

MI3.03a Data Mining

Tran Giang Son, tran-giang.son@usth.edu.vn

ICT Department, USTH



Course Introduction



Goals

- Be familiar with the principles of data mining algorithms
- Implement data mining algorithms
- Apply data mining techniques for new problems
- Do the project
 - and (hopefully!), pass the exam



Content

- Introduction to Data Mining
- Data Pre-processing and Post-processing
- Clustering
- Classification
- Multi Layer Perceptron
- Convolutional Neural Network



Format

- 2.5 ECTS = 25 hours
- Lecture / Practical work : 10h / 15h
- Prerequisites: C/C++ Programming
- Environment: Linux/Mac
 - Windows isn't supported
- Assessment:
 - Project / Final Exam



Policy

- Collaborations!
 - During exercises
 - Not in the project
 - Nor in the final exam
- **All** students finish, we move to the next exercise



Labwork

- Several Python Programs
 - Understand what you learnt
 - Show your ability to apply it to new problems
- Don't copy paste. I have checker tools ☺
- Should be well organized and well written

Bad programmers worry about the code. Good programmers worry about data structures and their relationships.

- Linus Torvalds

Labwork

- Git : Version Control System
- Github
 - Initial repository and instruction
 - <https://github.com/SonTG/dm2021>

	COMMENT	DATE
○	CREATED MAIN LOOP & TIMING CONTROL	14 HOURS AGO
○	ENABLED CONFIG FILE PARSING	9 HOURS AGO
○	MISC BUGFIXES	5 HOURS AGO
○	CODE ADDITIONS/EDITS	4 HOURS AGO
○	MORE CODE	4 HOURS AGO
○	HERE HAVE CODE	4 HOURS AGO
○	AAAAAAAAA	3 HOURS AGO
○	ADKFJSLKDFJSDKLFJ	3 HOURS AGO
○	MY HANDS ARE TYPING WORDS	2 HOURS AGO
○	HAAAAAAAAAANDS	2 HOURS AGO

AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.

Don't be lazy with your commit messages

Exams

- 3 sheets of A4 documents are allowed
- No laptop / mobile phone / internet
- No discussion, of course



References

- Books
 - Introduction to Data Mining, 2nd Edition, Pang-Ning Tan et al., Addison Wesley
 - Mining of Massive Datasets, 3rd Edition, Jure Leskovec et al., Cambridge University Press

