

Warsaw University of Technology  
The Faculty of Electronics and Information Technology

Data Mining (EDAMI)  
Project Documentation

---

# CLUSTERING BASED ON DENSITY

---

*Author:*

ALEKSANDRA KURDO  
ADAM STELMASZCZYK

June 13, 2013

## Project task

Implementation and experimental evaluation of DBSCAN [?] and DEN-CLUE [?] algorithms.

## Solution specification

### Data set

[?]

Set with information about geometrical properties of kernels belonging to three different varieties of wheat was chosen.

### Data set attribute information

To construct the data, seven geometric parameters of wheat kernels were measured:

- area A,
- perimeter P,
- compactness C,
- length of kernel,
- width of kernel,
- asymmetry coefficient
- length of kernel groove.

All of these parameters were real-valued continuous.

## Algorithms description

### DBSCAN algorithm

### DENCLUE algorithm

Denclue algorithm is based on the idea that the influence of each data point can be modeled using a mathematical function (influence function). The overall density of the data space can be calculated as the sum of the influence function of all data points. Clusters can be determined mathematically by identifying density-attractors, which are the local maxima of the overall density function.

The Denclue algorithm works in two steps.

Step one:

- It is preclustering step, in which a map of the relevant portion of the data space is constructed. The map is used to speed up the calculation of the density function which requires to efficiently access neighboring portions of the data space.

Step two:

- It is the actual clustering step, in which the algorithm identifies the density-attractors and the corresponding density-attracted points.