficient crop yield prediction using machine learning algorithms. International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, p-ISSN: 2395-0072
3. S.Veenadhari, Dr. Bharat Misra, &Dr. CD Singh (2014). Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).
4. E. Manjula, & S. Djodiltachoumy (2017). A model for prediction of crop yield. International Journal of Computational Intelligence and Informatics, ISSN: 2349-6363
5. Balaghi, R., Tychon, B., Eerens, H., &Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10, 438–452.

6. Donohue, R. J., Lawes, R. A., Mata, G., Gobbett, D., &Ouzman, J. (2018). Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Research*, 227, 79–90.
7. Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., et al. (2003). DSSAT cropping system model. *European Journal of Agronomy*, 18, 235–265
8. Kantanantha, N., Serban, N., & Griffin, P. (2010). Yield and price forecasting for stochastic crop decision planning. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 362–380.
9. Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., et al. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267–288.
10. Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., et al. (2017).