Effect of water quality and rainfall on agricultural production

Shashank Sistla B.E. in Chemical Engineering(Hons) with Minor in Data Science

Rushabh Musthlaya B.E in Computer Science and Engineering

Vivek Varma B.E. in Computer Science and Engineering

Harshit Gakhar B.E. in Computer Science and Engineering

*Abstract*— India is primarily an agrarian country. This paper presents an analysis of the relationship between crop yield, ground-water quality and rainfall using various techniques.

Keywords— rainfall, crop yield, water quality

# Motivation

Agriculture is a vital part of the Indian economy, and is in fact the most common industry. Crop yields depend mainly upon the rainfall in the region as well as the quality of the water. The aim of the research project is to analyse data and establish an inter-relation betweenwater quality, rainfall and crop yield. This report deals with the pre-processing of aforementioned data.

# Objectives

The objectives of the project are as follows:

1. To pre-process rainfall data across various states of the nation from years (1901 – 2015) by filling missing values in the data and subset selection.
2. To pre-process ground water quality data across various water sources( such as bores, lakes, etc.) in the nation. (2014)
3. To pre-process crop yield data across various districts from years (1997-2014) by binning and filling missing values in the data.

And finally, to relate the crop production in the datasets in the year 2014 in specific. This way, it will be possible to decide the ideal parameters for agriculture for various crops.

# Background

There is good reason to expect a relationship between both the rainfall and ground water quality with the production of crops since water is an integral part of agriculture as a whole.

# Methodology

Rainfall data:

Firstly, columns which were superfluous to the data were dropped. These columns included simple addition of the other columns. Secondly, the initial dataset had multiple missing values which had to be filled before further pre-processing. In order to fill these missing values, we interpolated neighboring values of the dataset since this made the most sense in this type of data. In this dataset attribute-subset selection and filling up of missing values was implemented.

Ground water quality:

This dataset has several columns which have been condensed using PCA (Principal Component Analysis). However, before being able to use PCA, the missing values had to be dealt with. Ideally, stochastic regression imputation would work best for this dataset. As of this report, we have implemented a precursor to this, which is random imputation. Further, PCA was implemented on the datasets of all states. The number of principal component was 5-7 depending on each dataset. In this dataset, PCA and random imputation was implemented.

Crop yield data:

This dataset has large numerosity with unnecessary segregation of the data. This data was made more useful and computable by binning the yields of the crop, both district-wise and season-wise, in bins of size 3 years. In this dataset, binning was implemented and the missing values were filled with average produce per unit area for that particular crop.

# Results and Discussion

This project will be very helpful in ascertaining relations and studying and analysing it. It will be very useful to analyse crop yield and might help in boosting the agriculture industry.

# Bibliography

1. Water Quality in India 2014 (<http://tiny.cc/np2bjz>)
2. District-wise, season-wise crop production statistics (<http://tiny.cc/el2bjz>)
3. Subdivision wise Rainfall and its departure from 1901 to 2015 (<http://tiny.cc/g43bjz>)

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**