ANALYSIS OF CROP PRODUCTION AND RAINFALL IN INDIA

Anmol Garg
Ashwin Siddhart S
B R Yogesh

Problem Statement

Analysis of agricultural data to better understand crop production in India. We collect data about crop production and rainfall over the years to analyze the trend and get insights, predictions and recommendation from the data.

RELEVANCE

Multitude of domains have benefited from the advancement of technologies but the usage of technology in Agriculture sector in India has been very limited. Indian agriculture is hassled by several problems which are resulting for the decline in production. One of the most innovative pieces of the digital transformation is the ability to use machine learning and advanced analytics to mine data for trends. The insight provided allows farmers to sow and harvest their crops at the optimum time, which maximizes crop yields.

How do you do the analysis?

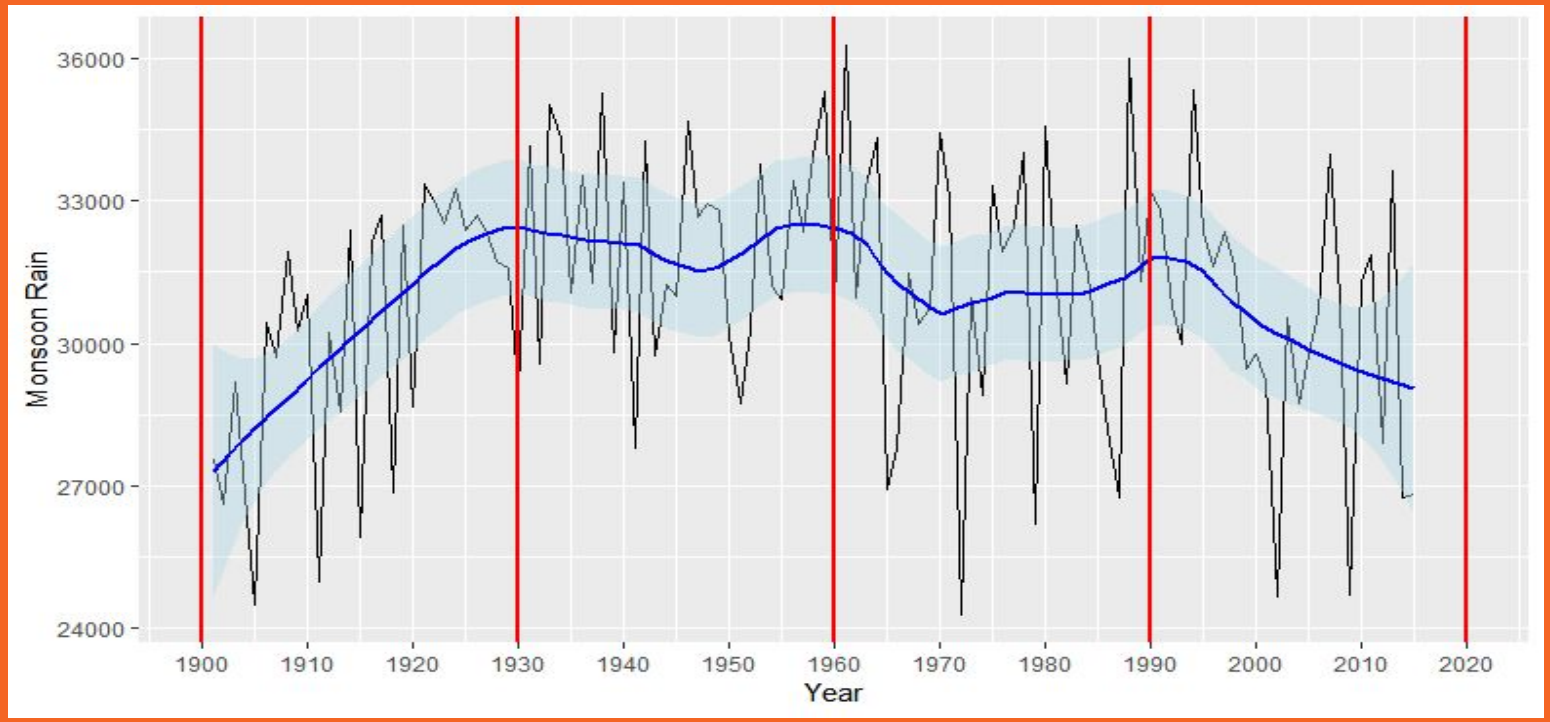
Lots and lots of DATA

Our Approach.

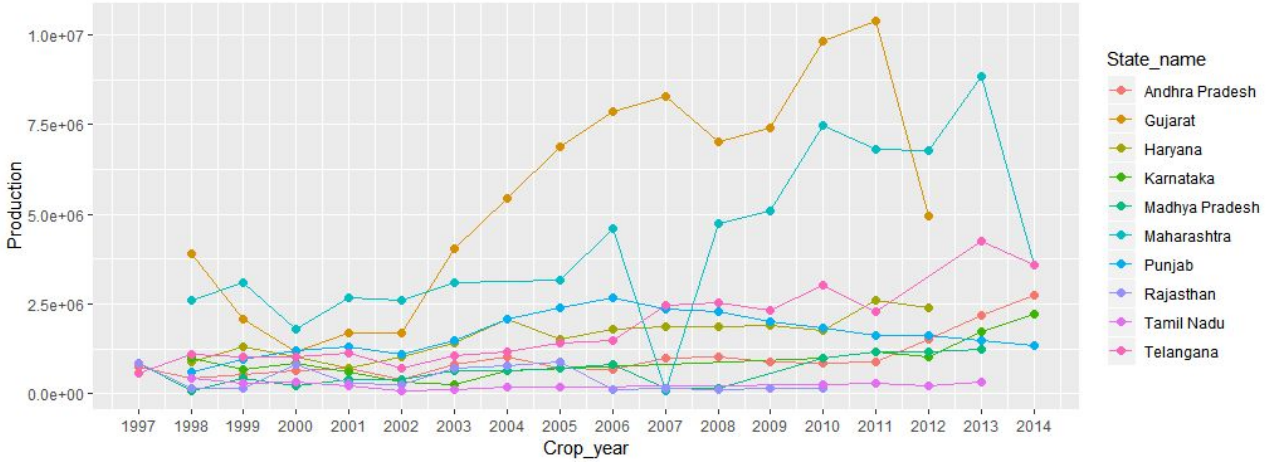
- Descriptive analysis
- Predictive analysis
- Crop Recommender

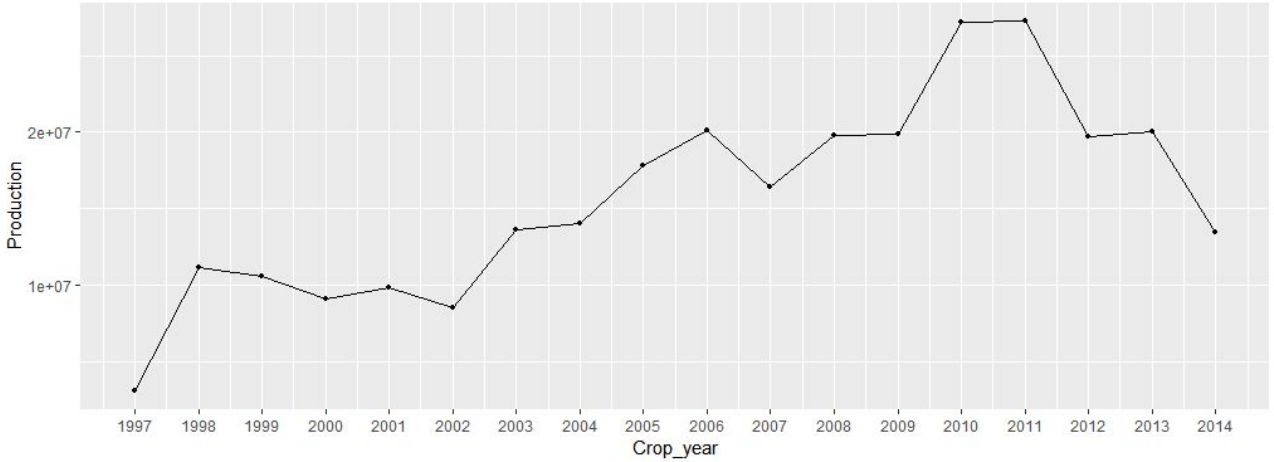
Descriptive Analysis

Descriptive analysis is used to describe the basic features of the data in a **study**.
Highly useful to get the trends and insights from the crop production and rainfall data to comprehend the needs of farmers and Industries.

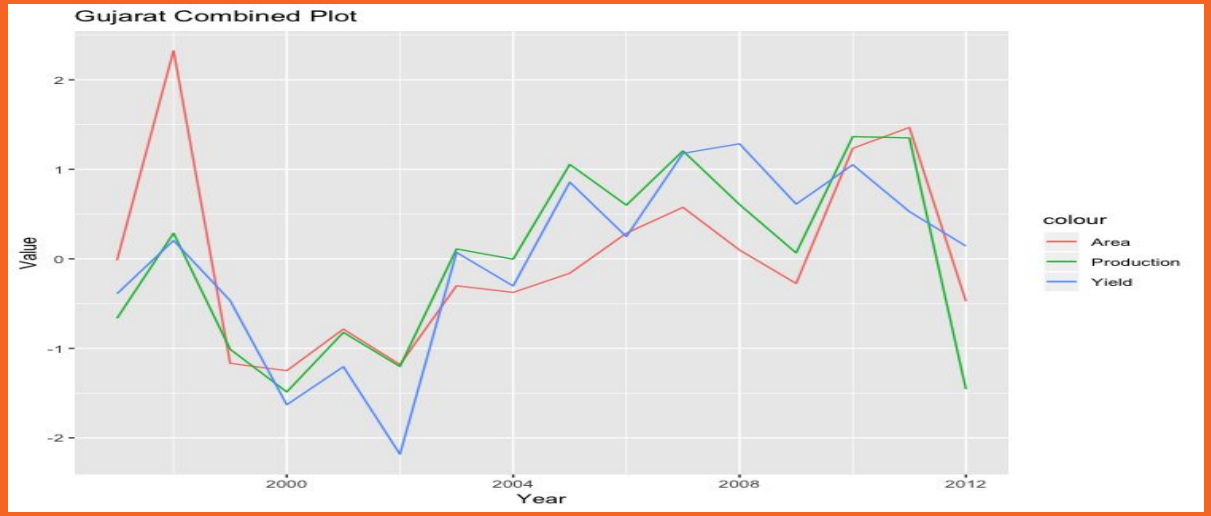


- ➔ The rainfall this century follows a 30 year cycle of high rainfall years followed by low rainfall spells. But the last decade completely reversed this trend.
- ➔ Rainfall was lower for the year 2012 with respect to other years which also resulted into failure of crops like sugarcane, cotton and rice.



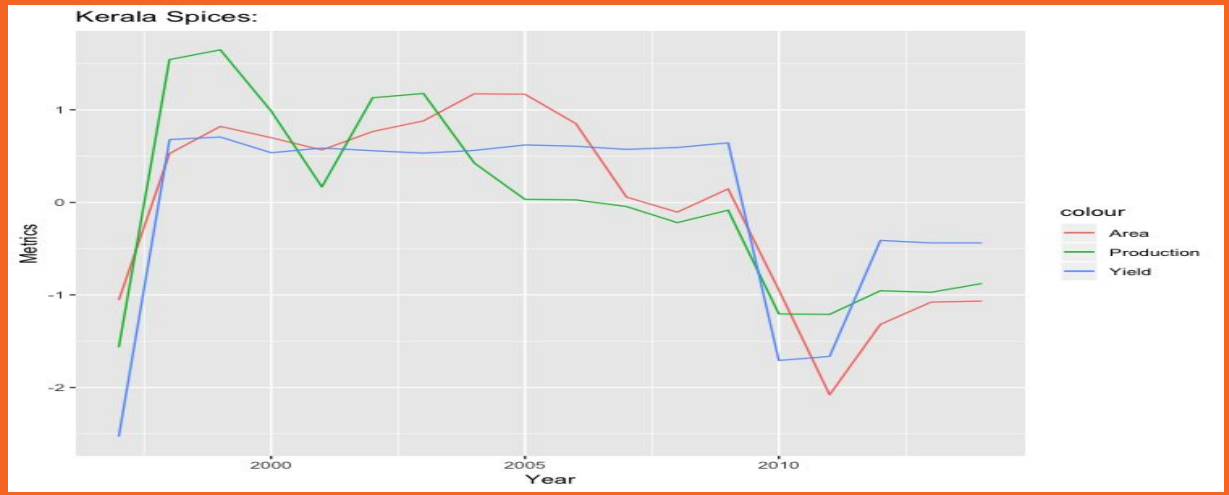


- Till the year 2009 India showed high cotton production growth. Largest contributors of this growth being Gujarat and Maharashtra. Though India become largest producer of organic cotton its production reduced after it's peak in 2009.
- Districts like Davangere, Raichur, Gulbarga in Karnataka showed decrease in cotton production compared to the previous years. The reason for this was, due to floods and heavy rain in north Karnataka districts which damaged the crop there.



➔ Gujarat recorded the highest decadal agricultural growth between 2000-01 to 2009-10.

➔ The area for spice cultivation has decreased in Kerala in 2008 whilst the production increased during 2011-2012.



Predictive analysis

Farming is a complex business with a lot of variables, particularly weather

So the starting point for any analytics solution is to understand the entire process and workflow through our descriptive analysis.

Used for agriculture, these methods allow for analysation of what has happened in the past on the farm, as well as what currently is happening and is going to happen, to make use of the data to predict the future and make decisions that impact the bottom line and end use of on-farm products

We are using XGBoost model to predict crop production.

We used XGBoost because tree boosting has empirically proven to be efficient for predictive mining for both classification and regression

As mentioned above we are using XGBoost as our predictor model. In order to take in account of rainfall variation, we are feeding the model both production and rainfall data. We are running the model for 80 rounds to get the least train-rmse.If we furthur increase the nrounds parameter,
the change is rmse is not that considerable. Our model predicts pretty well on the training set with accuracy mentioned

ME	0.01973366
RMSE	0.1087064
MAE	0.08393507
MPE	0.9123203
MAPE	3.70467

TABLE III
ACCURACY MEASURING METRICS

```
train-rmse:0.035644
train-rmse:0.030057
train-rmse:0.025423
train-rmse:0.021506
train-rmse:0.018218
train-rmse:0.015447
train-rmse:0.013085
train-rmse:0.011108
train-rmse:0.009418
train-rmse:0.007990
train-rmse:0.006787
train-rmse:0.005763
train-rmse:0.004903
train-rmse:0.004165
train-rmse:0.003545
train-rmse:0.003025
train-rmse:0.002587
train-rmse:0.002207
train-rmse:0.001894
train-rmse:0.001632
train-rmse:0.001412
train-rmse:0.001229
train-rmse:0.001078
train-rmse:0.000955
train-rmse:0.000854
train-rmse:0.000775
train-rmse:0.000707
train-rmse:0.000600
train-rmse:0.000557
train-rmse:0.000514
train-rmse:0.000512
train-rmse:0.000510
train-rmse:0.000510
train-rmse:0.000509
train-rmse:0.000509
train-rmse:0.000509
train-rmse:0.000509
```

Recommendation System

Recommendation System is a filtering system that seeks to predict the rating or preference of an item.

A better metric for a farmer or someone from the industry to look at a smaller subset of crops rather than a huge set of crops.

It gives them an idea as to what crops they can look at , and can prepare with a contingency plan.

An integral parameter for the recommendation system is the season-crop matrix with yield as the similarity metric.

Conclusions:

An amicable robust analysis system with a wide range of applications.

Interactive plots and figures for making decisions and taking close calls.

Crop production predictor that predicts crop yield for a region , with high accuracy.

Crop and Rainfall correlation for better understanding of trends.

Recommendation System to predict top 5 crops that can be grown in the region.

THANK YOU

Story for illustration purposes only