

SYLLABUS

MLE5257 Big Data Analysis

University year 2025-2026

1. Information regarding the programme

1.1. Higher education institution	Babeş-Bolyai University
1.2. Faculty	Faculty of Psychology and Educational Sciences
1.3. Department	Department of Psychology
1.4. Field of study	Psychology – Cognitive Sciences
1.5. Study cycle	Bachelor level
1.6. Study programme/Qualification	Psychologist
1.7. Form of education	Full-time

2. Information regarding the discipline

2.1. Name of the discipline	Big Data Analysis			Discipline code	MLE5257
2.2. Course coordinator*	Ioana-Georgiana Ciuciu, Senior Assistant Professor, PhD Raluca D. Szekely-Copîndean, Lecturer, PhD				
2.3. Seminar coordinator*	Ioana-Georgiana Ciuciu, Senior Assistant Professor, PhD Raluca D. Szekely-Copîndean, Lecturer, PhD				
2.4. Year of study	2	2.5. Semester	3	2.6. Type of evaluation	E
				2.7. Discipline regime	DS
*Courses and Seminars have a multidisciplinary approach; both instructors are in charge of courses and seminars.					

3. Total estimated time (hours/semester of didactic activities)

3.1. Hours per week	3	of which: 3.2 course	2	3.3 seminar/laboratory	1
3.4. Total hours in the curriculum	42	of which: 3.5 course	28	3.6 seminar/laboratory	14
Time allotment for individual study (ID) and self-study activities (SA)					hours
3.5.1. Learning using manual, course support, bibliography, course notes (SA)					28
3.5.2. Additional documentation (in libraries, on electronic platforms, field documentation)					15
3.5.3. Preparation for seminars/labs, homework, papers, portfolios and essays					10
3.5.4. Tutorship					2
3.5.5. Evaluations					2
3.5.6. Other activities: <i>research activities</i>					1
3.7. Total individual study hours					58
3.8. Total hours per semester					100
3.9. Number of ECTS credits					4

4. Prerequisites (if necessary)

4.1. curriculum	Quantitative Research Methods and Data Analysis (I and II) Algorithms and Programming Cognitive Psychology Cognitive Neuroscience Advanced Data Science
4.2. competencies	Basic programming skills Basic skills in data analysis and knowledge of main research methods in cognitive science Basic knowledge of main models and paradigms in cognitive science

5. Conditions (if necessary)

5.1. for the course	Classroom
---------------------	-----------

	Video projection system (PC/laptop, video projector) Audio system
5.2. for the seminar /lab activities	Classroom Video projection system (PC/laptop, video projector) Audio system

6. Specific competencies acquired

Professional/essential competencies	<p>Knowledge and understanding</p> <ul style="list-style-type: none"> • Understanding of Big Data particularities in the context of cognitive science research • Knowledge of main sources of Big Data for cognitive science research and main procedures of processing Big Data <p>Explaining and interpretation</p> <ul style="list-style-type: none"> • Explaining specific steps in a Big Data processing and analysis pipeline • Interpretation of data analysis results <p>Instrumental - applicative</p> <ul style="list-style-type: none"> • Developing skills for advanced querying over distributed information resources • Developing skills for processing, evaluating, testing and validating real-world data <p>Attitude</p> <ul style="list-style-type: none"> • Interest in continuous personal development in programming and statistical training • Openness to contributions of Big Data in the field of Cognitive Science
Transversal competencies	<ul style="list-style-type: none"> • Working with (Big) Data in a transparent, reproducible and ethical manner • Developing Data Science and Data Engineering basic competencies • Asking appropriate questions both regarding research objectives and data processing objectives

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	Familiarizing students with handling and analysing large amounts of data specific to cognitive science
7.2 Specific objective of the discipline	<ul style="list-style-type: none"> • Introducing students to Big Data architectures and methodologies to retrieve, process and analyse large amounts of data • Understanding processing and analysis challenges of different types of Big Data in Cognitive Science • Implementing processing and analysis pipelines using Big Data

8. Content

8.1 Course	Teaching methods	Remarks
Course 1: Fundamentals of Data Science and Big Data – part 1	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 2: Fundamentals of Data Science and Big Data – part 2	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 3: Big Data Pipeline	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 4: Data Systems and the Lambda Architecture for Big Data – part 1	lecture, demonstrative example, knowledge synthesis, inductive method	

Course 5: Data Systems and the Lambda Architecture for Big Data – part 2	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 6: Big Data Storage and Processing – Batch Layer	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 7: Big Data Storage and Processing – Speed Layer	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 8: Particularities of Human Big Data	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 9: Big Data Quality in Cognitive Science	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 10: Feature Engineering and Modelling Cognition	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 11: Sensor-Based Cognition	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 12: Working with Big Brain Data	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 13: Extracting Meaning from Big Text Data	lecture, demonstrative example, knowledge synthesis, inductive method	
Course 14: Ethical Challenges of Big Data	lecture, demonstrative example, knowledge synthesis, inductive method	
8.2 Seminar / laboratory	Teaching methods	Remarks
Seminar 1: The Data Science Process in the context of Big Data Applications	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 2: Tutorial – API Interrogation for Data Science	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 3: Tutorial – MongoDB-part 1	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 4: Tutorial – MongoDB-part 2	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 5: Tutorial – Data Ingestion and Cleaning – part 1	conceptual clarification, case studies, guided discovery, group activities	
Seminar 6: Tutorial – Data Ingestion and Cleaning – part 2	conceptual clarification, case studies, guided discovery, group activities	
Seminar 7: Data Processing using MapReduce	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 8: Psychometric Challenges of Big Data	presentation, knowledge synthesis, conceptual	

	clarification, case studies, guided discovery, group activities	
Seminar 9: Drawing Insight from Big Team Science Data	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 10: Basic Concepts of Machine Learning	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 11: Physiological Data Processing Tutorial	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 12: Reproducibility Challenges of fMRI Data Processing Pipelines	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 13: Elements of Natural Language Processing	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	
Seminar 14: Ethics, Privacy and Mental-State Inference	presentation, knowledge synthesis, conceptual clarification, case studies, guided discovery, group activities	

Bibliography

Part 1:

- Banker, K., Bakkum, P., Verch, S., Garrett, D., & Hawkins, T. (2016). *MongoDB in action* (2nd ed.). Manning Publications.
- Borucki, A. (2024). *MongoDB in action* (3rd ed.). Manning Publications.
- Cielen, D., Meysman, A. D. B., & Ali, M. (2016). *Introducing data science: Big data, machine learning, and more, using Python tools*. Manning Publications.
- Damji, J. S., Wenig, B., Das, T., & Lee, D. (2020). *Learning Spark*. O'Reilly Media.
- Grigorev, A. (2021). *Machine learning bookcamp*. Manning Publications.
- Grus, J. (2019). *Data science from scratch: First principles with Python*. O'Reilly Media.
- Holmes, A. (2015). *Hadoop in practice* (2nd ed.). Manning Publications.
- Khalil, M. (2025). *Effective data analysis*. Manning Publications.
- Marz, N., & Warren, J. (2015). *Big data: Principles and best practices of scalable real-time systems*. Manning Publications.
- Perrin, J. G. (2020). *Spark in action* (2nd ed.). Manning Publications.
- Rioux, J. (2022). *Data analysis with Python and PySpark*. Manning Publications.
- Sadalage, P., & Fowler, M. (2013). *NoSQL distilled: A brief guide to the emerging world of polyglot persistence*. Pearson Education.
- White, T. (2009). *Hadoop: The definitive guide*. O'Reilly Media.
- Zečević, P., & Bonači, M. (2017). *Spark in action*. Manning Publications.
- Zelenin, A., & Kropp, A. (2025). *Apache Kafka in action*. Manning Publications.

Part 2:

- Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, 73, 899–917. <https://doi.org/10.1037/amp0000190>

Bainbridge, W. A. (2025). *Big data in the psychological sciences*. Cambridge University Press.

Coviello, L., Sohn, Y., Kramer, A. D. I., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PLoS One*, 9(3), e90315.
<https://doi.org/10.1371/journal.pone.0090315>

Coles, N. A., Perz, B., Behnke, M., Eichstaedt, J. C., Kim, S. H., Vu, T. N., Raman, C., Tejada, J., Huynh, V.-T., Zhang, G., Cui, T., Podder, S., Chavda, R., Pandey, S., Upadhyay, A., Padilla-Buritica, J. I., Barrera Causil, C. J., Ji, L., Dollack, F., ... Saganowski, S. (2025). Big team science reveals promises and limitations of machine learning efforts to model physiological markers of affective experience. *Royal Society Open Science*, 12(6). <https://doi.org/10.1098/rsos.241778>

Jones, M. N. (Ed.). (2017). *Big data in cognitive science*. Routledge.

Lin, Q., Chan, S. H. M., & Chan, D. (2018). Big data in social and psychological science: Theoretical and methodological issues. *Journal of Computational Social Science*, 1(1), 59–66.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4), 1689–1696.
<https://doi.org/10.3758/s13428-020-01516-y>

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncavles, M., Jwa, A., & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *eLife*, 10.
<https://doi.org/10.7554/eLife.71774>

Poldrack, R. A., Markiewicz, C. J., Appelhoff, S., Ashar, Y. K., Auer, T., Baillet, S., Bansal, S., Beltrachini, L., Benar, C. G., Bertazzoli, G., Bhogawar, S., Blair, R. W., Bortoletto, M., Boudreau, M., Brooks, T. L., Calhoun, V. D., Castelli, F. M., Clement, P., Cohen, A. L., ... Gorgolewski, K. J. (2024). The past, present, and future of the brain imaging data structure (BIDS). *Imaging Neuroscience*, 2. https://doi.org/10.1162/imag_a_00103

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjeh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences*, 121(34).
<https://doi.org/10.1073/pnas.2308950121>

Vezzoli, M., & Zogmaister, C. (2023). An introductory guide for conducting psychological research with big data. *Psychological Methods*, 28(3), 580–599. <https://doi.org/10.1037/met0000513>

Woo, S. E., Tay, L. E., & Proctor, R. W. (2020). *Big data in psychological research*. American Psychological Association.

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

The proposed lecture and seminar offer central topics in fundamental and applied research in the fields of cognitive sciences, and their approach is based on the most recent results found in the literature. The course also offers state of the art research skills that are transferable to any scientific and applied field of knowledge.

Synergies with various local and EU initiatives: local industry, European Data Science Academy (EDSA, <https://edsa-project.eu/>), EU projects such as LETHE (<https://cordis.europa.eu/project/id/101017405>), FARE (<https://cordis.europa.eu/project/id/853566>), the Human Brain Project (<https://www.humanbrainproject.eu/en/>), SoBigData (<http://project.sobigdata.eu/>), etc.

Synergies with the Big Data course from the Faculty of Mathematics and Computer Science (possible collaboration between the Computer Science students and the Psychology students, where the former could play the role of data engineers and the latter could play the role of data consumers or even data analysts).

10. Evaluation

Activity type	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Percentage of final grade
---------------	--------------------------	-------------------------	--------------------------------

10.4 Course 10.5 Seminar/laboratory	Demonstrating understanding of core concepts and principles regarding Big Data analytics applied to Cognitive Science. Evidence of Big Data-related critical thinking and problem-solving. Demonstrating ability of approaching Big Data challenges from Cognitive Science, implementing queries, analyses and drawing information relevant to Cognitive Science use cases.	Written examination	100%
10.6 Minimum standard of performance			
A minimum of 50% in the written examination.			

11. Labels SDG (Sustainable Development Goals)

	General label for Sustainable Development							
								
								

Date:
October 2025

Signature of course coordinator

Ioana-Georgiana Ciuciu

Signature of seminar coordinator

Ioana-Georgiana Ciuciu

Raluca D. Szekely-Copindean

Raluca D. Szekely-Copindean

Date of approval:

Signature of the head of department

.....