



# *The Hebrew University of Jerusalem*

## *Syllabus*

### *Big Data Mining. - 52002*

*Last update 09-03-2022*

*HU Credits: 3*

*Degree/Cycle: 1st degree (Bachelor)*

*Responsible Department: Statistics*

*Academic year: 0*

*Semester: 2nd Semester*

*Teaching Languages: Hebrew*

*Campus: Mt. Scopus*

*Course/Module Coordinator: Or Zuk*

*Coordinator Email: [or.zuk@mail.huji.ac.il](mailto:or.zuk@mail.huji.ac.il)*

*Coordinator Office Hours: Wed. 16:15-17:15*

*Teaching Staff:*

---

Dr. Or Zuk

Course/Module description:

*We will learn methods for analyzing big datasets*

Course/Module aims:

*Acquiring statistical and computational tools for performing statistics on large-scale data*

Learning outcomes - On successful completion of this module, students should be able to:

*Analyze datasets with millions of records and thousands of variables. Use in an efficient manner programs with parallel/cloud computing. Extract data from the web.*

Attendance requirements(%):

0

*Teaching arrangement and method of instruction: Lectures, hands-on examples on the computer*

Course/Module Content:

*Working remotely in a unix environment/cloud computing, SQL, acquire data from the web.*

*Finding similarities: Hash functions, nearest neighbours*

*Distributed computing in a cloud environment*

*Analyzing network data: finding communities, sampling large graphs*

*Streaming data: online algorithms, A/B testing*

Required Reading:

None

Additional Reading Material:

*Leskovec, Rajaraman&Ullman (2014). Mining of massive datasets, Cambridge University Press*

*Tan, Steinbach, Karpatne and Kumar (2005). Introduction to Data Mining. Pearson Addison Wesley*

---

*Liu (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications). Springer*

*White (2015). Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale. O'Reilly Media*

*Course/Module evaluation:*

*End of year written/oral examination 0 %*

*Presentation 0 %*

*Participation in Tutorials 0 %*

*Project work 75 %*

*Assignments 25 %*

*Reports 0 %*

*Research project 0 %*

*Quizzes 0 %*

*Other 0 %*

*Additional information:*

*There will be a mid-term project during the semester that will comprise 25% of the course grade.*

*After the end of the semester there will be given a final project that will comprise 75% of the course grade.*

*In addition, there will be a few exercises for self-practice (not for grade)*