**Construction of an index to evaluate the democratic level of countries**

**Titre alternatif : Statistical study and model explanation of Democratic indexes**

**Selection of the variables which are the most relevant to explain the democratic level**

**I.** **Abstract**

**II.Introduction**

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

1. Democracy index from The Economist
2. Dataset built with variables from *Wikipedia*
3. Dataset built with variables from the website *Our World in Data*
4. Subdatasets with obvious democracies and autocracies

**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. Management 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**

**B. Reduction of the problem’s dimension**

**C.** **Explanation of the democratic level (Gscore) according to multiple features**

1. Study of the democratic level as an explicative quantitative variable

2. Study of the democratic level as an explicative binary variable

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

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