

PSS718 - Data Mining

Policy and Strategy Studies

Asst.Prof.Dr. Burkay Genç

Hacettepe University

September 26, 2016

Who Am I?

Asst. Prof. Dr. Burkay Genç
(Industrial Engineer, Computer Scientist)
Institute of Population Studies
Policy and Strategy Studies

Computational Geometry, Game Technologies, Data Analysis, Social
Networks



What is Data Mining?

Data Mining is the science of extracting information hidden in structured or unstructured data. Data usually comes dirty, noisy and unstructured. We have to clean it, remove noise, and structure, so that we can process it. Then, we can dive deep into the data to extract information out of it.

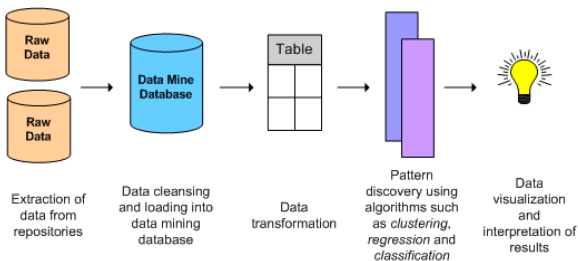


Figure: Steps of Data Mining



Why?

- Determine relations in data
- Detect possible improvements
- Understand the system
- Model
- Predict



Course Identity

Course Name	PSS718	Semester	The.	App.	Credit	ECTS
Data Mining	PSS718	G/B	3	0	3	10



- Assignments (A): 2 or 3 assignments \rightarrow 50%
- Project or Final (P) \rightarrow 50%
- Overall (O) \rightarrow (A) + (P) \rightarrow 100%
- Opinion Grade (K) \rightarrow 0.8 - 1.2
- Assigned Grade (G) \rightarrow (O) * (K)
- Pass Grade \rightarrow 60

What does that mean?

- O = 100 \rightarrow 80 - 100
- O = 75 \rightarrow 60 - 90
- O = 50 \rightarrow 40 - 60



Content to be covered:

- Working with Data
- Loading Data
- Exploring Data
- Graphics
- Descriptive and Predictive Analytics
- Cluster Analysis
- Association Analysis
- Decision Trees
- Random Forests
- Boosting
- Support Vector Machines



- **Data Mining with Rattle and R**, Graham Williams, Springer
- The Web



- R: Free! -> <https://www.r-project.org/>
- R-Studio: Free! -> <https://www.rstudio.com/>
- Rattle: Free! -> Install in R-Studio

