Vaibhav Jha

DS 4002

28 April 2025

Clustering Countries by Constitution: A Vernacular Study

As human society transitioned from agrarian to industrial means of production, the establishment of a new social order became necessary. Nation-states sought to formally declare their political systems to the world to gain legitimacy. This gave rise to the Constitution in the post-Modern era shortly after the revolutions of the late-18th century.

Constitutions capture the essence of national identity. They underscore economic, religious, and political systems of power. They reflect the will of the people.

With the prominence of the nation-state rising in contemporary society, it is important to understand how nation's compare in constitution. This allows for a greater understanding of their international relations, providing insight into their position in the world.

This project conducts a constitution similarity analysis. Studying the vernacular of 202 countries' constitutions, the project uses a hierarchical clustering data science technique to group countries by constitution similarity. The end result is two-fold: 13 distinct clusters, each with word clouds displaying the frequency of words used.

Deliverable

Professors in the Department of Sociology at the University of Virginia seek to create a course series on constitutions and how effectively they are followed in contemporary society. The professors do not know how to group countries by course, lacking understanding of the subtopics appropriate to divide and conquer such a vast study. They contacted the UVA School of Data Science for input on how to deploy this course series.

Following data acquisition from The Constitute Project, data cleaning techniques should be used to remove stop-gap words and prepare for analysis. A brief exploratory data analysis should be conducted to allow researchers to understand what words are most prevalent in the data set. The analysis will begin with feature extraction, giving the preambles numerical value in the form of a cosine similarity matrix. A hierarchical clustering algorithm will follow, grouping the constitutions by similarity.

GitHub link: https://github.com/vaibhavjha06/cs3-casestudy