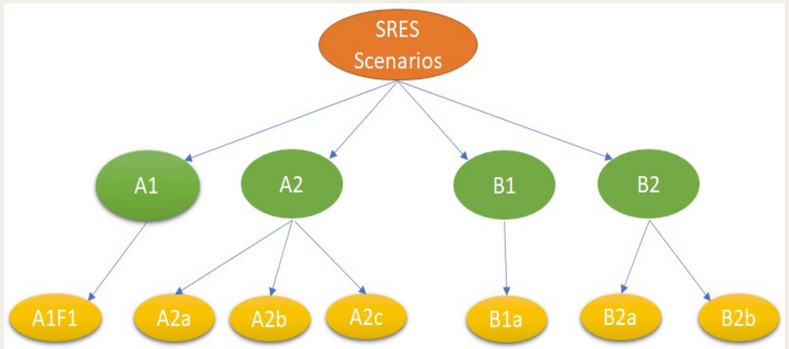
# Analyzing effects of climate change in the global food production using k-means clustering and DBSCAN

J. Kovoor and S. Neupane

## Motivation

Agricultural sector will be facing alarming challenges originating from climate change in coming decades.
Currently, the global food production able to cope with the world population growth but in this century the climate change could affect the global food production and availability in many parts of the world, especially those area which are prone to drought and famine. In this project, we are studying the potential impact of climate change on the global food production using the clustering methods such as K-means clustering and DBSCAN.

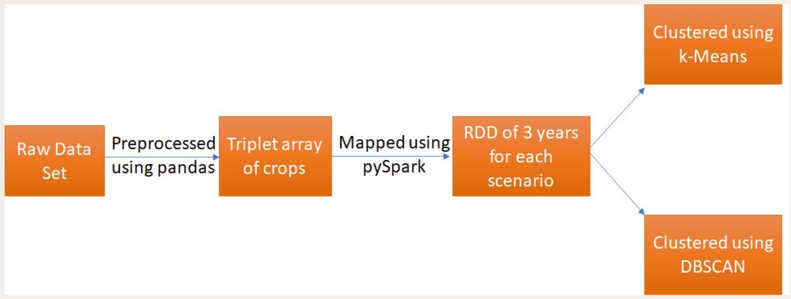


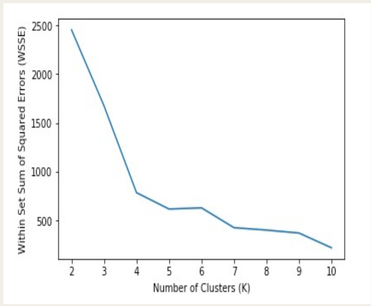

## Data Set

This data set provides an assessment of the potential effect of climate change on world major crop production; wheat, maize and rice. It primarily focuses on quantitative estimates of global food yield changes based on different climate scenarios.
It is published by NASA Socioeconomic Data and Applications Center (SEDAC) and it has food yield information for 166 countries modeled under 7 different scenarios for years 2020, 2050 and 2080 based on the baseline data from 1990 production.
It basically provides two types of yield information: total production change and percentage change in yield. For this project, we are considering only the second one.

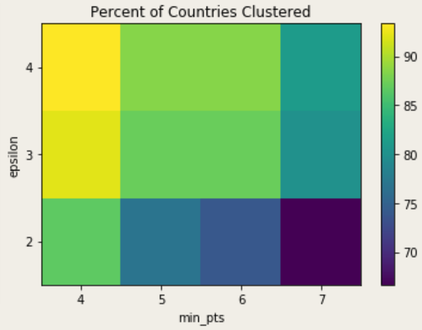
## Methodology

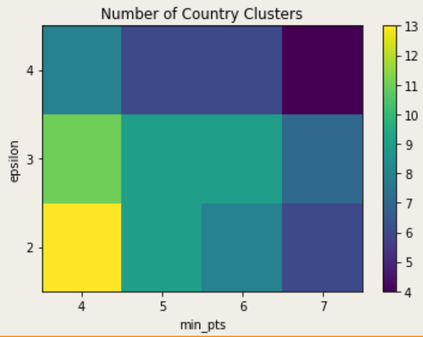
- Data was preprocessed using pandas. Missing values were filled with zeros.
- A dataset of the three crops were selected for a particular scenario and year.
- Clustered countries using two clustering methods: K-means clustering and DBSCAN.
- Optimal number of clusters determined using elbow-method for K means clustering and using heatmap epsilon and minimum
points were determined for DBSCAN.



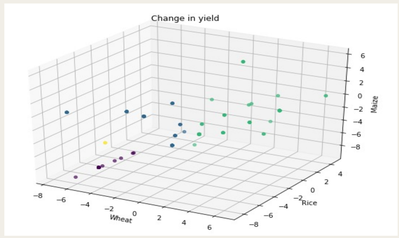



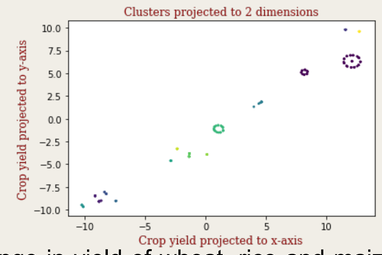

Finding the optimal number of clusters using elbow method. In the plot above the optimal number of clusters found was 4. Plot on the right shows the heatmap to determine the optimal value of epsilon and minimum point required for DBSCAN. For this particular case, epsilon and minimum points found were 2 and 4.



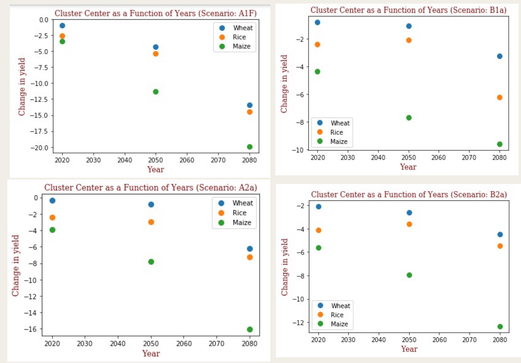



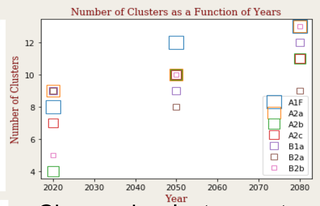

## Results and Discussion




Clustering the countries based on change in yield of wheat, rice and maize for year 2020 under the scenario A1F using K-means clustering (left) and using DBSCAN (right).




Change in cluster centers for four different scenarios using k-means (left four) and change in cluster numbers for all seven scenarios using DBSCAN (top).

## Summary

- We clustered the countries based on the change in yield of three major crops using K-means as well as DBSCAN.
- We see the decreasing pattern of yield over the years for all three major crops under all scenarios.
- The number of clusters are increasing over the years under all scenarios indicates the climate change climate affecting some countries most than others.

## References


- Iglesias, A., and C. Rosensweig. 2009. Effects of Climate Change on Global Food Production from SRES Emissions and Socioeconomic Scenarios. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- Parrya, M.L.,C. Rosenzweig, A. Iglesias, M. Livermored, and G. Fischer. 2004. Effects of Climate Change on Global Food Production Under SRES Emissions and Socio-economic Scenarios. Global Environmental Change 14 (1): 53-67.


THE UNIVERSITY OF TENNESSEE KNOXVILLE