Project Proposal **Effrosyni Simou**
EE-558 Date: 28/11/2016

Aim of Project

Since the 2016 Presidential Elections in the USA, the interest of people with regards to climate change and the correct environmental policy has reached an all-time high. In this project the aim is to use global temperature data from 1750 to 2015 [1] and let them give the answers.

Data Acquisition

For this project we will use a dataset, which was put together by *Berkeley Earth*, which is affiliated with *Lawrence Berkeley National Laboratory*. The dataset combines 1.6 billion temperature reports from 16 pre-existing archives. It is nicely packaged and allows for slicing into interesting subsets (for example by country). The data will have to be cleaned before they can be used since i) there are measurements that are missing in some cases ii) older measurements were less reliable *e.t.c.*.

Data Exploration

As far as data exploration is concerned, it would first be good to verify with a Chi-Squared test of independence the intuitive correlation between global temperature and year. Furthermore, interesting visualizations can be created with the given data. Such visualizations could be for instance i) histograms of average temperature in cities over the years ii) average global temperatures over the years ii) histograms of how many cities reached their maximum temperature ever in a given year *e.t.c.*..

Data Exploitation

Once we have a better understanding of the data offered in [1], we would like to use it in order to build a model that can predict the temperature at a given year. The predicted temperature could be for a city, a country or for the global temperature. Our model would have to regress the temperature, since temperatures assume values from a continuous set.

Evaluation

The machine learning model that we will build will be evaluated based on its accuracy in predicting the temperature. In order to do that, we will split the data in [1] in training data and test data. After training our model using the training data, we will evaluate it using the test data.

References

[1] Climate Change: Earth Surface Temperature Data. Exploring global temperatures since 1750. Available from World Wide Web: (https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data).