DATA MINING 2017/18

List of project proposals

1. An algorithm for classification using decision trees (including the procedure for evaluating the value of a test).
2. Apriori Some algorithm for finding sequential patterns.
3. Apriori All algorithm for finding sequential patterns.
4. Parallel algorithm for constructing decision tree – partitioned approach.
5. Parallel algorithm for constructing decision tree – synchronous approach.
6. Agglomerative Clustering with k-means final reclustering.
7. Divisive Clustering with k-means final reclustering.
8. Clustering of web-site visitors based on their URL traversal patterns.
9. Sales forecasting using data mining methods.
10. Identification of stock trading rules from historical market data.

Project Requirements:

- Each of the above project proposals is prepared for a group of two students;
- You are welcome to choose projects outside this list or to create your own project proposal BUT the project proposal must be approved (projects must have direct relevance to the material covered during the lecture);
- The final version of the project should consist of:

-the computer program,
-data sets for tests,
-the program documentation.

The documentation should contain:

1. Description of the considered problem.
2. Background (main known results, used known algorithms).
3. Description of the own solution (new algorithms, methods, techniques or modifications of some algorithms).
4. Algorithmic complexity and correctness (quality) analysis.
5. User Manual.
6. Technical documentation (comments in source, description of main data structures and procedures).
7. Description of tests (used data sets, results, conclusions).
8. Conclusions, comments.
9. List of references to literature, web pages.

Project Grading: (Total 40 points max.)

- Project documentation **(20 points max.)**
  - Description of the problem and the solution (sections 1., 2., 3.) **(0-4 pts)**
  - Algorithmic complexity and correctness analysis (section 4.) **(0-4 pts)**
  - User's manual (section 5.) **(0-4 pts)**
  - Technical documentation (section 6.) **(0-4 pts)**
  - Description of tests and conclusions (sections 7., 8.) **(0-4 pts)**
- Computer program, preparation of the data sets and tests, presentation.**(20 points max.)**

Deadline for the final submission of the project: **Friday, January 19, 2018**.
Penalty for late submission: **-4 points** per late week (max. -16 pts) (starting from Monday, January 22, 2018, penalty stays the same for the whole week).
Both members of each group are expected to contribute equally towards theirs team progress!