|  |  |
| --- | --- |
| Title | Notes |
| [Predicting Trends in Air Pollution in Delhiusing Data Mining](file:///home/damian/Dokumenty/studia/masters-thesis/thesis/articles/delhi_air.pdf) | Analysis of air pollutants in Delhi using:   * linear regression * multi-layered perceptron   Not much about methodology. The article covers mainly results of the analysis. |
| [ANALIZA I PROGNOZA ZANIECZYSZCZENIA POWIETRZANA PRZYKŁADZIE AGLOMERACJI MIEJSKIEJ KRAKOWA](file:///home/damian/Dokumenty/studia/masters-thesis/thesis/articles/krakow_regresion_Fuksa.pdf) | Prediction of air pollution in Krakow based on historical data from 3 observation points. Prediction based on various types of regression techniques:   * linear * exponential * hyperbolic * logarithmic * power   No additional machine learning methods have been covered. |
| [Air quality prediction in Milan: feed-forwardneural networks, pruned neural networks andlazy learning](file:///home/damian/Dokumenty/studia/masters-thesis/thesis/articles/milan.pdf) | The article discusses 3 modelling methods:   * feed-forward neural networks * pruned neural networks * lazy leaning (an auxilliary article: lazy\_learning.pdf)   It points out the need of aggregate some variables (for example wind speed – it’s possible to use an average speed based on a 6 hour period but not each of the 24 daily records).  How to choose variables – cross-correlation.  Problem: non-linear model, linear cross-correlation analysis, but it’s an established practice in the field.  Measured of prediction quality:   * Mean Absolute Error * Mean Bias Error * true/predicted correlation.   It has been found that the level of PM10 shows a periodic behavior: it is highest during winter and significantly lower on weekend. Despite this fact it hasn’t been proven that data deseasonalization causes the prediction efficiency to increase. |
| [Forecast of Air Quality Based on Ozone byDecision Trees and Neural Networks](file:///home/damian/Dokumenty/studia/masters-thesis/thesis/articles/mexico_city.pdf) | The article presents data gathered in 4 meteorological stations across Mexico City. Authors propose a few models for the prediction of the level of air pollutants based on:   * neural networks * decision trees (C4.5, Random Forest)   The paper describes the process of data preprocessing and selection of parameters taken into consideration while building a predictive model. The variable selection procedure is based on the chi-square metric.  Pollution level value needs to be discretized (based on the NOM-1993 norm) before being applied in the subsequent phases.  Data has been divided into chunks based on the time of the year.  Best results for the Random Forest algorithm (although not by a large margin).  Best accuracy in the noon when the level of pollution is low.  // The article contains multiple grammatical and stylistic errors. |
| [Extensive evaluation of neural network models for theprediction of NO 2 and PM 10 concentrations, comparedwith a deterministic modelling system and measurementsin central Helsinki](file:///home/damian/Dokumenty/studia/masters-thesis/thesis/articles/helsinki.pdf) |  |
|  |  |
|  |  |
|  |  |
|  |  |