An Analysis of the rise of the Far-Right in Europe: A Data Mining Project

Beltrán Castro Gómez

April 3, 2022

---

1 Introduction

Europe has seen a boom of Far-Right and Populist political parties and personalities in the recent years. Some examples are Matteo Salvini in Italy, Marine Le Pen in France, Fidesz and Viktor Orban in Hungary, The Freedom Party in Austria or Vox in Spain [1].

Most of the experts highlight two main causes: people's discontent over the economic recession derivated from the 2007–2008 financial crisis [2] and the success of populist speeches with regard to the 2015 refugate crisis [3] and the increasing globalization of the european society in general.

In this project, I will try to use Machine Learning and Data Mining to study this problem and its relation with the said causes.

2 Problem Understanding

Nowadays, Data Science can be a double-edged sword in politics. At the same time it can be used for understanding problems and its implications, it can also lead to populist speeches based on sentiment analysis or even massive political advertising like we have seen not long ago in the Facebook–Cambridge Analytica data scandal [4].

This project does not only aim to show that Data Science can be used for the good when applied to politics and social studies but also follows a greater reason: studying our history and trying to understand it is the best way to avoid past mistakes.

3 The dataset

3.1 Data Understanding

The dataset is composed of four kinds of data (electoral data, political parties data, economic data and demographic data) from 28 countries: Austria, Belgium, Bulgaria, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, Finland, France, Greece, Hungary, Ireland, Iceland, Italy, Luxembourg, Latvia, Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovakia, Turkey, Ukraine and the United Kingdom.

1. **Electoral data**: It consists in the electoral results to the respective parliamentary elections of all of the above countries in the XXI century. Due to the fact that an already existing dataset was not found, the electoral data was collected from the Election Resources on the Internet webpage by Manuel Álvarez-Rivera [5]. It contains both regional and nationwide results.

2. **Political parties data**: It contains basic information about Political Position (i.e. Left-wing or Right-wing) and Ideology (i.e. Socialdemocracy or Liberalism) of all of the political parties present in the Electoral data. Since a political party dataset could not be found in the internet, this dataset was created by adding the information manually from each Party's Wikipedia entry.

3. **Economic data**: This part is composed of the evolution of the Unemployment Rate in different European countries from 1992 until 2020 provided by the Eurostat (European Statistical Office) database [6].

4. **Demographic data**: This section involves the Total Population by country and the Total Number of Immigrants from 1990 to 2020 and it has also been obtained from the Eurostat database.

All of the data used for this project is available in farRightEurope-analysis.

3.2 Data Preparation

This was the part of the project that took the longest. Specially for the electoral data, since it was not coming from a real dataset and a lot of adjustments had to be done. These are some of the issues that had to be addressed:

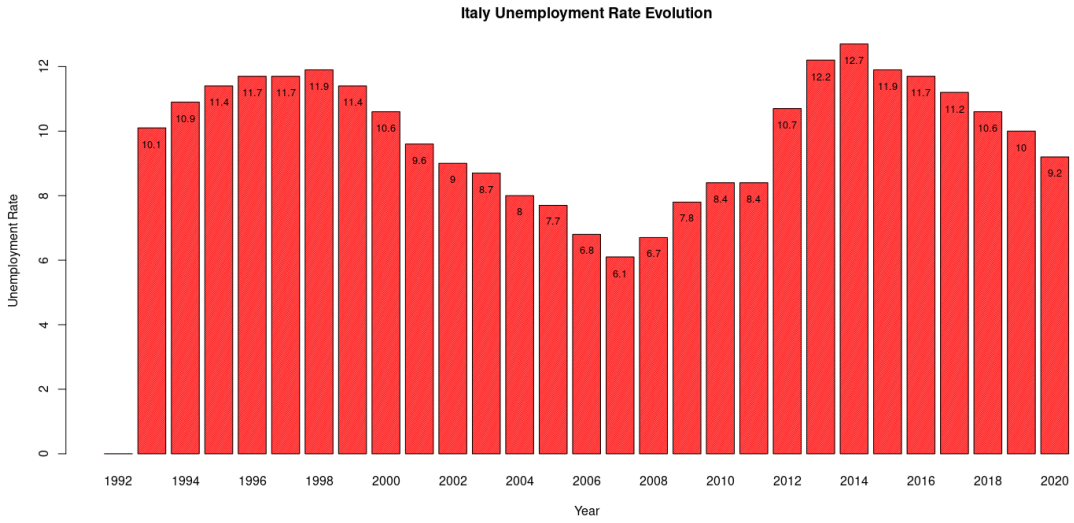

Figure 1: Example of Economic data from Italy.

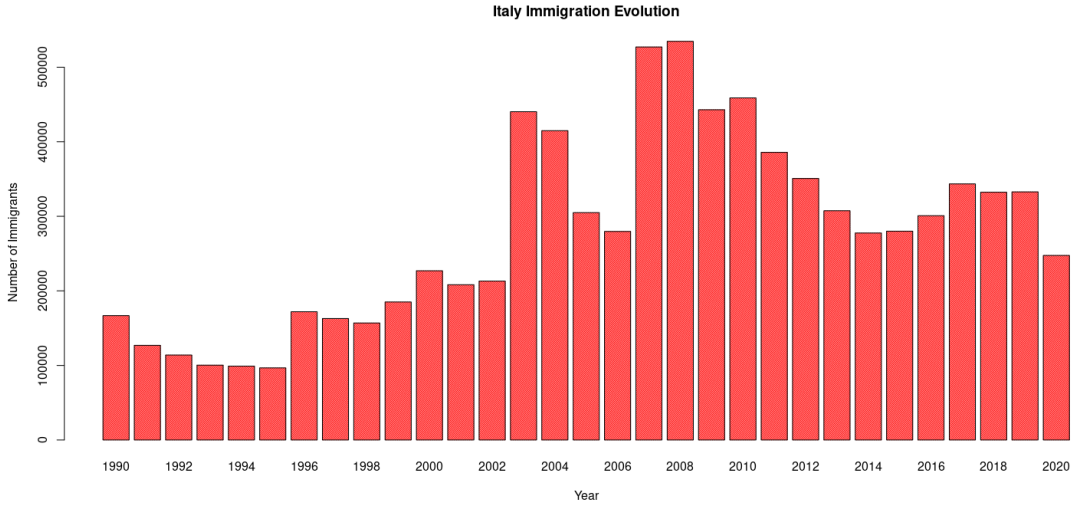

Figure 2: Example of Demographic data from Italy.

- Most of the data was available in CSV format, but for some of the countries it was not the case and the data had to be downloaded by performing web scrapping via Python directly from the webpage.
- Some corrections have to be done in the text (i.e. the name of the political parties). Because the name of the political parties were written in different languages, some special characters were not being encoded correctly in the webpage.
- The fact that every country has its own electoral system and this can even change through the years put more constraints into the format of the electoral data and made it necessary to almost process every country's data differently.
- For this project, only national results were taken into account and so regional ones were avoided.

With respect to the data taken from Eurostat, some scaling needed to be done in the Immigration data, as it was expressed in total numbers. Thus, this was scaled by dividing it by the total population of the countries to obtain the desired proportion.

3.3 Data Exploration

After collecting and processing all of the data from the different sources, this was merged into a dataframe with the following structure:

- **Region**: The abbreviation for the country.
- **Year**.
- **Unemployment**: The unemployment rate registered for that country in that year.
- **Immigration**: The proportion of immigrants in the total population of the country that year.
- **FarRightSeats**: The percentage of seats achieved by all Far-right parties in that country and in that year.

Finally, after removing NaN records, a very small dataset was achieved considering two key factors: electoral processes take part each 4 or 5 years and not every year and that the quantity of records about seats achieved by these kind of parties may not be representative enough.

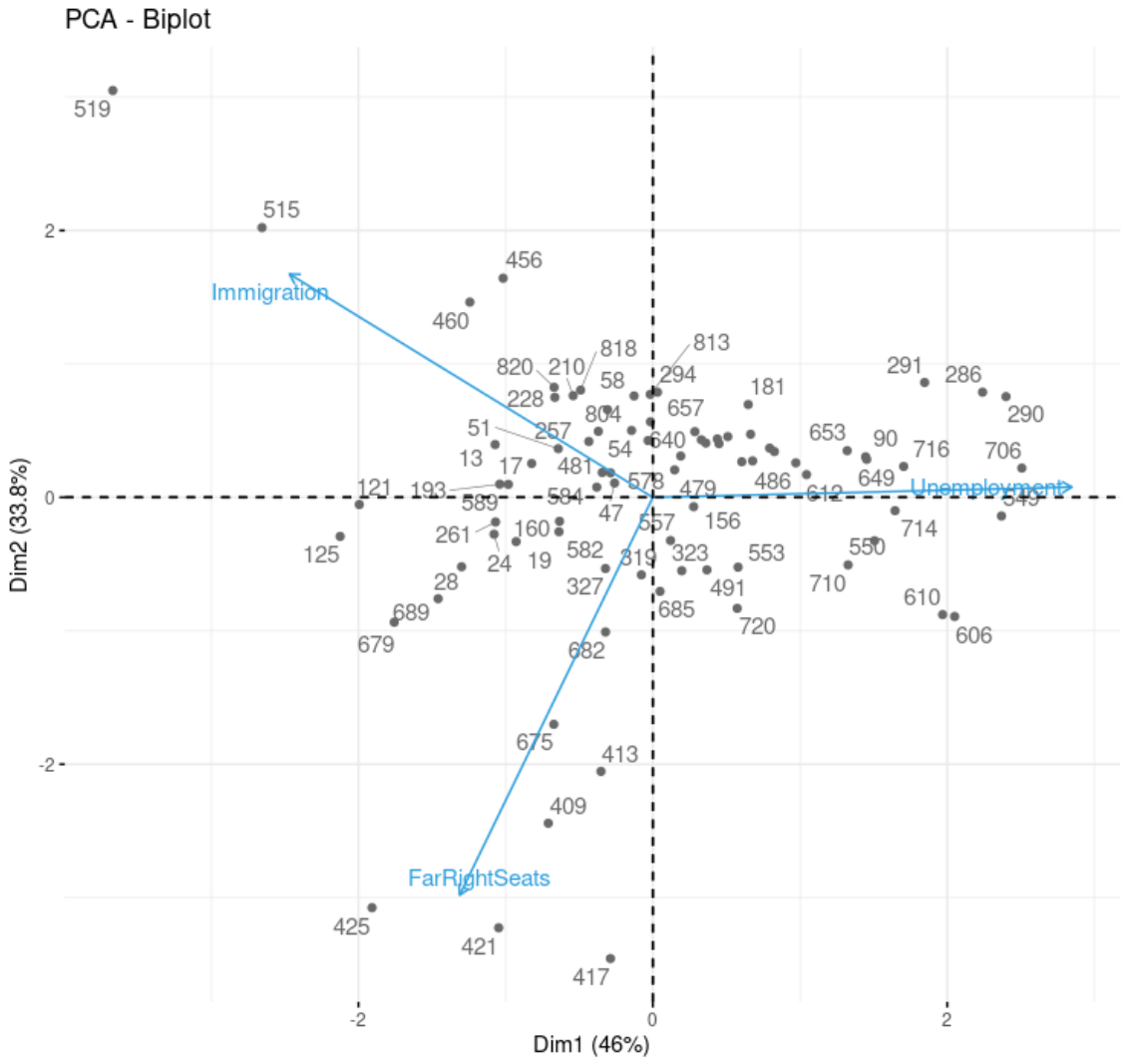

Figure 3: PCA representation of the final dataset.

4 Analysis

Considering the final dataset achieved and the little amount of time available after preparing the data, the dataset was analyzed following two models: regression trees and a hierarchical clustering.

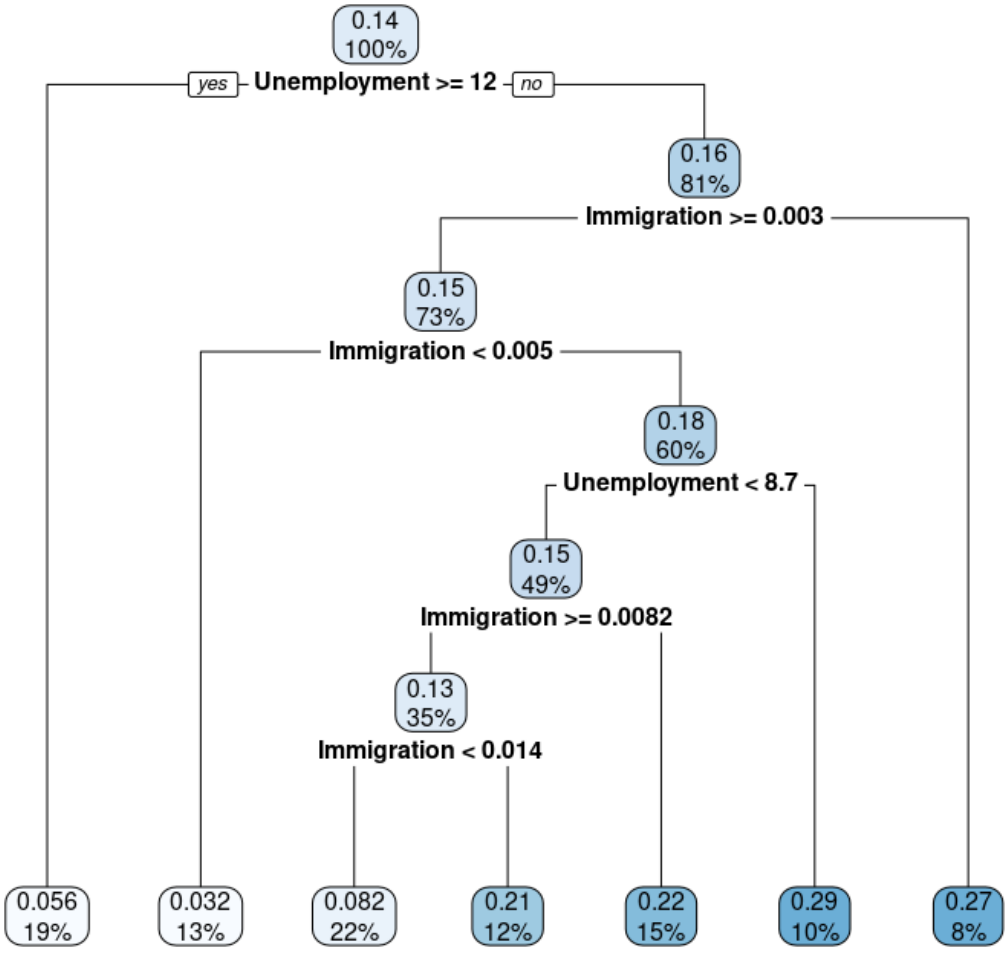

Figure 4: Results of applying Regression trees to the dataset.

5 Conclusion

The time spent for collecting and preparing the data in this project was so big compared to the amount dedicated to modelling and analyzing it that no conclusions can be made on the initial considerations and goals.

Furthermore, the final amount of useful data is not sufficient to perform any strong analysis. Nevertheless, a lot of improvements could still be made in that regard, like for example try to expand it by considering also regional results and not only the national ones or by trying to augment the dataset by generating artificial data coherent with the existing one.

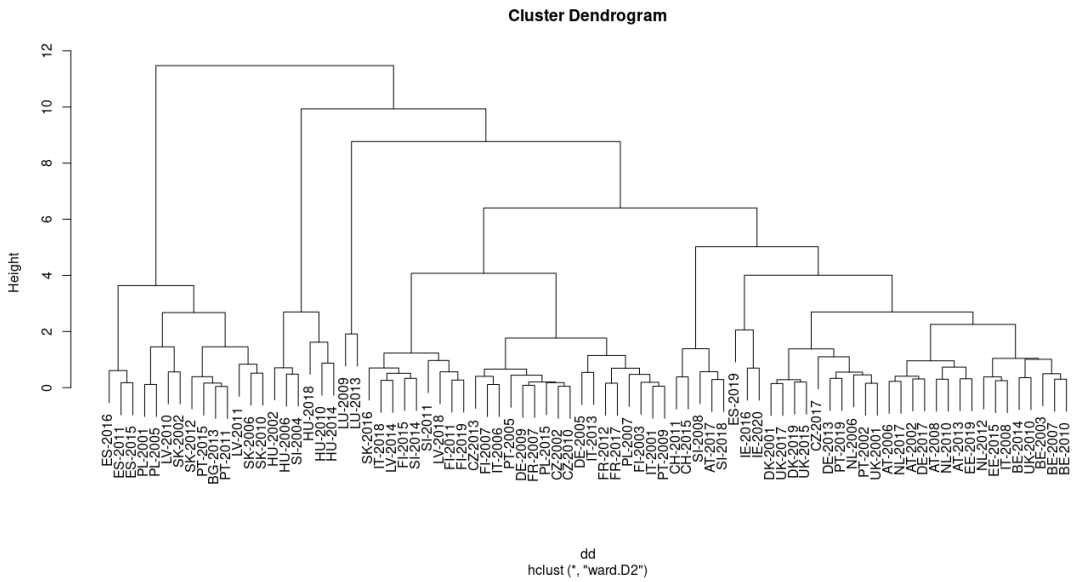

Figure 5: Results of applying Hierarchical clustering to the dataset.

6 Bibliography

References

[1] BBC. "Europe and right-wing nationalism: A country-by-country guide". In: *https://www.bbc.com/news/world-europe-36130006* (2019).

[2] the balance. "2007 Financial Crisis Explanation, Causes, and Timeline". In: *https://www.thebalance.com/2007-financial-crisis-overview-3306138* (2022).

[3] Elspeth Guild et al. *The 2015 refugee crisis in the European Union*. Centre for European Policy Studies Brussels, 2015.

[4] The New York Times. "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far". In: *https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html* (2018).

[5] Manel Álvarez-Rivera. "Election Resources on the Internet". In: *https://electionresources.org* (2022).

[6] Eurostat. "Eurostat Database". In: *https://ec.europa.eu/eurostat/web/lfs/data/database* (2022).