



Subject card

| | | | | | | | |
|---|--|--|---|-------------------------------------|--|------------|-----|
| Subject name and code | BIG DATA , PG_00060930 | | | | | | |
| Field of study | Economic Analytics | | | | | | |
| Date of commencement of studies | October 2023 | | Academic year of realisation of subject | | 2024/2025 | | |
| Education level | second-cycle studies | | Subject group | | Optional subject group Subject group related to scientific research in the field of study | | |
| Mode of study | Part-time studies (on-line) | | Mode of delivery | | at the university | | |
| Year of study | 2 | | Language of instruction | | Polish | | |
| Semester of study | 3 | | ECTS credits | | 4.0 | | |
| Learning profile | general academic profile | | Assessment form | | exam | | |
| Conducting unit | Katedra Statystyki i Ekonometrii -> Faculty of Management and Economics | | | | | | |
| Name and surname of lecturer (lecturers) | Subject supervisor | | dr hab. Michał Pietrzak | | | | |
| | Teachers | | dr hab. Michał Pietrzak | | | | |
| Lesson types and methods of instruction | Lesson type | Lecture | Tutorial | Laboratory | Project | Seminar | SUM |
| | Number of study hours | 8.0 | 0.0 | 24.0 | 0.0 | 0.0 | 32 |
| | E-learning hours included: 0.0 | | | | | | |
| Learning activity and number of study hours | Learning activity | Participation in didactic classes included in study plan | | Participation in consultation hours | | Self-study | SUM |
| | Number of study hours | 32 | | 8.0 | | 60.0 | 100 |
| Subject objectives | Demonstrates in-depth comprehensive preparation for the analysis of large data sets | | | | | | |
| Learning outcomes | Course outcome | | Subject outcome | | Method of verification | | |
| | [K7_W03] demonstrates in-depth preparation in the application of analytical methods and techniques for formulating and solving problems | | uses advanced technologies to handle large data sets, preparing them to solve complex problems | | [SW1] Assessment of factual knowledge | | |
| | [K7_U01] creates innovative solutions to complex and unstructured problems, taking into account the variability of the environment by synthesising information from many sources | | formulates innovative solutions based on large data sets using in-depth machine learning algorithms | | [SU4] Assessment of ability to use methods and tools | | |
| Subject contents | Overview of Big Data. Types of Digital Data, Introduction to Big Data Big data programming tools (e.g., Hadoop, MongoDB, Spark, etc.). Using Spark with R Big data extraction and integration Big data storage; Technologies for Handling Big Data Introduction to Hadoop HDFS (Hadoop Distributed File System) Dig Deep to understand the fundamental of MapReduce and HBase Hadoop MapReduce in R; Integrating Hadoop and R RHIPE; RHadoop Data Analytics with R and Hadoop data preprocessing, visualising data Big Data Analysis and Machine Learning supervised and unsupervised ML algorithms. Spark Machine Learning with R Importing and exporting data from various DBs (RMySQL, RSQLite, RHive, RHBase). Using SparkSQL with R Big Data Analytics with BigR Deep learning algorithms with R & H2O | | | | | | |
| Prerequisites and co-requisites | | | | | | | |
| Assessment methods and criteria | Subject passing criteria | | Passing threshold | | Percentage of the final grade | | |
| | Exam | | 60.0% | | 60.0% | | |
| | Test | | 60.0% | | 40.0% | | |

| | | |
|--|--------------------------|--|
| Recommended reading | Basic literature | Hamstra, M., & Zaharia, M. (2013). Learning Spark: lightning-fast big data analytics. O'Reilly & Associates Densmore, J. (2021). Data pipelines pocket reference. O'Reilly Media Drabas, T., & Lee, D. (2017). Learning PySpark. Packt Publishing Ltd Haines, S. (2022). Modern Data Engineering with Apache Spark: A Hands-on Guide for Building Mission-critical Streaming Applications. Apress |
| | Supplementary literature | Warren, J., & Marz, N. (2015). Big Data: Principles and best practices of scalable realtime data systems. Simon and Schuster Ilijason, R. (2020). Beginning Apache Spark Using Azure Databricks: Unleashing Large Cluster Analytics in the Cloud. Apress |
| | eResources addresses | Adresy na platformie eNauczanie: |
| Example issues/ example questions/ tasks being completed | | |
| Work placement | Not applicable | |