Business Intelligence

General data

Course code:	
ECTS credits:	6
Type of the course:	General core course (A)
Semester:	Fall Semester 2
Course restrictions:	-
Course leader (with availabilities):	*Ferenc KRUZSLICZ, Associate Professor* *+ 36 72 501 599/ 23113* *kruzslic@ktk.pte.hu* *office: B122*
Further lecturer(s) (with availabilities):	-

1. Description and aims

This course provides an introduction to the concepts of decision support information systems focusing on components and functionality of business intelligence (BI). It explores how business problems can be solved effectively by using operational data to create and use data warehouses, and then applying data mining tools and analytics combined with machine learning (ML) to gain new insights into organizational operations. Detailed discussion of the analysis, design and implementation of systems for BI, including the different types of visual and automatic reporting and analytics. Wide range of BI related real world business problems and approaches are investigated in order to introduce their standard solution schema. Also methods and techniques are demonstrated how the effectiveness of such various solutions can be measured. The major topics covered like enterprise data warehousing, big data and data/text mining helps to understand the new challenges in data, information and knowledge management. Theoretical knowledge is applied and tested in practice using BI software tools and modeling. Case studies are used to explore the use of the CRISP-DM approach, the success and limitations of BI application, as well as their technical, social and ethical issues.

2. Intended Learning Outcomes (ILOs)

Upon the successful completion of this course, students should be able to:

CILO 1. know the basic concepts and methods of data-driven decision support required mid- and high-level corporate decision-making; recognize the types, requirements, possibilities and techniques of machine learning based data analysis related to corporate performance measurement and human resources. (PILO1)

CILO 2. recall the standards of BI applications and data analytic processes; identify the information technological and organizational needs and resources; understand the requirements and evaluation methods of business application of machine learning; follow the current research results and publications. (PILO1)

CILO 3. participate in BI projects; review the quality of BI applications for businesses; estimate the validity and up-to-dateness of implemented BI solutions; contribute effectively to data preparation and cleansing processes. (PILO3, PILO4)

CILO 4. follow the development of data warehouse and data/text mining technologies; independently master the use of the BI software tools and their functions; solve specific BI tasks according to the standards (PILO6)

CILO 5. professionally mediate between the business and the analyst side; effectively manage or participate in BI projects; identify and categorize BI related problems; formulate meaningful conclusions and proposals of BI problems utilizing relevant methodologies. (PILO5)

CILO 6. design and create data-driven decision processes; compile and build machine learning models; interpret and qualify the results obtained and determine their performance. (PILO4, PILO5, PILO6)

CILO 7. take into consideration data protection and ethical aspects of BI solutions; plan compliance programs to respond to the continually changing BI risks and expectations. (PILO7)

CILO 8. support their opinion with professional arguments, confidence and responsibility; constantly follow the development of BI methods; control the corresponding processes with consequences taken into account. (PILO8)

3. Content, schedule

1. Introduction: BI and Big Data
   (RA: BInDM chapter 1.1, 2.2 & PAnDM chapter 13)
2. Data Warehouses, OLAP operations, KPIs, Balanced Scorecards
   (RA: BInDM chapter 1.2, 13.2, 3 & PAnDM chapter 1)
3. Data Exploration and Visualization, CRISP-DM
   (RA:BInDM chapter 1.4, 2.1, 4.4 & PAnDM chapter 2)
4. Data Preprocessing, Data Quality, Data Cleansing
   (RA: BInDM chapter 4.1, 13.1 & PAnDM chapter 3)
5. Data Analytics, Data Mining Techniques, Lazy and Eager Methods
   (RA: BInDM chapter 1.3, 4.3 & PAnDM chapter 4.4)
6. Classification Model, Inductive Learning, Decision Tree Classifiers
   (RA: BInDM chapter 5 & PAnDM chapter 4.1)
7. Classification Evaluation, Validation, Performance Metrics
   (RA: BInDM chapter 4.2 & PAnDM chapter 8, 4.2)
8. Statistical and Machine Learning, Naive Bayes and Artificial Neural Networks
   (RA: BInDM chapter 7 & PAnDM chapter 4.5, 4.6)
9. Association analysis, Frequent Itemset and Association Rule Generation
   (RA: BInDM chapter 9 & PAnDM chapter 6)
10. Segmentation, Prototype Based and Hierarchical Methods
    (RA: BInDM chapter 8 & PAnDM chapter 7.1)
11. Segmentation Evaluation, Density Based, DBSCAN and SOM methods
    (RA: BInDM chapter 6 & PAnDM chapter 7.2, 7.3)
12. Text Mining, Web Mining, Sentiment Analysis
    (RA: BInDM chapter 10, 11 & PAnDM chapter 9)

4. Learning and teaching strategy, methodology

Principal teaching methodologies: Lectures and Case studies

Theoretical foundations strictly follow the text book. Reading assessments are short quizzes used for measuring the preparation level. Presentations and other supplementary course materials are provided. (CILOs 1,2,7,8) In class the selected topic is discussed and small scale illustrative examples and exercises are solved. (CILOs 3,4,5) The corresponding computer lab exercises are simulated data mining projects of different fields and real data sets. Students have to read the case study and do the basic data exploratory steps in advance, and group work is allowed. (CILOs 5,6,7) Along with the course progresses less detailed instructions are given and the data analytic goals are getting more ambiguous. Midterm exams based on similar exam cases, where students have to understand and build a working model on computer, evaluate it, and interpret the results. (CILOs 4,5,8)

5. Assessment

Formative assessment elements:

Learning progression is measured and visualized with a "robo-dog" gamification tool. Correct answers of RAs are revealed after evaluation referring to the appropriate chapter of the core materials. Complete clues are given for weekly group works how the problems was modeled, but with the final exact solution, which must be reproduced. Students are involved in course and faculty level competitions, where the results can be compared to the others in ranking lists.

Summative assessment elements:

Individual Assessment	100%	**Group Assessment**	0%

Name of the element	**Weight**	**Type**	**Details**	**Retake opportunity**	**Req.***	**Related CILOs**
Reading assessment	20%	individual written	weekly quizzes	no	no	1,2,7
Midterm exam	40%	individual written	computer exercise	yes	yes	3,4,6
Final exam	40%	individual written	essay questions	yes	yes	1,2,5,7,8
Home work bonus	*(20%)*	*optional group work*	*computer task*	*no*	*no*	*4,5,6*
Competition bonus	*(20%)*	*optional individual*	*modeling tasks*	*no*	*no*	*3,4,5,6*

** Req.: Completion of the element is required to pass the course, irrespective of the performance in other elements.*

6. Learning materials

- Essential
  - —Anil Maheshwari: Business Intelligence and Data Mining (Big Data and Business Analytics) Business Expert Press 2014, (1st Edition) ISBN: 978-1631571206, pp. 180
  - —Bala Deshpande, Vijay Kotu: Predictive Analytics and Data Mining (Concepts and Practice with RapidMiner) Morgan Kaufmann Publishing 2014, (1st Edition) ISBN: 978-0128014608, pp. 446

- Recommended
  - —Ramesh Sharda, Dursun Delen, Efraim Turban: Business Intelligence, Analytics, and Data Science: A Managerial Perspective, Pearson 2017, (4th Edition) ISBN: 978-0134633282, pp.512.
  - —Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Anuj Karpatne: Introduction to Data Mining, Pearson 2012, (2nd Edition) ISBN: 978-0-13-312890-1, pp. 792
  - —Markus Hofmann, Ralf Klinkenberg: RapidMiner: Data Mining Use Cases and Business Analytics Applications, Chapman & Hall/CRC, 2013 ISBN: 1482205491, pp. 525

7. Further information

International aspects embedded with the course
Data type localization problems are discussed as a part of data cleansing and integration. Text mining is clearly language dependent which requires different NLP tools. Course competitions are shared across different other programmes.
Ethics, Responsibility & Sustainability (ERS) aspects embedded with the course
Explainable artificial intelligence (XAI) is discussed as part of the responsibility of AI systems. Privacy issues are considered as the ethical aspects of data analysis, anonymization methods are used to demonstrate statistical data security.
Connections to the world of practice of the course
Case studies are based on real world problems, only company data is obfuscated and reduced to a manageable size. Students are encouraged to develop their own BI solutions projects.