Analysing Atmospheric Parameters and it's impacts on Agricultural Production

Ayush Lodha[1] , Hardik Khichi[1], Kamlesh Sawadekar[1], Yogesh Dhawan[1], Udit Bhatia[2] (Instructor)

[1]Civil Engineering , IIT Gandhinagar, India

[2]Civil and Environmental Engineering, IIT Gandhinagar, India

Motivation

- India is primarily an agriculture based economy and the production is highly dependent upon atmospheric parameters.
- It is necessary to understand how these parameters affect the agriculture production in a telescopic view.
- Temperature being the primary parameter, is extensively used to understand the correlation between atmospheric parameters and agricultural production / yield.
- Other parameters such as precipitation, rainfall, cloud cover, etc. are derivatives of temperature originated changes.

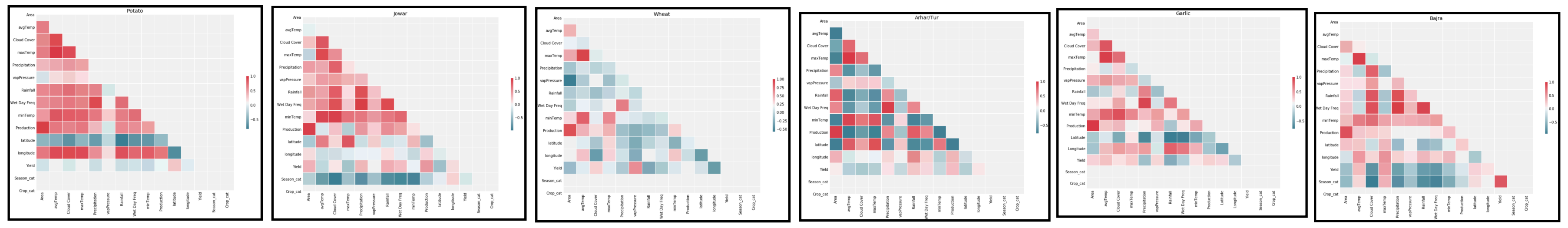
Data

- We have based our study on Gujarat region because of the easy availability of data and also the agriculture data for it is very well organised.
- **Atmospheric parameters** : Avg. Temperature, Rainfall, Precipitation, cloud cover, max. Temp, vapour pressure, wet day frequency, min. Temp, Crop_cat, Season_cat, longitude and latitude.
- **District regions** : Ahmedabad, Surat, Rajkot, Kutch, Kheda.
- **Crops** : Wheat, Jawar, Bajra, Garlic, Arahar and Garlic.
- **Season** : Kharif, Rabi, Summer, Whole Year.

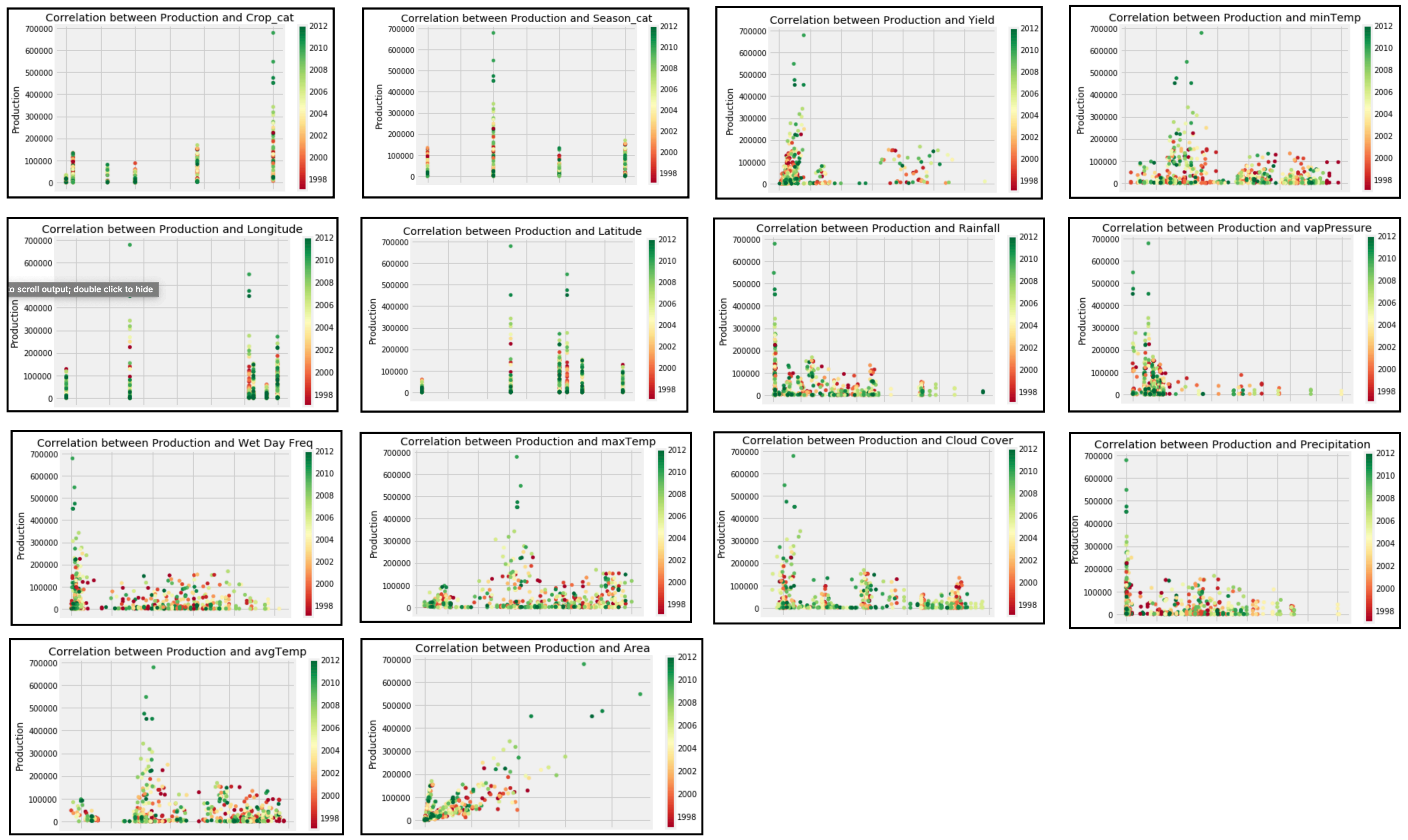
Methodology

- We collected the number of parameters that could affect the yield of a crop.
- We studied the heat map to detect any possible correlation between the parameters and yield.
- During EDA, we came to know there might be many possible outliers in the data. Therefore, to keep it simple we penalised some values of yield.
- Further we standardised the data and applied PCA over our parameters data-frame.
- Finally, we used Linear Regression, Lasso Regression, Support Vector Regression and Random Forest, on each and every crop time series per city, and filtered out the best fit and predictions.

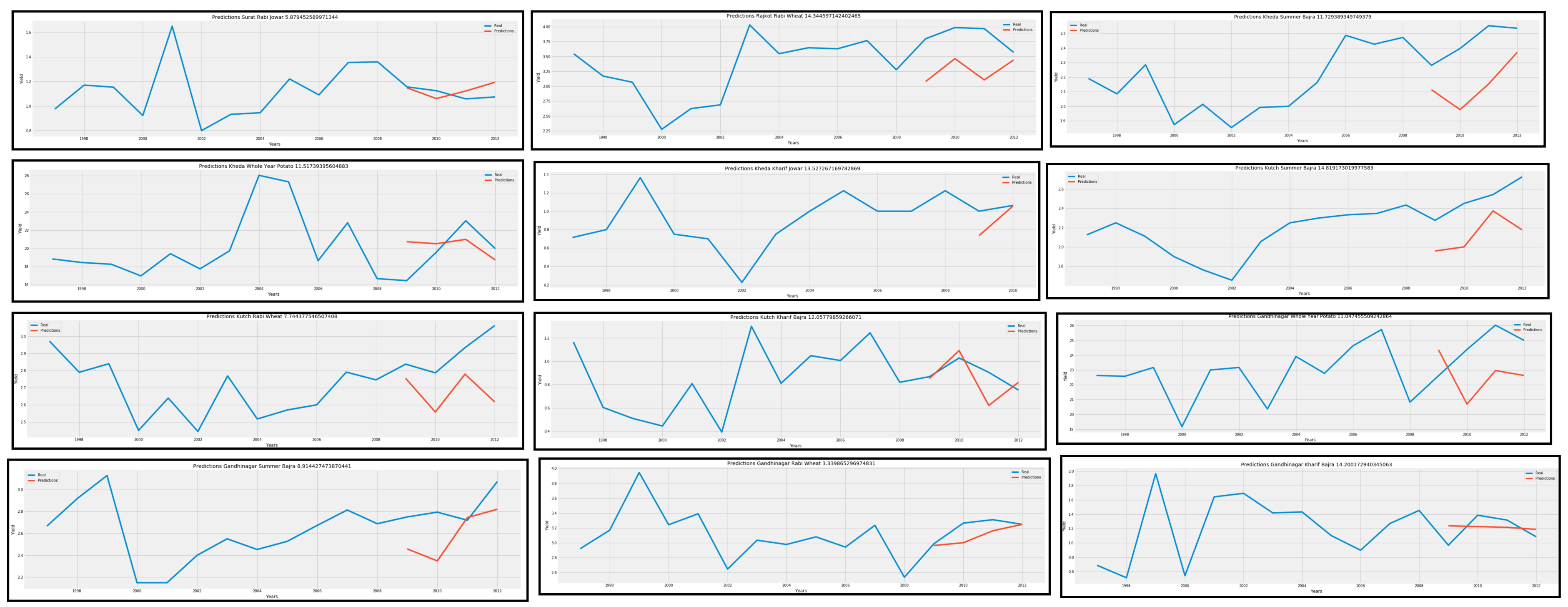
Results



Exploratory Data Analysis



The best fitted Predictions



Results(Continued)

Conclusions

- We have tried deriving the best features, removed features which have correlation with each other explicitly but we faced some problem in doing so because there were many combinations of cities, season and crops.
- Hence we applied PCA. By PCA we found the minimum dimension for our dataset and hence applied the further models.
- We reduced the complexity of the model to certain extent by the minimum dimension reduction which gives significant amount of variance (90%)
- We have tried out here various models such as multi linear regression, lasso regression, SVR and Decision Trees we have observed that each
- Regressor is performing well for some combination of city, crop and season than the others, which are listed above.

Limitations

- We did the cleaning of dataset by penalising the yield to be less than 5 because maximum density of yield is between 1 to 5.
- The average value of yield (5.38) is high compared to large number of yields because certain regions produce crops which require less amount of land. We can further dig out more ways of improvising the model via cleaning dataset.
- The dataset size is small, hence we cannot apply deep learning or neural network model because it might lead to overfitting.
- If in near future, we can get a larger dataset we can improvise model performance by using some deep learning techniques and hyper parameter tunings.

References

- https://www.sciencedirect.com/science/article/abs/pii/S0308521X01000580
- https://www.sciencedirect.com/science/article/abs/pii/S0308521X02000434
- SEASONAL ARIMA MODEL FOR FORECASTING OF MONTHLY RAINFALL AND TEMPERATURE by Inderjeet Kaushik and Sabita Madhavi Singh.