

# Introduction to Machine Learning and Data Mining

Course contents & “rules of the game”

## 0. (INTRODUCTORY) LECTURE

---

assoc. prof. Branko Kavšek

# About the course

---

**Course name:**

Introduction to Machine Learning and Data Mining

**Teacher:** Branko Kavšek

**Assistant:** Domen Vake

**Course type:** elective/compulsory

**Students:**

- Students of the study program **Computer Science** and **Bioinformatics** (1. level) UP FAMNIT – 2. and 3. year

**Prerequisites:** no prior knowledge needed

useful: basic statistics, basic programming (in Python)

**e-classroom:** <https://e.famnit.upr.si/course/view.php?id=5620>

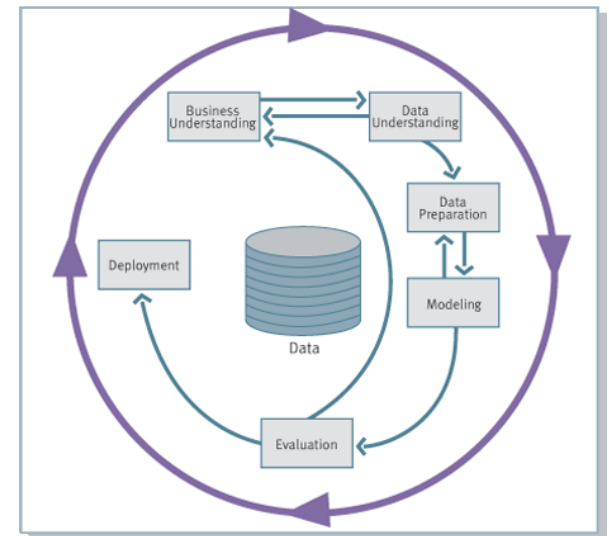
# Course contents – theory

Knowledge discovery (searching for patterns) in data using the CRISP-DM methodology:

- Problem understanding
- Data understanding
- Data preparation
- Modeling – pattern discovery
- Evaluation
- Deployment

Source:

[https://en.wikipedia.org/wiki/Cross\\_Industry\\_Standard\\_Process\\_for\\_Data\\_Mining](https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining)



# Course contents – practice sessions

---

Using data mining techniques on real data:

- “manually”
- with aid of (computer) tools:
  - Advanced text editors (gedit, Notepad++, ...)
  - Spreadsheet editors (Calc, Excel, ...)
  - Open-source data mining toolboxes ([WEKA](#), Python, ...)

# Rules of the game

---

The final grade is “composed” of:

- The **written exam** grade – or 2 mid-terms (**50%**)
  - Condition: written exam has to be “positive” ( $\geq 50/100$  points),  
mid-terms have to be both “positive” ( $\geq 40/100$  points);
- The **homeworks** grade  $\sim$  1 homework/week (**15%**)
  - Condition: at least 50% of all points achieved;
- The **programming project** on tutorials (**30%**)
  - Condition: submitted on time and positively graded project;
- The **oral exam** grade (**5%**):
  - Condition: a “satisfactory” answer to each of the 3 questions;

Detailed information in the **Course Syllabus**:

<https://e.famnit.upr.si/mod/page/view.php?id=107177>

# The path towards the final grade

Activity	Grade part	Condition
Written exam / mid-terms	<b>50 %</b>	At least half of all points achieved, not less than 40% for each mid-term.
Programming project	<b>30 %</b>	Working program + documentation, submitted on time = 100 % late submissions = – %.
Homeworks	<b>15 %</b>	At least 50% of all points achieved, submitted on-time = 100 % 1 week late = 50 % > 1 week late = 0 %.
Oral exam	<b>5 %</b>	At least satisfactory knowledge demonstrated at answering each of the 3 questions.
<b>Total:</b>	<b>100 %</b>	<b>All activity conditions fulfilled</b>

**95 to 100** points = excellent (**10**),

**85 to 94.99** points = very good (**9**),

**75 to 84.99** points = very good (**8**),

**65 to 74.99** points = good (**7**),

**50 to 64.99** points = sufficient (**6**),

**less than 50** points = insufficient (**fail**)

# Literature and other sources

---

1. Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition, Morgan Kaufmann, 2016.
- 
2. Mohammed J. Zaki, Wagner Meira, Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 2014.
3. David J. Hand, Heikki Mannila and Padhraic Smyth, *Principles of Data Mining*, MIT Press, Fall 2000.
4. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Verlag, 2001.
5. Tom Mitchell: *Machine Learning*, McGraw Hill, 1997.
- 
6. UCI ML Repository: <http://archive.ics.uci.edu/ml/>
7. Kaggle: <https://www.kaggle.com/>
- 
8. WEKA software: <http://www.cs.waikato.ac.nz/ml/weka/>