Fakultät für Informatik
Facoltà di Scienze e Tecnologie informatiche
Faculty of Computer Science

unibz

COURSE DESCRIPTION – ACADEMIC YEAR 2020/2021

Course title	**Development of Data Products**
Course code	73007
Scientific sector	INF/01
Degree	Master in Computational Data Science (LM-18)
Semester	1
Year	2
Credits	6
Modular	No

Total lecturing hours	40
Total lab hours	20
Attendance	Attendance is not compulsory but highly recommended due to the strong practice-based and experiential contents, even during theoretical hours. Non-attending students will therefore have to contact the lecturer at the start of the course to agree on the modalities of the independent study.
Prerequisites	Sufficient computer skills to autonomously create a data science solution, using previous knowledge acquired in the respective studies (Data Analytics or Data Management) and any programming, BI or Data Science tool the student is comfortable with (R, Python, Java, C#, etc.)
Course page	https://ole.unibz.it/

Specific educational objectives	The course belongs to the type "caratterizzanti – discipline informatiche" in the curricum "Data Analysis".
	In data science, a lot of attention has been focus on the development techniques, with pure programming or with data science-ready environments (like R, SAS, Matlab or Python-related environments like Anaconda, PyCharm, etc.). Less attention has been devoted to the management of the whole lifecycle for the creation of a data product. The educational objectives of the course are therefore related to the ability to define, manage, and deliver data products considering the whole lifecycle of a product, not only the software development phase. A data product is a complete, fully functional, installable, and maintainable software solution that facilitates an end goal of its users through the prominent and intensive use of data. Under this perspective, the development of a data product using whatever software solution (pure programming, BI tools, or Data Science tools) is like any other productization: it must be well understood since the very beginning (for example, stakeholders' perception of the solution and their business objectives, not only the technical aspects of it), and must have all the characteristic of a product ready to be sold, not just a shiny piece of software mostly used by the creators of it. Creating data products, delivering, and maintaining them has some peculiarities of the IT processes of software engineering, some others of the data science tools and techniques, but many aspects are similar to any other product creation. This means that looking at a data product from the perspective of the project management world will be very beneficial for any data scientist. The best-of-breed combination

of data science and project management techniques are needed to successfully lead and profitably deliver a data product.

The course therefore will introduce the student to the Data Science Project Management field, a promising and highly requested set of competencies in today's data science projects, a mix of data science, project management and software engineering tools and techniques. The course aims at providing a solid understanding of the project management standards and aspects in which the data product creation (analysis, design, implementation, test, delivery, etc.) is a relevant part, but not the whole story. Scope, time, cost, resources, risk, communication, and stakeholder management are typical areas of knowledge for the project management world, that perfectly fit projects for data product's creation. Some examples of questions where project management tools and techniques are fundamental, and data science/software engineering background are not enough, are the following:

- how much does this data product cost to my company, and what final price will the customer pay?
- When will the final data product be delivered? Are we able to maintain promised deadlines?
- When will I have data available so to start the development?
- As another concurrent project needs part of our resources, how many resources do we need during the third month of our project?
- Do we have a risk management plan that is to be presented in the BoD or our company?
- What happen if task XYZ will have to stop because data is not available?
- How do we re-schedule the product tests if the development will be delayed by 3 weeks and how do we reallocate resources due to this delay?
- What is the impact of this delay on the other projects our organization is running?

These and many other questions normally receive general or no answer from data scientist, normally too concentrated on the technical aspects and less (or in no way) interested in these vital aspects for any paying customer, and for the company's financial health and reputation. As brutal and simple as it seems, this is where data science and project management collide, because a bullet-proof, innovative ML solution is nothing if not delivered in time, within the promised cost, and respecting many other functional and non-functional constraints.

The course will provide adequate project management tools and techniques for the role of IT manager, CTO, CIO, CDO, Data Scientist, Data Engineer etc., where deep technical competences must be flanked by strong project management competencies. The course provides a comprehensive theoretical background to project management, based on the current standard of the Project Management Institute (PMI), that will allow participants to access the first level of Project Management certification (CAPM). This theoretical knowledge about Project Management applied to the data science industry will be integrated by a practical knowledge about software that can helps the IT Manager / CTO / CIO etc. in managing projects.

Lecturer	Andrea Molinari
Contact	E3.10, andrea.molinari@unibz.it, +39 0471 013278 / 013279
Scientific sector of lecturer	ING-INF/05
Teaching language	English
Office hours	Before or after the lecture: arrange beforehand by email.
Lecturing Assistant (if any)	Damiano Somenzi
Contact LA	POS 1.04, Damiano.Somenzi@unibz.it
Office hours LA	Before or after the lecture: arrange beforehand by email.
List of topics	• Data centric software engineering • Methods and practices for data product development • Domain engineering for data product • Managing the software quality of data products • Methods for testing • Maintenance of software Specific topics: • What is a data product • Recap of different approaches to data products' creation • Specific methods and practices for data product development • Project Management fundamentals, theories, and practices for data products projects • Data Science Project Management • Using software tools to manage the entire lifecycle of a data product
Teaching format	Frontal lectures using slides produced by the teacher, exercises in lab, projects.

Learning outcomes	Knowledge and understanding: • D1.3 - Knowledge of principles, methods, and techniques for managing the entire lifecycle of a project for data product development, from business scope definition to budgeting to delivery, and understanding of the challenges in this field Applying knowledge and understanding: • D2.1 - Practical application and evaluation of tools and techniques in the field of data science • D2.4 - Ability to follow the complete life cycle of the creation of a data product, applying the appropriate methodology to the respective phase of the data product life cycle • D2.8 - Practical knowledge about project management techniques for data products' creation Making judgments • D3.1 - Ability to plan, manage, reschedule and adapt to changing conditions of a data product development • D3.2 - Ability to select the appropriate sources of information needed for a given action • D3.3 - Ability to define milestones, constraints, deadlines, and budget of a project, and the respective resources needed to achieve these objectives Communication skills

- D4.1 - Ability to use English at an advanced level with particular reference to disciplinary terminology.
- D4.3 - Ability to structure and draft scientific and technical documentation

Learning skills
- D5.1 - Ability to autonomously extend the knowledge acquired during the study course
- D5.3 - Ability to deal with problems systematically and creatively and to appropriate problem-solving techniques

Assessment	Lab project presented all its planning and delivering details during the oral exam: • The student will develop a lab project for the development of a data product, following the whole life cycle: conception, discussion with stakeholders and sponsors (simulated by the lecturers), planning, executing and monitoring the project, and finally reporting the end of the project. This data product will therefore be developed during the course and the whole project management activities related to the data product will be presented during the oral exam.
Assessment language	English
Assessment Typology	Monocratic
Evaluation criteria and criteria for awarding marks	The final mark is 100% related to the evaluation of the project of data product development. The whole life cycle, from the initiation phase to the closing phase will be evaluated during the oral exam. Evaluation criteria follow. • effective application of project management techniques to the data product development; • good presentation and relational capabilities of interacting with the different stakeholders of the project • Qualities of the applied development process and the delivered data product. • Effective usage of data science project management techniques seen during the course • degree of understanding of theoretical aspects related to data product development; • ability to put into place theoretical aspects and select appropriate project management tools and techniques to the final data product; • effectiveness and quality of project management choices made during the life cycle of the project for the creation of the target data product

Required readings	No required readings are provided, but the course heavily relies on previous academic-level knowledge on a) data science b) software engineering. For those participants with lacks in these topics, adequate readings will be suggested by the lecturer during the course on a personalized base. Refer to the lecturer for any clarification Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it

Supplementary readings	Supplementary readings will be provided during the course.
Software used	The student will develop the data product using any preferred tool/language/library s/he is comfortable with. The only new software that will be used is Microsoft Project, to support the student in conceiving, planning, executing, monitoring, and reporting the whole life cycle of the data product development, including scope, time and budget definition, project analytics, risk management and all communication and stakeholders management of a data product. Microsoft Project is available for all UniBZ students or can be accessed through the Vmware Horizon Client.