Project proposal Data Mining and Machine Learning

Where you found the dataset?

The World Happiness Report is an annual publication of the United Nations. The report is a measurement of life quality in the different countries all around the world.

This happiness measurement is based on different features that correlates with life factors. The different features of the dataset are Health information, the efficiency of social support, freedom level, democratic quality and others.

There are also information about the resource distribution between the residents of the countries. The dataset[1] can be found on the official website of the world happiness report[2].

The dimensionality of the data set is 1704 and the cardinality is 26.

The second source of data will be from OECD, The Organisation for Economic Co-operation and Development. It's an international organisation that works to build better policies for better lives. Their goal is to shape policies that foster prosperity, equality, opportunity and well-being for all. They draw on almost 60 years of experience and insights to better prepare the world of tomorrow.

What data mining machine learning algorithm you will use?

We are going to use different methods and tools that we learned during this class in order to extract and analyse the data from the dataset from the World Happiness Report and compare with other datasets from OECD.

We chose to work on a selection of countries from the original dataset that we have selected, USA, Brazil, China, Australie, France, Spain and Switzerland.

We will choose the most relevant features from the original dataset by choosing the ones that have the higher correlation with the happiness score (Life Ladder).

Then we'll take datasets from OECD to see what correlation exists for these features and the impact on the happiness score.

Finally, we'll try to do some predictions about the influence of the development slope (the features selected earlier) on the happiness score.

- Correlation between features
- Classification : gathering same characteristics among the countries

[1] https://s3.amazonaws.com/happiness-report/2019/Chapter2OnlineData.xls
[2] https://worldhappiness.report/ed/2019/

- Comparison : comparison between different countries and continents
- Analyse : predictions about future results

Any other related background about the usefulness and impact of this approach?

We will try to link our dataset with others to see if we can detect correlations with political and/or economic decisions from the different governments of the countries.

With the gathered information we'll be able to predict with some accuracy whether the development slopes would increase, decrease or stay the same.

The objective is to predict a development slope but also to analyse how and why this development occurs.

Team Members :

Haque Anika
Boudry Alexandra
Ahmed Shahzeb
Bouzerda Samy