Task 2. Business understanding

Identifying your business goals

Background

Idea of this project is to predict parts of world in the future with most less polluted natural resources like water and air. It is possible that extra characteristics will be pursued. Since pollution created by big industries is major problem on this always forward rolling ball of dirt, it may be wise to start looking sweetspots on planet for human inhabitation.

Business goals

Find correlations between clean energy production increase by regions and their pollution levels. Create model to predict Planet's 'clean' area in numbers of square kilometers during this century.

Business success criteria

Find obvious correlations and also hidden from sight ones. Succeed in building the model to predict 'clean' area on Planet up to 50 years with 10 year increments.

Assesing your situation

Inventory of resources

- Mid to high end laptop.
- High end personal computer.
- Complete dataframes from Kaggle.
- Personal: One dataminer.

Requirements, assumptions, and constraints

Need to find more dataframes for pollution levels to match the 20 year dataframe of renewable energy.

Risks and contingencies

Loss of data on computer hard drive failure. Solution is to duplicate project through version control system and make extra copies to external drive.

Terminology

Clean energy dataframe – Dataframe from Kaggle containing 20 years of data.

Pollution dataframe – Merged dataframe from multiple in Kaggle available pollution dataframes.

Complete dataframe – One dataframe to have all clean energy and pollution data.

Predictive model – Final model to display results.

Costs and benefits

costs

- Energy to run hardware – 20 euros.
- Worker wages – 0 euros.

benefits

- Succeeding in building predictive model – priceless.
- Datamining techniques knowledge increase – priceless.

Defining your data-mining goals

Data-mining goals

Create the predictive model and find good ways to present all findings on the globe in 'Notebook'.

Data-mining success criteria

Succeed in making the model with acceptable accuracy between 55-75%.

Task 3. Data understanding

In works….

Task 4. Planning your project

- Find more pollution dataframes to match 20 year period of renewable energy dataframe.(4h)
- Clean pollution dataframes of unwanted data.(4h)
- Merge all dataframes into one complete dataframe with all data.(4h)
- Look for correlations between pollution and renewable energy.(10h)
- Build predictive model up to 50 years relying on complete dataframe.(20h)
- Presentation of data.(5h)
- Work on Github.(3h)

Methods and classifiers to be determined yet.