COURSE DETAILS

"BIOINFORMATICS AND DATA SCIENCE"

DEGREE PROGRAMME: AGRI-ENVIRONMENTAL AND FOOD BIOTECHNOLOGY
ACADEMIC YEAR 2021-2022

GENERAL INFORMATION – TEACHER REFERENCES

TEACHER: EDOARDO PASOLLI
PHONE: 081 2539397
EMAIL: EDOARDO.PASOLLI @UNINA.IT

GENERAL INFORMATION ABOUT THE COURSE

INTEGRATED COURSE (IF APPLICABLE): BIOINFORMATICS AND DATA SCIENCE (U3300)
MODULE (IF APPLICABLE): DATA SCIENCE (U3302)
CHANNEL (IF APPLICABLE):
YEAR OF THE DEGREE PROGRAMME: I
SEMESTER: II
CFU: 6

REQUIRED PRELIMINARY COURSES (IF MENTIONED IN THE COURSE STRUCTURE "ORDINAMENTO")

None

PREREQUISITES (IF APPLICABLE)

None

LEARNING GOALS

The integrated course "Bioinformatics and Data Science" aims to provide the theoretical foundations and knowledge on the main methods of analysis of biosequences and scientific data; outline the mathematical/computational principles and the methodological basis on which the main tools for data analysis are based. The course aims to provide students with theoretical notions in addition to practical exercises in the laboratory.

EXPECTED LEARNING OUTCOMES (DUBLIN DESCRIPTORS)

Knowledge and understanding

At the end of the Bioinformatics module the student:
1. will consolidate the basic principles and concepts of molecular biology;
2. will have knowledge of the transversal and interdisciplinary nature of bioinformatics and will know the main areas of application of bioinformatics;
3. will know how the data are stored and organized;
4. will know the methodological bases of the main basic bioinformatics tools;
5. will be able to independently interpret the results obtained by applying basic bioinformatics tools.

At the same time, at the end of the Data Science module, the student:
1. will acquire the basic knowledge and skills for data analysis;
2. must demonstrate knowledge and use of statistical approaches;
3. must be able to select and implement IT tools for data analysis;
4. will be able to develop and apply machine learning techniques.

At the end of the course the student will have also to demonstrate that they are able to apply the knowledge, skills and competences provided for in this syllabus. The student must also be able to obtain information or broaden their knowledge by independently drawing on texts, scientific articles belonging to the bioinformatics and data science sector.

Applying knowledge and understanding

The Bioinformatics module aims to provide skills so that the student autonomously interrogates the main databases of biological interest and use the main bioinformatics tools for the analysis of biosequences.
At the end of the course the student will be able to:
1. query databases of scientific literature on the topics of interest and adopt a research-oriented attitude;
2. query primary and secondary databases of biosequences and extract useful data and information;
3. use the most popular bioinformatics tools for the analysis and comparison between biosequences;
4. assess the degree of reliability of the bioinformatic evidence collected;
5. solve analytical problems of increasing complexity in the fields of basic bioinformatics and molecular biology.

The Data Science module aims to provide skills so that the student can extract knowledge from large dataset (big data) using machine learning and data mining techniques.
At the end of the course the student will be able to:
1. develop simple computer programs in R environment;
2. solve analytical problems of increasing complexity in the field of Data Science.

COURSE CONTENT/SYLLABUS

The course is divided into two modules: BIOINFORMATICS (SSD BIO11/molecular biology) and DATA SCIENCE (SSD ING-INF/03)

DATA SCIENCE

1. **Theory classes** (3 CFU)
   - ✓ Introduction to data science for biological systems; Data acquisition and pre-processing; Data exploration and visualization (1 CFU)
   - ✓ Machine learning algorithms (1 CFU)
   - ✓ Hidde Markow model; The big data era; Final quiz and discussion (1 CFU)
2. **Practice classes** (3 CFU)
   - ✓ Introduction to R & The first commands; How to build a Markdown & Data reading and writing; Subsetting & Package dplyr (1 CFU)
   - ✓ Control structures & Package ggplot2; Principal component analysis & Multidimensional scaling (1 CFU)
   - ✓ K-means clustering & Hierarchical clustering; Package caret & Supervised classification; Linear and non-linear regression (1 CFU)

READINGS/BIBLIOGRAPHY

1. Course slides
2. “R programming for data science”, Roger Peng, 2016
3. “Exploratory data analysis with R”, Roger Peng, 2016

TEACHING METHODS

The course alternates between a) lecturers for about 50% of the total hours, b) practical sessions (hands on) in order to deepen the theoretical aspects, verify what is reported during the lecturers and strengthen and "fix" the knowledge (remaining 50% of total hours).

EXAMINATION/EVALUATION CRITERIA

a) Exam type:

Exam type	
written and oral	
only written	X
only oral	
project discussion	
Other	

In case of a written exam, questions refer to:	Multiple choice answers	X
	Open answers	X
	Numerical exercises	X

b) Evaluation pattern:

In the case of multiple choice questions, the number and correctness of the answers will be assessed.

The final grade will be weighted on the CFU of each course and therefore composed as follows: BIOINFORMATICA module 6 CFU 50%, DATA SCIENCE module 6 CFU 50%.