

Machine Learning Agriculture Yield Prediction

Submitted by Data Pirates:

Kashish Shah-AU1841014

Harsh Patel-AU1841015

Hriday Nagrani AU1841042

Yugamsinh Chavda AU1841090

Abstract—This document focuses on building a predictor which predicts the agriculture yield using the regression model. The agricultural yield of any crop depends on different Agro-Climatic factors like rainfall, temperature, soil type and the area and production of that crop. All these factors need to be analysed to predict the yield. As there are so many factors on which the yield depends, a polynomial regression model is needed instead of a linear model to find a better predictor.

Index Terms—Supervised Data, Crop Yield, Regression, Data Augmentation, Data Normalization.

I. INTRODUCTION / MOTIVATION, AND BACKGROUND

Agriculture sector is one of the most important sector in Indian economy. With the continuing expansion of the human population and urbanization, understanding agricultural yield is central to addressing food security challenges

Studying the impacts of climatic change on crop yield has become one of the most important topics to look forward as our surroundings are changing due to the impact of global warming.

The Agricultural yield mainly depends on weather conditions like temperature and rainfall, type and quantity of pesticides used, and the soil type.

II. LITERATURE SURVAY

We visited many official government websites like : 1) <https://iasri.icar.gov.in/>

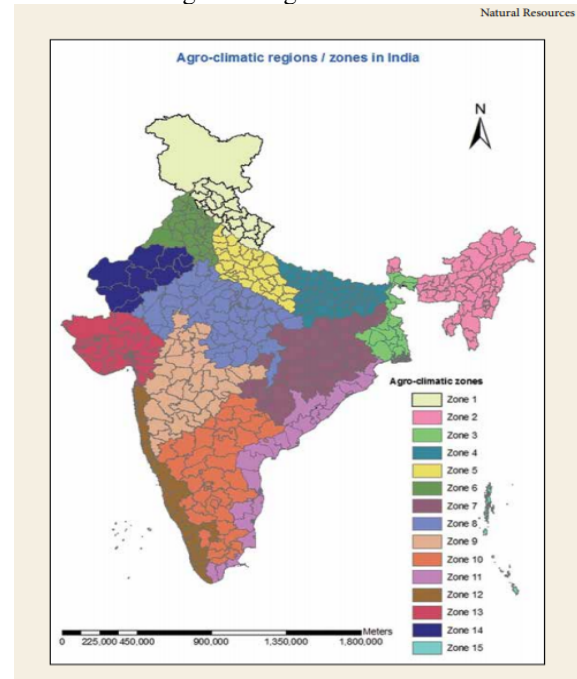
2) <https://agricoop.nic.in/en>

We explored through these websites and read about the agricultural scenario of India. We also went over the latest official Government agricultural report and also the estimate prepared by the government.

Here are the main learning outcomes :

- 1) Agro-climatic regions/zones in India : India is divided into 15 major agro climatic regions with multiple parts

of state covering each region.



- 2) Agro-ecological regions in India:

Arid
Semi Arid
Sub Humid
Humid-PerHumid
Coastal
Island

- 3) Broad soil groups:
- 4) Major Soil categories:
- 5) State and zone wise fertilizer and pesticides consumption:
- 6) Production yield and growth of major crops like rice and wheat.
- 7) Export and import trends and the position of India in world agricultural frameworks
- 8) Estimated and targeted yield sheet for 2020-21 by Ministry of Agriculture and Farmers Welfare

III. IMPLEMENTATION

A. Data Preprocessing

After collecting data for different factors such as rainfall, soil type, fertilizers, there were many outliers in the data, missing values, Variable transformation and many other issues.

a) We removed the outliers by replacing the values with the mean value. Missing values are replaced with mean value.

In soil data we have different components of soil in a single cell. We divided the soil data individually and then created new columns and filled the values.

We transformed multiple columns to a single column by taking average of them or just by adding them. For example, in the fertilizer data, we had two columns for Nitrogen fertilizer(variable), one for kharif season and other for rabi season but as we are considering annual data, we added them into a single column for Nitrogen fertilizer. :

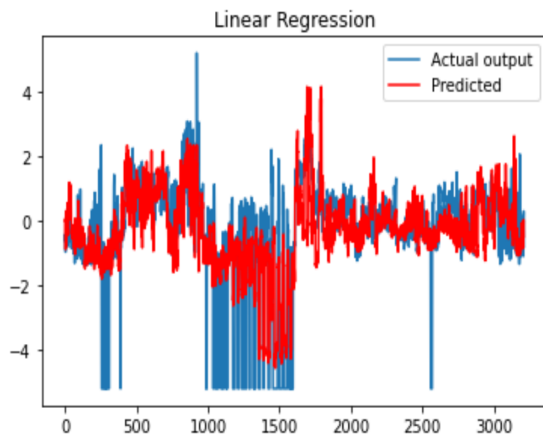
B. Implementing Regression

After the final data was prepared we used different regression techniques such as Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression and observe the results.

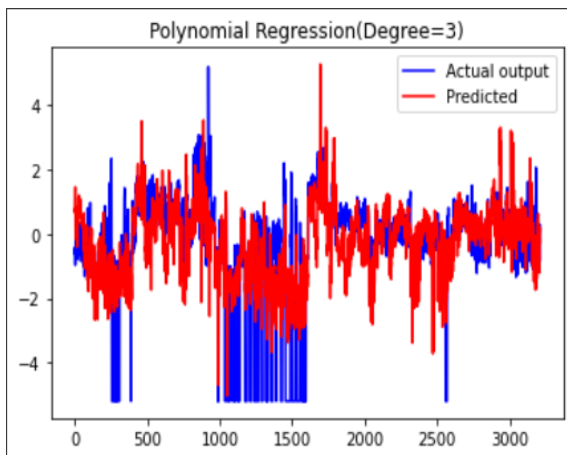
Results are shown in the next section.

IV. RESULTS

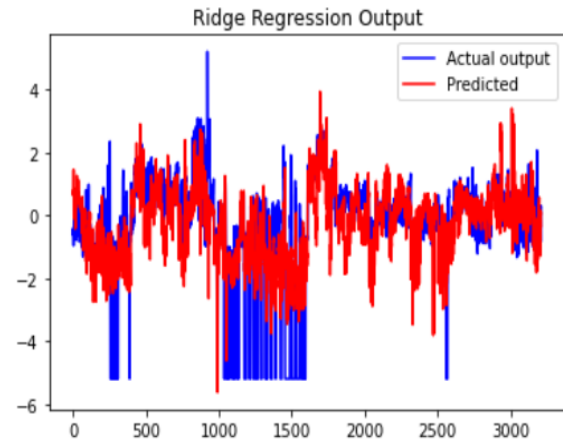
A. Linear regression



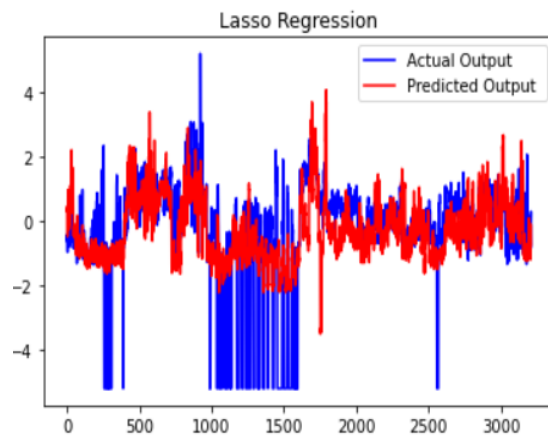
B. Polynomial regression with degree=3



C. Ridge Regression Output



D. Lasso Regression Output



REFERENCES

- [1] <https://iasri.icar.gov.in/>
- [2] <http://data.icrisat.org/dld/src/crops.html>
- [3] <https://data.gov.in/>
- [4] https://scikit-learn.org/stable/modules/linear_model.html

V. CONCLUSION

After applying four different regression methodologies which are: Linear Regression, Polynomial regression, Lasso regression and Ridge regression, we found out that the Lasso and Ridge regression techniques gave a better accuracy than a simple Linear regressor.