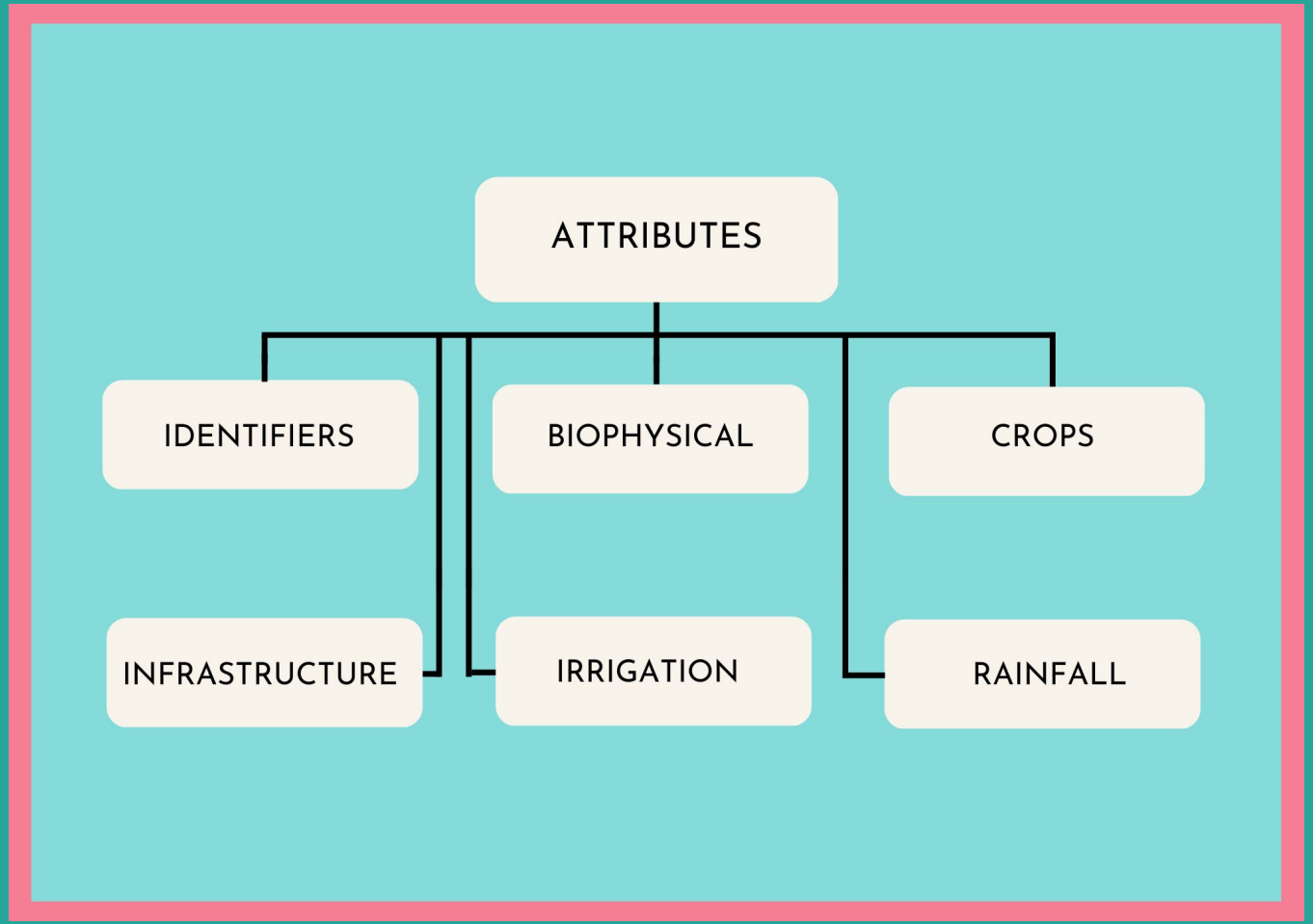
Analysis of Total Geographical Land Use and Prediction of Crops using Various ML Models

Review - 1

20BCE1043 (Vishal N)
20BCE1360 (Prathiba Narayan)
20BCE1317 (Jyothssena GS)

Datasets are a combination of

- Crops (http://data.icrisat.org/dld/src/crops.html)
- Biophysical (http://data.icrisat.org/dld/src/biophysical.html)
- Irrigation - http://data.icrisat.org/dld/src/irrigation.html
- Soil - https://data.icrisat.org/dld/src/soil.html
- Environment - http://data.icrisat.org/dld/src/environment.html (Optional because of data redundancy)
- Infrastructure (http://data.icrisat.org/dld/src/infrastructure.html) (Only bank column)
- Crop based on environment factors (https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset)


ATTRIBUTES
IDENTIFIERS
BIOPHYSICAL
CROPS
INFRASTRUCTURE
IRRIGATION
RAINFALL

Identifiers

- Year
- State
- District

Biophysical

- Total Area
- Forest Area
- Barren And Uncultivable Land Area
- Land Put To Non-Agricultural Use Area
- Cultivable Waste Land Area
- Permanent Pastures Area
- Other Fallow Area
- Current Fallow Area
- Net Cropped Area
- Gross Cropped Area
- Cropping Intensity

Infrastructure

- No. Of Banks
- No. Of Post Offices

Crop based on factors

- Nitrogen
- Phosphorus
- Potassium
- Temperature
- Humidity
- pH
- Label

Crops

- Area
- Production
- Yield

The 3 Attributes Are Present For The Following Crops

- Rice
- Wheat
- Kharif
- Sorghum
- Maize
- Chickpea
- Pigeonpea
- Minor Pulses
- Sesamum
- Rapeseed
- Oilseed
- Sugarcane
- Fruits
- Vegetables
- Fodder Area

Rainfall

- Jan To Dec Rainfall (In Mm)

Outcomes

The project can help in identifying crop patterns, predicting crop yields, evaluating model performance, understanding factors influencing crop growth, creating interactive visualisations, and providing decision support for farmers, policymakers, and other stakeholders involved in crop management and planning. These outcomes can lead to improved crop yields, optimised resource utilisation, and more sustainable agricultural practices.

ML Models and Predictions

- Predicting cultivable land for future years using regression techniques (linear).
- Change in cultivable land in 2018/2017 in all districts, and comparing with actual data to show accuracy
- Finding correlation between various factors(rainfall, irrigation and area) affecting crop yield
- Multiple regression for crop yield in future years by predicting the individual factors

Procedure

1. We assess the quality of the data before we work on it, and decide whether it meets the standards needed for running an ML model on it using Python, as well as visualising it using R/Python.

2. Pre-process the data and remove the unnecessary columns.

3. In the other columns, impute the data using various imputation methods, and then choose the most suitable option.

4. We can also remove the outliers, to prevent it from affecting the predictions by a large margin.

Hypothesis (Expected Output)

Possible Hypothesis and Visualizations

1. There is a positive correlation between the cropping intensity and yield for certain crops.
2. There is a positive correlation between the rainfall and the production of certain crops.
3. Time-series visualisations showing the trends between biophysical attributes.
4. There is a positive correlation between the area under cultivation and the production of certain crops.
5. There is a positive correlation between the area under cultivation and the yield of certain crops.
6. Certain states or districts have higher production and yields for certain crops compared to others.
7. The data provides insight into which crops are more suitable for cultivation in certain regions.
8. The data provides insight into how to increase the production and yield of certain crops.
9. The crop grown in a region depends on the Nitrogen, Phosphorus, Potassium contents of the soil, its pH, the temperature, humidity and rainfall

Thank You