Crop yield production using machine learning- Data Analytics Phase 1 Report

Chand Kumar Adhaduk  
Computer Science & Engineering  
PES UniversityBangalore,India  
adhadukchand@gmail.com

Mahendra N  
Computer Science & Engineering  
PES UniversityBangalore,India  
mahendranerella2002@gmail.com

Bhavana Chowdary J  
Computer Science & Engineering  
PES UniversityBangalore,India  
bhavanachowdaryjetti420024@gmail.com

***Abstract*— In India agriculture is, no doubt, the major and least paid occupation in the country. Not only that it is the backbone of India. With the help of Machine learning , we can help the farmers by predicting the yield of the crop so that they can plan in advance, so that they can be prepared beforehand. With the help of the model, they can decide which crop to grow so that they can get the required yield. By using attributes like State Name ,Season ,Crop ,Area we will be able to predict the yield of the crop.**

***Keywords—agriculture, machine learning, yield of the crop***

# Introduction

Agriculture in India dates back to Indus valley civilization,with its allied sectors, is unquestionably the largest livelihood provider in India, more so in the vast rural areas.Sustainable agriculture, in terms of food security, rural employment, and environmentally sustainable technologies such as soil conservation, sustainable natural resource management and biodiversity protection, are essential for holistic rural development. Also India ranks second when it comes to agriculture. Total GDP(Gross Domestic Product) of forestry and fishery combined is 15.4 % whereas agriculture alone contributes 31% of GDP which is very large compared to other occupations and also the workforce is 59%. In terms of net cropped area, India comes first and then US and China comes after India.India is an agriculturally developed country.

Predicting the yield of the crop using machine learning will surely help. Farmers can plan beforehand and decide which crop to grow. There are a large number of machine learning models which predict the crop yield using different input attributes, some use attributes like temperature and rainfall, while other use static attributes like season, area, state. To make it simple we will be using attributes like crop, area , state, Season to predict the crop yield.

# REVIEW OF LITERATURE

In [1] authors predict the yield of almost all kinds of crops that are planted in India. Their data contains attributes like State, district, season, area and year that are used to predict the crop yield Methodology used by them is Stacked Regression . This is a kind of ensembling but a little of enhancement of averaging. In this, we add a meta model and use the out-of-fold predictions of the other models used to train the main meta model. Performance metric used by them is Root mean square error. When the models applied individually, for ENet it was around 4%, Lasso had an error of about 2%, Kernel Ridge was about 1% and finally after stacking it was less than 1%.

In [2] authors tried to focus on predicting the yield of the crop by applying various machine learning techniques. Their dataset contains attributes Temperature and Rainfall .Methodology implemented by them are ensemble learning algorithms like Random Forest Classifier and XGBoost , KNN Classifier, Logistic Regression, Linear Regression and Artificial Neural Networks. Performance metric used by them is MEAN ABSOLUTE ERROR .With 22.17 for Simple RNN , with 34.14 for LSTM.

In [3] authors predict the yield of the crop using the Random Forest algorithm based on existing data. Data from TamilNadu was used as a training set for construction of the models.Number of attributes were 7 namely Soil Type, Soil value of ph, Climate Parameters Like Wind Rainfall Humidity Temperature, The cultivation expenses ,and Manufacturing Support Vector machines were also used for this project. The methodology used here is CSM.This is helpful as it tracks the seasonal yield of crops and tries to predict based on that. At the end the Random Forest Algorithm is used as it shows good accuracy. Accuracy of the model built was 75%.

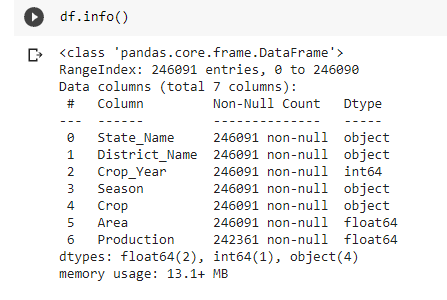
In [4] authors predict crop yield and success rate for the crops as per input given by the farmer. The data used for research is resourced from the Indian government website, consisting of records from the year 1997 to 2014 Maharashtra state was considered as a study area. Number of rows were 12000 Columns were DISTRICT,SEASON,CROP,YEAR,PRODUCTION(tons per hectare in lakh) Model used is a 3 Layer ANN with linear regression with forward and backward propagation. Methodology used was both binning and encoding for analysis. dataset was Split into dev-test and train set using 80:20 assignment. Model is trained Using linear Regression with Neural Network. The Learning rate for each layer is kept constant i.e. 0.001 Reduces MSE by using Adam optimizer, RELU activation function and gradient descent. Accuracy of 82% was obtained.

In [5] authors predict crop yield in India using ML techniques. The Dataset in this research is collected from www.mospi.gov.in.Columns were rainfall, area, area under irrigation, crop names, seasons, production, and yield for the year 1950 to 2018. Models used were ML techniques such as Decision Tree, Linear Regression, Lasso regression, and Ridge Regression. The prediction is made for five crops which are Rice, Wheat, Jowar, Bajra, Tobacco, and Maize.MAE, RMSE were used to validate.The prediction is attained using decision tree and random forest. Accuracy obtained is as follows, Decision Tree:98.62% ,Linear Regression:89.38% ,Lasso regression:86.33% ,Ridge Regression:89.53%. The Decision tree outperforms other machine learning techniques.

DATA ACQUISITION

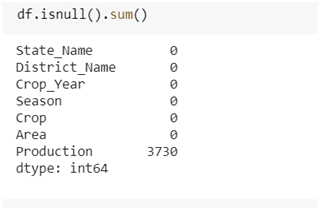
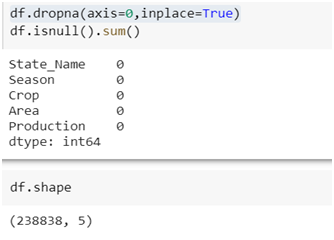
The selected data has a huge amount of information on crop production in India ranging from several years. The data was taken from<https://data.world/thatzprem/agriculture-india> .

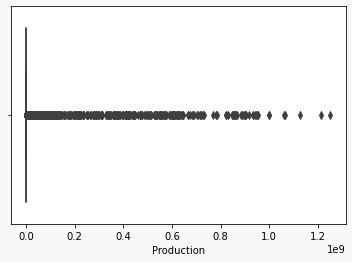
Dataset has 246091 entries and 7 columns in which four are categorical. Based on the information the ultimate goal is to predict the crop production using machine learning techniques.The following attributes must be given to the Dataset:State\_Name,District\_Name,Crop\_Year,Season,Crop,Area,Production.



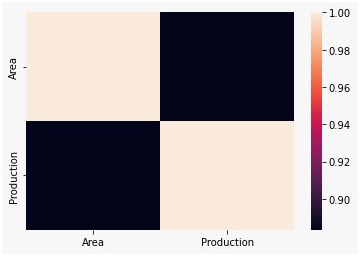
As we are not working with Crop\_Year and District\_Name we have dropped it from the dataset.

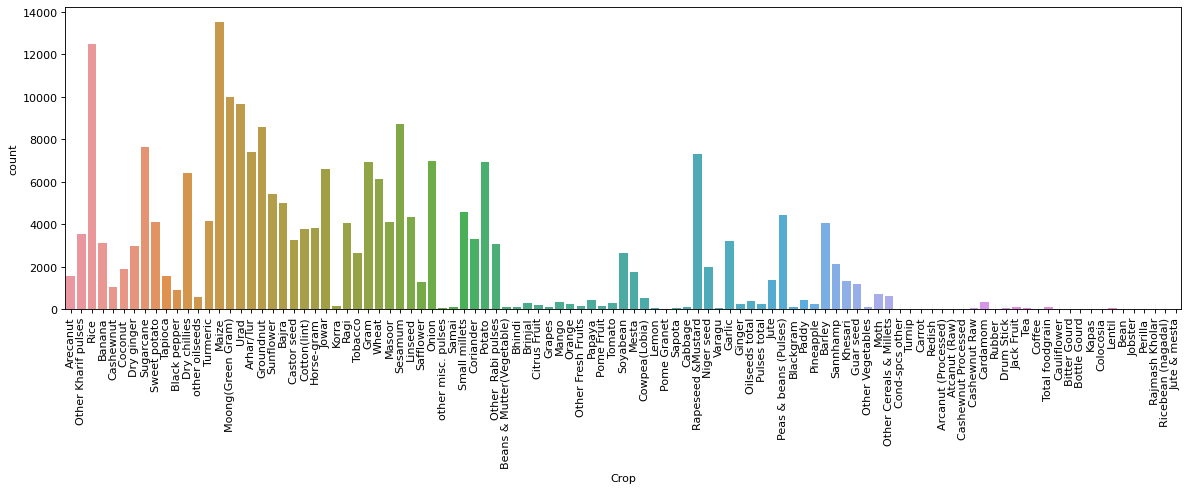
METHODOLOGY

Null values in data before preprocessing:Null values are replaced with Nan values in the data frame and the Nan values are dropped when needed. Null values in data after pre-processing:

It's clear from the graph that there are many outliers but removing them wouldn't be a good approach because we would be losing a lot of data. Therefore we decided to remove only the most extreme values.

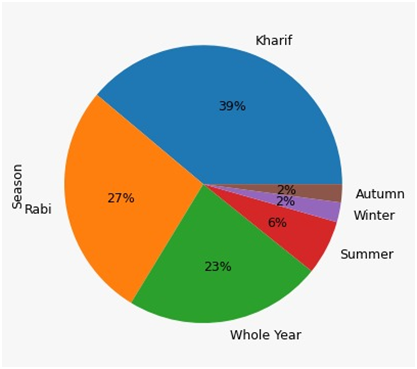
# Few observations and inferences:

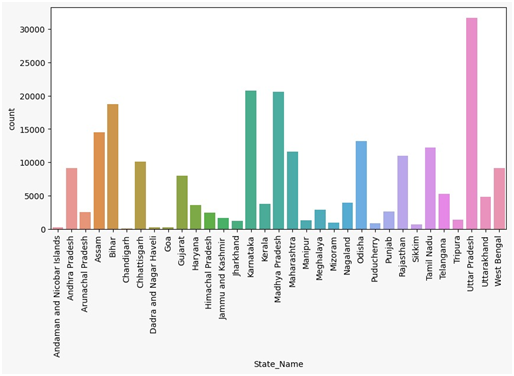
Area and Prediction columns of the dataset are plotted to check for correlation as shown below:Inference: Area and Production are highly correlated.

A bar plot showing the count of different crops is plotted as shown below:

Inference: Top 5 crops grown in India are Maize,Rice,Moong,Urad,Sesamum

A pie chart showing the distribution of crop growth in different seasons is plotted as shown below:

Inference:In most parts of India Kharif and Rabi are the season for cultivation whereas Autumn and Winter are least and In few places crops are grown whole year.

A bar plot showing distribution of crop growth in different states is plotted as shown below:

Inference:Uttar Pradesh,Karnataka,Madhya Pradesh are the top agricultural states in India due to high-yielding varieties of seed,greater availability of fertilizers and increased use of irrigation

##### Acknowledgment

We would like to acknowledge our Data Analytics Course Professor Dr. Gowri Srinivasa,assistant professors and teaching assistants for providing constant guidance during each phase of our project.

References

1. Potnuru Sai Nishant, Pinapa Sai Venkat ,Bollu Lakshmi Avinash, B. Jabber, “Crop Yield Prediction based on Indian Agriculture using Machine Learning”, In:2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.
2. Aruvansh Nigam, Saksham Garg, Archit Agrawal, “Crop Yield Prediction Using Machine Learning Algorithms”, In:2019 Fifth International Conference on Image Information Processing (ICIIP).
3. Namgiri Suresh,N.V.K.Ramesh,Syed Inthiyaz,P. Poorna Priya,Kurra Nagasowmika,Kota.V.N.Harish Kumar,Mashkoor Shaik, “Crop Yield Prediction Using Random Forest Algorithm”,In:2021 7th International Conference on Advanced Computing & Communication Systems (ICACCS).
4. Shivani S. Kale,Preeti S. Patil, “A Machine Learning Approach to Predict Crop Yield and Success Rate”, In:2019 IEEE Pune Section International Conference (PuneCon) MIT World Peace University, Pune, India. Dec 18-20, 2019.
5. Ms Kavita ,Pratistha Mathur, “Crop Yield Estimation in India Using Machine Learning”, In:2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) Galgotias University, Greater Noida, UP, India. Oct 30-31, 2020.