An examination of the role of social characteristics in the prevalence of COVID-19 in Ireland using machine learning techniques.

Introduction

COVID-19 has had a huge impact globally and Ireland, like most other nations has been hugely impacted having one of the worlds strictest lockdowns and restrictions. We are currently in the process of a vaccine rollout with the aim to reopen the country slowly and carefully. With this research I hope to understand the impact of the current COVID-19 pandemic throughout Ireland given associated characteristics and study the related patterns in different clusters obtained from demographic information (Census Data).

Research Questions

What meaningful clusters can be identified using machine learning techniques grouping spatial data from electoral divisions in Ireland?

What demographic information can be identified from these various clusters?

Can we define an area likely to have high risk of transmission of COVID-19 or other epidemiological viruses in the future using this spatial and socio-demographic data extracted from census data?

Significance

The information obtained from this research could perhaps help in the management of restrictions and help control the spread of this virus or further epidemiological viruses in the future by identifying clusters of locations linked to transmission of the disease. As a result of this research, it could also be possible to focus public health campaigns at the demographics identified.

Data

Spatial Data - points representing the location of electoral division areas throughout Ireland of which there are 3444. These represent the smallest legally defined administrative areas in the state.

Socio-demographic data – obtained from the census data 2016 consists of over 750 features for e.g. (age, martial status, education level, house share, social class)

COVID-19 data – reported at an electoral division level and using confirmed cases of covid-19 for each electoral division.

Method

Use of PCA **Principal Component Analysis** for feature extraction. As the census data has over 750 features dimensionality reduction removing redundant variables is going to be an important step.

Extracted features will be fed into a clustering model and different clusters will be identified with the aim of clustering various electoral divisions with various demographic characteristics from the census data so any similarities between them can be maximised.
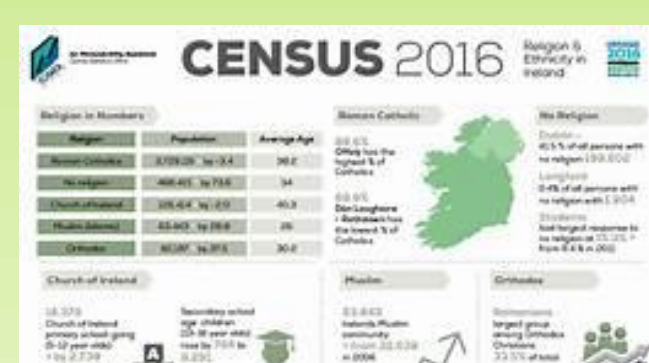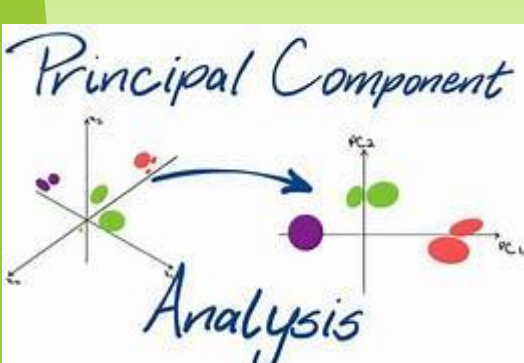
A machine learning based **spatial clustering model** will analysis links between the number of confirmed covid cases and the various socio-demographic determinants.

Implement **unsupervised learning** approach based on **artificial neural networks** that can transform georeferenced data into information.

Using **Topology Data Analysis** along with the machine learning model I hope to get better results than are possible by either technique individually.

Technologies

Artificial neural networks
Principal Component Analysis
Kernel PCA
Spatial Clustering
Topology Data Analysis (TDA)
TensorFlow
Sklearn

References
https://www.cso.ie/en/census/census2016reports/census2016boundaryfiles/
https://maproom.net/shop/political-map-of-ireland/
https://www.cso.ie/en/media/csoie/methods/censusofpopulation/Census_2016_Quality_Report_rev_0918.pdf
https://data.cso.ie/

Author: Elaine Koyce
Email: elainekoyce@gmail.com

Supervisor: Dr. Greg Doyle