Arable Land clustering and prediction

Anzar Naseer, 21087748
Github Link: https://github.com/anzarnaseer/Applied-Data-Science-As-3.git

ABSTRACT

The World Bank's collected data on arable land will be analyzed as part of this study, and clustering and prediction algorithms will be used to acquire insight into altering patterns in the utilization of arable land. The initial phase of the project consists of obtaining data from the World Bank and processing it to extract key characteristics. The KMeans clustering algorithm is then used to identify trends in the data pertaining to arable land over an extensive time span. In order to present the results, scatter plots and three-dimensional graphs are employed. In the final stage, a Linear function is applied to the country's arable land data. This offers us the opportunity to predict future trends in the utilization of arable land. This study aims to highlight how techniques such as data analysis and machine learning may be utilized to get insight into major environmental indicators, such as arable land utilization.

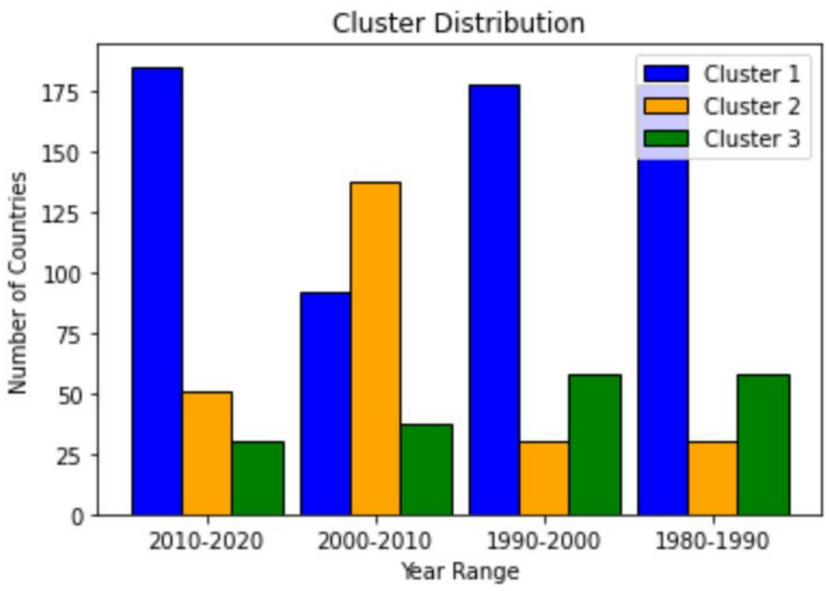
Introduction

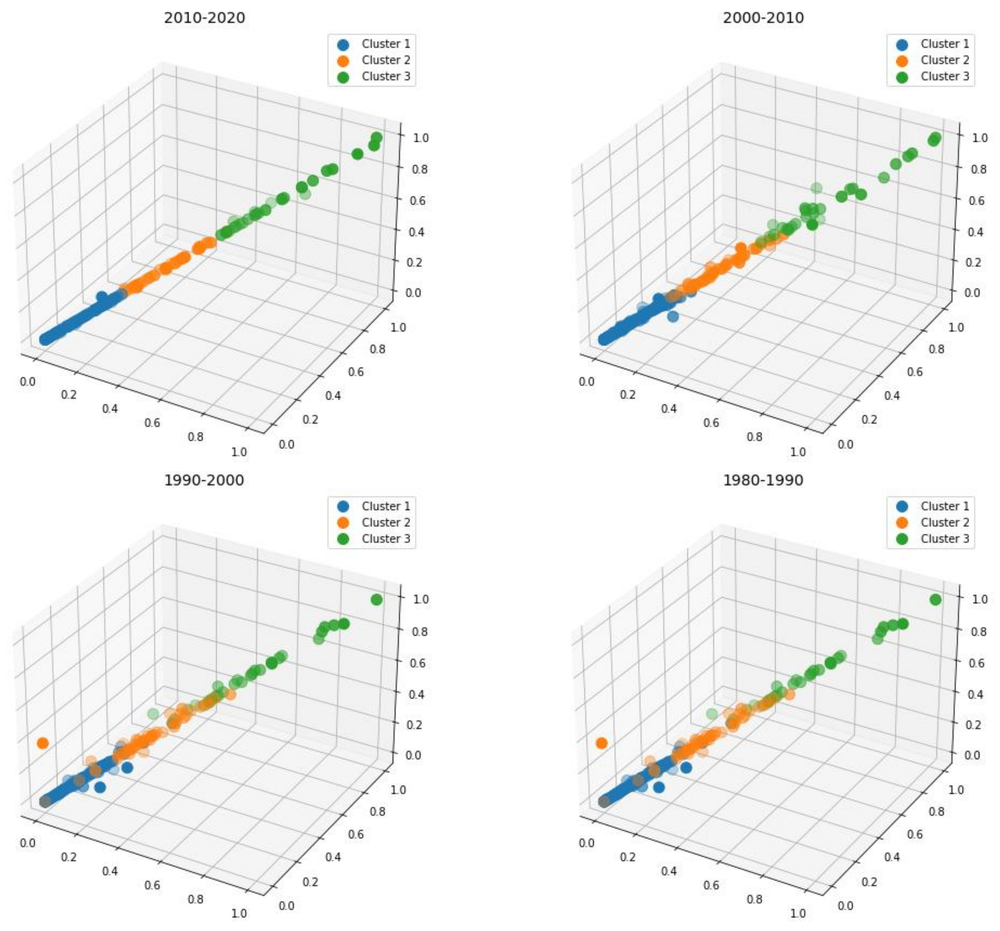
The most significant resource for agriculture and food production is arable land. It refers to land that is appropriate for cultivating crops and is crucial to the global food supply. However, because to urbanization, industrialization, and climate change, arable land is scarce and is under increasing pressure. This project seeks to employ clustering techniques to categories nations based on their arable land patterns over time and to construct predictive models for future arable land changes in order to gain a better understanding of the global state of arable land. in order to gain a better understanding of the global state of arable land. The research will utilize World Bank data on arable land area as a percentage of total land area from 1960 to 2019 and employ unsupervised clustering techniques to uncover patterns and commonalities in the trends of arable land across nations. Based on past data, it will also employ supervised learning algorithms to anticipate future changes in arable land. This study will yield vital insights into the existing condition of arable land and potential future threats to food security.

Clustering

To cluster and predict arable land across different time intervals (2010-2020, 2000-2010, 1990-2000, and 1980-1990), we divided the data of each time interval into 3 distinct clusters, as depicted in bar graph. Over time, the number of countries in each cluster varies. Cluster 1 contains the most countries throughout all time spans, while Clusters 2 and 3 have fewer nations. The number of countries in each cluster varies according on the interval of time.

A 3D scatter plot is a method of displaying the relationship between three variables. In the context of clustering and predicting arable land using data from the World Bank, the 3D scatter plot can be used to depict each cluster for each time interval. By assigning distinct colors or symbols to each cluster, the scatter plot may demonstrate how the clusters evolve over time and how they differ from one another. The 3D scatter plot can provide a clear and straightforward display of clustering results and aid in the identification of patterns and trends in the data.

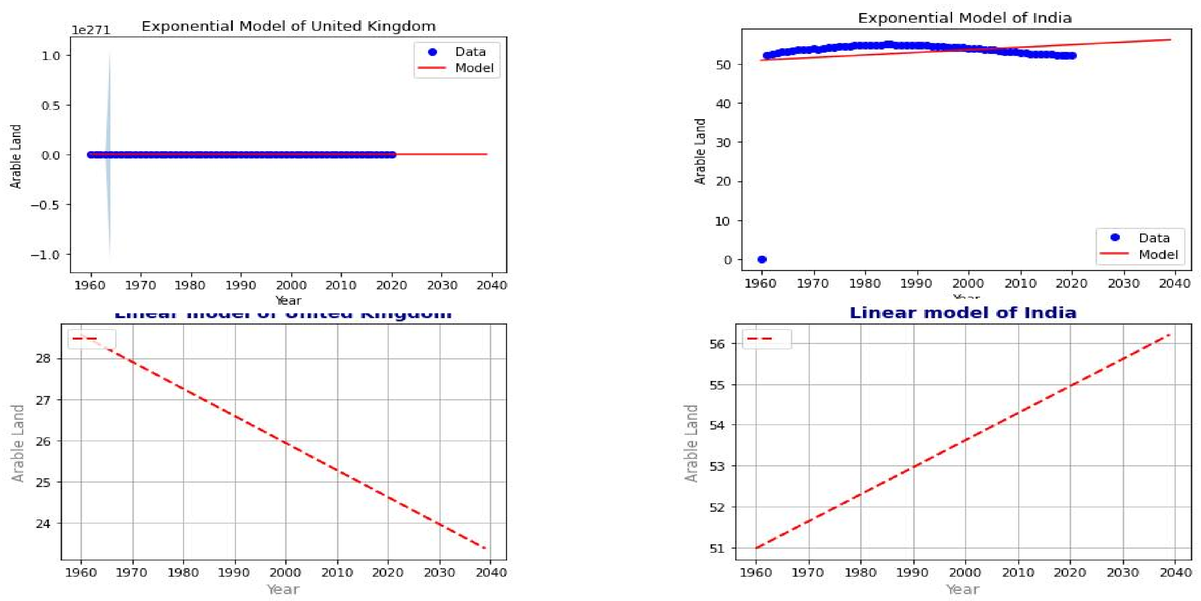




Prediction

The figures illustrate the expected values of India's and the United Kingdom's arable land in 2025, 2030, 2035, and 2039. The value of India's arable land is expected to increase from 55.3% in 2025 to 56.2% in 2039. The percentage of arable land in the United Kingdom is projected to fall from 24.3 percent in 2025 to 23.4 percent in 2035, with a modest uptick to 23.6 percent in 2039. Using a curve-fitting model that takes into consideration the historical data of arable land for each country and predicts future values based on the trends identified in the data, these results were achieved.

Country	2025	2030	2035	2039
India	55.27649503	55.60766903	55.93884303	56.20378223
UK	24.30098887	23.97357454	23.64616022	23.38422876



Conclusion

Due to urbanization, industrialization, and climate change, arable land, which is vital to global food security, is becoming increasingly rare. This research groups nations based on their patterns of arable land over time and creates forecast models for future changes in arable land using World Bank data from 1960 to 2019. A scatter plot in three dimensions exhibits each cluster for each time period and illustrates how they evolve and diverge from one another. In the coming years, the results indicate a minor rise in India's arable land and a decrease in the United Kingdom's arable land. This data can assist policymakers and other stakeholders in ensuring the availability of arable land for food production and global food security.