Data Mining Lab

Lothar Richter & Juan Miguel Cejuela

Apr. 29th

Institut für Informatik I12
## Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr 15th</td>
<td>Kick-off</td>
<td>Jun 3rd</td>
<td>Predictive Mining I</td>
</tr>
<tr>
<td>Apr 22th</td>
<td>Data Set Presentation</td>
<td>Jun 10th</td>
<td>Predictive Mining II</td>
</tr>
<tr>
<td>Apr 29th</td>
<td>Data Set Selection</td>
<td>Jun 17th</td>
<td>Predictive Mining III</td>
</tr>
<tr>
<td>May 6th</td>
<td>Descriptive Mining I</td>
<td>Jun 24th</td>
<td>Predictive Mining IV</td>
</tr>
<tr>
<td>May 13th</td>
<td>Descriptive Mining II</td>
<td>Jul 1st</td>
<td>Predictive Mining V</td>
</tr>
<tr>
<td>May 20th</td>
<td>Descriptive Mining III</td>
<td>Jul 8th</td>
<td>Final Presentation</td>
</tr>
<tr>
<td>May 27th</td>
<td>Descriptive Mining IV</td>
<td>Jul 15th</td>
<td>Final Presentation</td>
</tr>
</tbody>
</table>
Infrastructure

- wiki: [https://i12r-studfilesrv.informatik.tu-muenchen.de/dmlabss2015/](https://i12r-studfilesrv.informatik.tu-muenchen.de/dmlabss2015/)

- desktop computers for the course:
  - student lab: 01.08.021
  - you can pick up access cards for a deposit of 20 Euro from our secretary
  - hosts: i12k-biolab[01-14].informatik.tu-muenchen.de
  - Linux: Ubuntu
Infrastructure

- computation host: dataminer.informatik.tu-muenchen.de
- accessible via ssh on port 8574
- 24 cores
- 128 GB main memory
Task of the Week

- complete group formation
- create an respective entry on the wiki and link from the main page
- add your names
- retrieve the data set
- start with a naive dataset inspection and describe your findings:
Task of the Week

- determine the value range of an attribute
- calculate simple statistics
- simple countings
- remember that our tools tuple based
- is a relational database an option?
- think about visualizations like e.g. Google Charts, www.d3js.org, a.s.f.

......