Exploration of socioeconomical datasets to model the democratic  
level of countries

**I.** **Abstract**

**II.Introduction**

**III. Materials and Methods   
 A. Datasets**

1. Democracy index from The Economist
2. *Wikipedia* Dataset
3. *Our World in Data* (OWD) Dataset

**B. Statistical tools**

      1.   Descriptive statistics and Principal Component Analysis (PCA)  
                   2.  Clustering   
       3.   Linear and logistic models  
       4.   Regression and classification trees  
       5.   Random Forest

**C. Handling of missing data**

**IV.** The most relevant variable explaining the democratic level

**A. An overview of the Data and correlations**

**1.**

**2. Comparison between the most and the least democratic countries**

**C.** **Modeling the democratic level (Gscore)**

1. Gscore as a quantitative variable

2. Gscore as a binary variable

**D. Highlighting the features which significantly explain the Gscore**

**V.** **Conclusion  
VI.Acknowledgement  
VII.References**