Bachelor

Bachelor

Bachelor in Software Development,
IT-University of Copenhagen

Jakob Melnyk, jmel@itu.dk
Frederik Lysgaard, frly@itu.dk

May 22st, 2012

Indholdsfortegnelse

1 Introducton

This report is the result of a project in Data Mining at the IT-University of Copenhagen during spring 2013. Source code, data sets and graphs can be found on the attached DVD-disk and on our GIT repository **https://github.com/esfdk/mdmi**

1.1 Goal of the project

This project attempts to cluster European nations based on population, balance of payments[1], unemployment rate and gross domestic product from 1975 to 2012. We visualise the resulting clusters to show if any one European nation always stays in the same cluster or if there is any nation that often changes cluster.
We will also see if there what countries that is most similar to the United States of America and Japan.
In addition we use a frequent pattern data mining algorithm to see if there are any frequent patterns in a nations values over the years.

1.2 Questions

[1] Balance of Payments is the method countries use to monitor assets and liabilities.[Investopedia, What Is Balance Of Payments?]

2 Data sets

3 Preprocessing

4 Algorithms

5 Results

5.1 K-means

Using the simpleKMeans clustering algorithm in Weka results in the graph in figure 5.1. The x-axis is the name of the country, the y-axis is the year and the color of the cross is the cluster the country belongs to at that time. The centroids of the clusters can be seen in the appendix REFERENCE!!!!!!

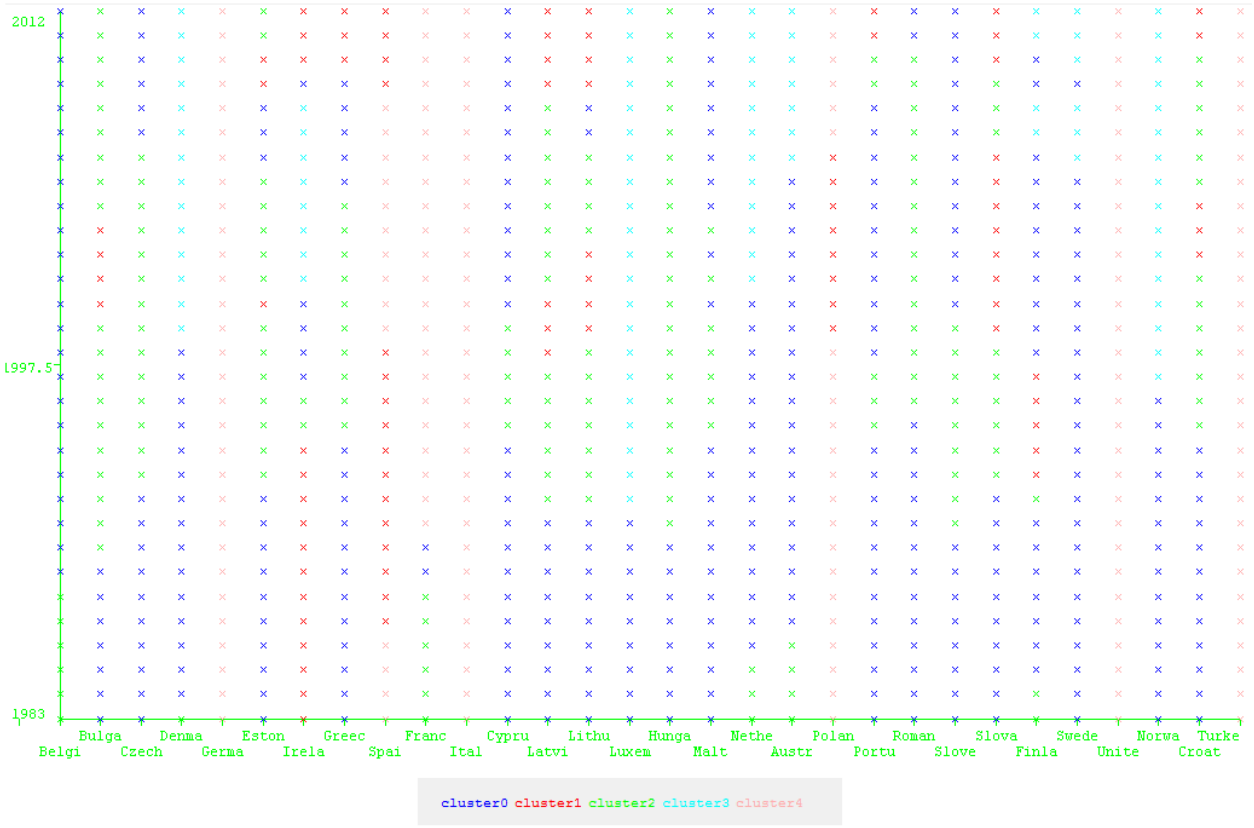

Figure 5.1: Graphical representation of what clusters the countries belong to at what year

6 Conclusion

Bibliography

[Investopedia] http://www.investopedia.com/articles/03/060403.asp, 8th of May

A Cluster Images

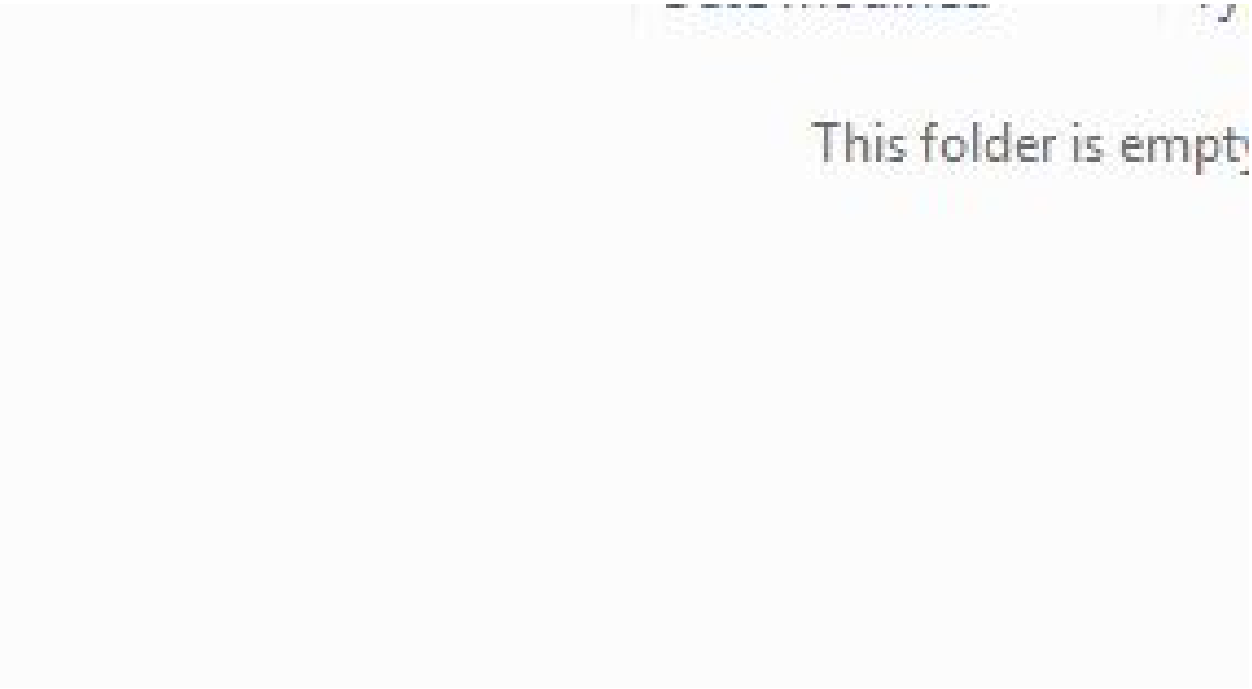

Figure A.1: Cluster Image 1