Course description

General information		
Course leader	**Nicolas LE HIR**	
Course title	**Advanced Visualization of Massive Data**	
Study programme	**Title of Expert in Information Technology**	
Course status	**Graduate Program**	
Year	2	
Number of credits and mode of teaching delivery	ECTS student workload coefficient	3
	Number of hours (L+E+S)	(14+0+60) 74

1. COURSE DESCRIPTION

1.1. Course objectives

The module "Visualization of Massive Data" aims at understanding the problematic of big data and its visualization. The questions "Why companies or other institutions have to address the processing of big data?" and "What is this problem specific?" will be discussed during the lecture. A first answer would be to say that it is due to the need for efficient algorithms which process online data and produce results which need to be understood by human being to make strategic decisions. In this sense, visualization of big data is an important way of presenting the results of the algorithms. It can give some precious information on the data structure and it can lead to find trending behavior.

1.2. Conditions for enrolment in the course

- Have elementary mathematical notions related to geometric data and calculus.
- Be familiar with Python 3 and with the installation of new libraries.

The following elements will be helpful:

- Have elementary notions of probabilities.
- Have elementary notions of linear algebra.
- Be familiar with the scientific computing libraries as numpy.

1.3. Expected learning outcomes of the course

LO1: Explain the problematics of big data and its visualization

LO2: Select a tool for visualization and exploratory analysis of a given problem and argue the choice analysing its functionalities.

LO3: Select appropriate techniques of analytical navigation and interaction for a given problem, and explain your choice

LO4: Link data from multiple sources for visual analysis and create complex interactive dashboard

LO5: Manage cloud based advanced visualization tools followed by AI supporting logic

LO6: Use data and advanced visualization tools combined with large data sets to create data-driven story

LO7: Join data from multiple sources, aggregate, filter and restructure tabular data, choose and apply appropriate method for handling missing values.

LO8: Apply and interpret the results of simple machine learning algorithms and statistical models

LO9: Achieve graphical integrity of given project example by using adopted tools and techniques

1.4. Course content

The lectures are structured on two days.

Lecture 1:

- Definition of big data
- Orders of magnitude
- Definition of big data visualization
- Data modelling
- Supervised learning
- Unsupervised learning
- Hierarchical clustering
- Reminder on random variables
- Overfitting
- Main component analysis

Lecture 2:

- Classical visualization methods
- Scatter matrices
- Parallel coordinate plots
- Miscellaneous : radar plots, heat maps
- Treemaps
- Visualization platforms
- Classification trees
- Reliability, metrics, prediction errors

The project consists in presenting or building a dataset and producing multiple visualizations of the data. At least, two types of visualization are to be produced by the students and the user of the visualizations should be able to fine-tune them. The visualizations should allow to identify the structure and trends in the data and the results should be commented by the students in a separate document.

A quantitative analysis of the data is also to be produced. The students have the choice between learning predictive model of a column as a function of another column (supervised learning), or learning a distribution of one or several columns of the dataset (unsupervised learning). This part of the project should also be commented in the same separate document as mentionned above.

1.5. Teaching delivery modes:	☒ lectures ☐ seminars and workshops ☒ exercises ☐ remote learning ☐ field work	☒ independent work ☐ multimedia and network ☐ laboratory ☒ mentoring ☐ other _____
1.6. Comments	Presentation on given topics are part of the modules. The presentations are interactive, and the students can ask questions. The students work on coding exercises during the class. All the material of the module– course slides, exercises and example of python scripts - is available on a repository.	
1.7. Student obligations		
STUDENT ATTENDANCE Class attendance is mandatory in the percentage prescribed by the Studies and examination regulations. PASSING EXAM Each groups of students must submit their results to the teachers and give an oral presentation based on their results. The results are reviewed by the academic staff during an oral examination. A justification of the project work can be explained by the students.		
1.8. Monitoring[1] student work		

[1] IMPORTANT NOTES: Next to each method of monitoring student work it is necessary to insert an adequate share of each activity in ECTS credits, so the total number of ECTS credits corresponds to the credit value of the course. You can use empty fields for additional activities.

Class attendance		Activity during class		Seminar paper		Experimental work	
Written exam		Oral exam		Essay		Research	
Project	100 %	Continuous assessment of knowledge		Student report		Practical work	
Portfolio							

1.9. Assessment and evaluation of student work during classes and the final exam

The students are assessed on the project results and the content.

A feedback will be given to the students, according to the results of the project.

CONCRETE REVIEW OF EVALUATION METHODS

The maximum number of points that a student can earn in a course is 100. Grades are calculated according to the following criteria table within which the distribution of passing grades in terms of the number of points is applied.

Points	Grade
0,00 - 50,00	(E) unsatisfactory
50,01 - 58,00	(D) sufficient
58,01 - 75,00	(C) good
75,01 - 92,00	(B) very good
92,01 - 100,00	(A) excellent

The method of accumulating points is determined in this course in accordance with the elements of scoring as follows:

Criterion	Maximum points
Project	100
TOTAL	100

1.10. Required reading (at the moment of submitting the joint study programme report)

Python Software Foundation.“Python“, 2020. https://www.python.org/

Sickit-Learn."Machine Learning in Python", 2020. https://scikit-learn.org/stable/

NumPy."NumPy", 2020. https://numpy.org/

SciPy."SciPy.org", 2020. https://www.scipy.org/

1.11. Additional reading (at the moment of submitting the joint study programme report)

Laurent Miclet, Antione Cornéjuols. "Apprentissage artificial". *Concepts et algorithmes*/ Eyrolles, May 2018.

- Bikakis, N. 'Big Data Visualization Tools'. *Encyclopedia of Big Data Technologies*, Sprigner, 2018.

- Olshannikova, E., Ometov, A., Koucheryavy, Y., and Olsson, T. Chapter 4 'Visualizing Big Data'. *Big Data Technologies and Applications (pp.101-131),* 2016.

- IBM Institute for Business Value. 'Analytics: The real-world use of big data'. Research reports.

1.12. Number of copies of required reading in relation to the number of students who currently attend a course

Title	*Number of copies*	*Number of students*

1.13. Methods of quality monitoring that ensure the acquisition of knowledge, skills and competencies.

The content of each modules is continuously revised to teach the students on the most up-to-date notions and concepts of IT. Indeed, the range of skills and knowledge in this sector is constantly getting broader, with a larger perspective of working in many different fields.
To ensure the quality of the teaching, a Steering Committee supervises the Quality Management System. The evolution of the teaching content is revised and validated by the Development Council. The teachers as well as the administration staff are evaluated by the students themselves. Finally, the teaching content is analysed and determined by evaluating the skills during the internships, by the partner companies.