Dataset and study area

The objective of the prediction module is to predict the emission rate at the stack/chimney. To build and test the prediction model dummy data was generated using Gaussian distribution to randomly generate values for the feature set. The feature set is chosen on the bases on what factors will correlate to the emission rate and the type of pollutants emitted from the stack.

The feature set consists of independent variables: day, month, year, type of industry, size of industry, and output efficiency of industry.

And independent variables: emission rate.

Definitions:

* Type of industry: what the industry/factory produces which will correlate to what gases are emitted out of the stack/chimney, this will be in the form of labeled classes.
* Size of industry: how big is the industry/factory which will correlate how much the maximum is outputted, this will be represented in the form of a scale from 1 to 10.
* Output efficiency: the amount of output it produces each day divided by the total amount of output it can ideally produce.
* Emission rate: this is defined as the amount of pollutants released from the stack per unit time.

The prediction module is also used on the collected metrological data to predict the air velocity and direction, this is then applied to calculate the dispersion of pollutants in air.