“Data Mining”

A Course for Doctoral students,
Doctorate School in Computer Science and Engineering,
University of Bologna

| Instructor and affiliation | Prof. Claudio Sartori  
Prof. Stefano Lodi  
All instructors belong to:  
DEIS, University of Bologna |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time span</td>
<td>10 hours</td>
</tr>
<tr>
<td>Final exam</td>
<td>Short essay on agreed topic</td>
</tr>
</tbody>
</table>

| Instructor names          | Claudio Sartori  
Stefano Lodi  
DEIS, University of Bologna |
|---------------------------|-----------------------------------------------------------------|

Course Outline
Data Mining principles and algorithms

Objectives
This course will introduce the problems related to data mining.
The amount of data collected by organizations in day-to-day operations is growing rapidly, and standard analysis tools, such as reports and queries, are not suitable to give interesting, general information to be exploited for strategic decisions. For this reason in the last decade the research on data mining focused on the discovery of general patterns from large amounts of data. At present data mining services are offered by most DBMS systems and specialised industry level software, nevertheless there are several problems yet to be solved, and the research on this topic is also very active.
In this course, first a formal definition of the various data mining objectives will be given, then some specific problems, such as clustering and classification, will be examined in depth. Some case studies will also be examined, to give an idea of the possible applications.
A more detailed schedule follows:

- Knowledge discovery process (2 hours)
  - Objective definition
  - Selection of data sources
  - Data filtering, reconciliation and transformation
  - Data mining
  - Data validation and presentation

- Classification (4 hours)
  - Definition of the problem
  - Decision trees
  - Support vector machines
  - Models evaluation
  - Research issues

- Clustering (4 hours)
  - Definition of the problem
  - Partitive methods
  - Hierarchical methods
  - Density-based methods
  - Research issues

Who should attend
The course aims to be of benefit for research students in the areas of databases and information systems, and also of computer science/computer engineering in general.

Learning and assessment modalities
The course will be taught in either Italian or English according to the preference of the attendants.
The final assessment consists of a short essay on a topic agreed between the student and the lecturer.

Materials
Slides and reference materials are in English and will be provided after the classes.

Detailed topics/Schedule

1. Monday 30-05-2011 h.14-16 – Room 0.6 - Knowledge discovery process
2. Tuesday 31-05-2011 h.09-11 – Room 5.1 – Classification, part 1
3. Monday 06-06-2011 h.14-16 – Room 0.6 – Classification, part 2
4. Wednesday 08-06-2011 h.14-16 – Room 4.1 – Clustering, part 1
5. Friday 10-06-2011 h.12-14 – Room 1.4 – Clustering, part 2

Reference to Google Calendar
https://www.google.com/calendar/ical/siqfj86pdgo04uet8pde2f0soo%40group.calendar.google.com/public/basic.ics