Password for secured document : PAR (only unlock if really necessary)

DESCRIPTION/Syllabi of Curricula/Module

Short Name of the University/Country code **Date (Month / Year)**	
TITLE OF THE MODULE	**Code**
Data Science for Bio-Engineering	

Teacher(s)	**Department**
Coordinating: Prof. Gur Yaari **Others:**	Engineering

Study cycle **(BA/MA)**	**Level of the module** **(Semester number)**	**Type of the module** **(compulsary/elective)**
BSc	6	compulsory

Form of delivery **(theory/lab/exercises)**	**Duration** **(weeks/months)**	**Language(s)**
theory/ exercises	12 weeks	Hebrew

Prerequisites	
Prerequisites: Probability and Statistics 1. Experience with programming is helpful, but not assumed	**Co-requisites (if necessary):**

ECTS (Credits of the module)	Total student workload hours	Contact hours	Individual work hours
6	146	48	98

Aim of the module (course unit): competences foreseen by the study programme		
In this course which is designed for advanced undergraduates, we will study the state of the art in big data management. It will focus on algorithms, statistical methods and analysis of data at a scale and efficiency that go well beyond the capabilities of conventional information technologies.		
Learning outcomes of module (course unit)	**aching/learning methods** **(theory, lab, exercises)**	**Assessment methods** **(written exam, oral exam, reports)**
The students at the end of the course should emerge with the following core competencies: 1. Graduates should have a broad understanding and hands-on experience in data analysis. 2. Should know how to use data to draw inferences about data and should be skilled to do that using R and BASH or any other related scripting language for data manipulation, visualization, and analysis.	theory/exercises	HW / midterm quiz / Final project : presentation and submitted work

Themes	Contact work hours	Time and tasks for individual work

	Lectures	Consultations	Seminars	Practiclwork	Laboratorywork	Placements	**Total contact work**	**Individual work**	**Tasks**
Foundations of R: Why use R, basic variables, datasets, functions, apply, R-markdown, visualization, basic plots, dplyr.	9			3			12	12	3
Inference: Random variables, NULL Hypothesis, t tests, Chi square, proportion tests in R, ANOVA, permutation test, bootstrapping	9			3			12	12	3
Regression analysis: Inference using regression models, simple linear regression, logistic regression, multiple linear regression, categorical predictors and Interactions	9			3			12	12	3
Machine Learning – KNN, cross validation, variable selection, hierarchical clustering , k- means, PCA, unsupervised CARET	6			2			8	8	2
Bash – Linux and command line tool	3			1			4	4	2
Final project								50	1
Total	36			12			48	98	14

Assessment strategy	Weight in %	Deadlines	Assessment criteria
HW	15		
midterm quiz	20		
Kahoot!	15		Each lesson starts/ends with an interactive quiz - the student with the highest score gets a bonus
Final Project	50		25- presentation 25- submitted work

Author
Compulsory literature

This course uses Open Source components.

You can find the source material of their open source projects along with license information below.

We acknowledge and are grateful to these developers for their contributions to open source.

Data Analysis for Genomics http://genomicsclass.github.io/book/

Copyright (c) 2013 Rafael Irizarry and Michael Love

License (MIT) https://github.com/genomicsclass/labs/blob/master/LICENSE

Statistical Modeling: A Fresh Approach http://project-mosaic-books.com

Second Edition, Copyright (c) 2012, v2.0 Daniel Kaplan

Data Science Specialization https://github.com/DataScienceSpecialization/courses

Contributors: Brian Caffo,Jeff Leek,Roger Peng,Nick Carchedi,Sean Kross License (CC-NC-SA) https://tldrlegal.com/license/creative-commons-attribution-noncommercial-sharealike-(cc-nc-sa)

Programming in R for Analytics http://www.contrib.andrew.cmu.edu/~achoulde/94842/

Contributors: Alexandra Chouldechova

License (CC BY-NC-SA 4.0) https://creativecommons.org/licenses/by-nc-sa/4.0/