DOI: https://doi.org/10.53052/9788367652018.13

Iva KOSTADINOVA[1], Vasyl MARTSENYUK[2] , Georgi DIMITROV[1] , Dejan RANCIC[3], Oleksiy BYCHKOV[4], Eugenia KOVATCHEVA[1], Vasil TOTEV[1]

WYTYCZNE DO SZKOLENIA PRAKTYCZNEGO W BIG DATA: PRZEWODNIK DLA STUDENTÓW A4.2 DO PROJEKTU IBIGWORLD

Streszczenie: Praca przedstawia wyniki projektu Erasmus+ „Innovations for Big Data in a Real World" (iBIGworld) 2020-1-PL01-KA203-082197 - Output IO4 - Piloting : A.4.2 - wytyczne dla studentów. Niniejszy artykuł zawiera ważne informacje dla pomyślnego ukończenia kursu Uczenie się oparte na kompetencjach w Big Data, zaproponowanego w wyniku pracy międzynarodowego zespołu naukowców nad Projektem iBIGworld. Niniejsza praca zawiera krótki opis struktury programu nauczania oraz inne istotne informacje związane z planowaniem szkolenia. Artykuł skierowany jest do osób szkolących się, które będą uczestniczyć w szkoleniu.

Słowa kluczowe: wytyczne, szkolenia praktyczne, Big Data, studenci, stażyści

GUIDELINES FOR PRACTICE TRAINING IN BIG DATA: STUDENTS GUIDE A4.2 FOR IBIGWORLD PROJECT

Summary: The work presents the result of Erasmus+ Project "Innovations for Big Data in a Real World" (iBIGworld) 2020-1-PL01-KA203-082197 - Output IO4 - Piloting: A.4.2 - students guidelines. This paper contains important information for the successful completion of the Competence based learning course in Big Data, proposed as a result of the work of an international team of scientists on Project iBIGworld. This work contains a brief description of the structure of the curriculum and other relevant information related to training planning. The paper is aimed at trainees who will follow the training course.

Keywords: guidelines, practice training, Big Data, students, trainees

[1] University of Library Studies and Information Technologies (ULSIT), Sofia, Bulgaria: (i.kostadinova, g.dimitrov, e.kovatcheva, v.totev)@unibit.bg
[2] Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: vmartsenyuk@ath.bielsko.pl
[3] University of Niš (UNi), Nis, Serbia: dejan.rancic@elfak.ni.ac.rs
[4] Taras Shevchenko National University of Kyiv (TSNUK), Kiev, Ukraine: oleksiibychkov@knu.ua

Introduction

Lately a lot of attempts have been made for the purpose of the development of student guidelines in the fields of data science and machine learning [1-5]
The given work presents the result of Erasmus+ Project "Innovations for Big Data in a Real World" (iBIGworld) 2020-1-PL01-KA203-082197 - Output IO4 - Piloting : A.4.2 - students guidelines [6].
Big data is a field of science that studies how to analyze, systematically extract information from, and generally work with, large data sets that are otherwise too large or complex for traditional data processing software applications.
Big Data and its analysis techniques are at the center of modern science and business. Extracting valuable knowledge from massive quantities of data is complicated due to the sheer volume of data generated every day.
Deepening the interdisciplinarity in the Big Data domain where Data Mining, Machine Learning, Data Science, and Advanced Analytics play a role as an approach palette to knowledge discovery. The lectures provide an overview of the Knowledge Discovery Paradigm based on Big Data, interdisciplinary links between fields, actors, and processes involved in Analytics, and potential applications, impact, and importance for business digital transformation, Industry 4.0, and Society 5.0.
Accelerating skill-building in Big Data Analytics by applying supervised and unsupervised approaches for regression, classification, clustering, and feature engineering through particular software tools (Orange, Tableau) following the learning-by-doing and project-based methods.
The work is aimed at students who follow the training course. Students should read the guideline carefully so that they can get acquainted in detail with the upcoming commitments. It is the student's responsibility to meet the requirements of the curriculum. The training aims to achieve two primary goals in the learning path of Big Data. The training is competences oriented.

1. Benefits of training in Big Data

The world's technological capacity per capita to store information has doubled every 40 months since the 1980s. The data can be seen in Fig. 1 and Fig. 2.

Relational database management systems and desktop statistical software are very difficult to process and analyse Big Data. Such tasks usually require software that runs in parallel on tens, hundreds or even thousands of servers.

Big data often includes data that exceeds the capacity of traditional software to process it in a timely manner. Challenges to analysing large data sets include data capture, storage, analysis, sharing, transfer, visualization, retrieval, updating and protection.

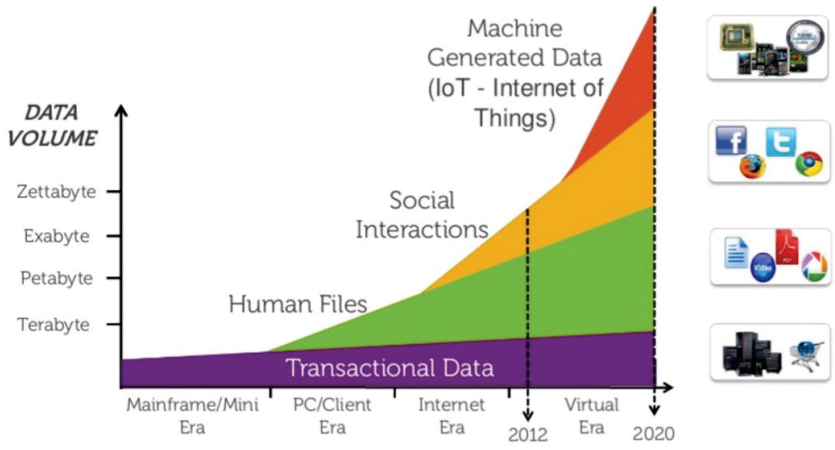

Figure. 1. Growth and digitalization of the global data volume to 2020 by type

Understanding the processes related to Big Data, the skills for processing and analysis of data extracted from Big Data are key to many areas of the global economy. Through Big Data analysis, analyses of processes such as consumer preferences, analyses of the quality of a product placed on the market, analyses and forecasts for the development of companies and many other applications can be made.

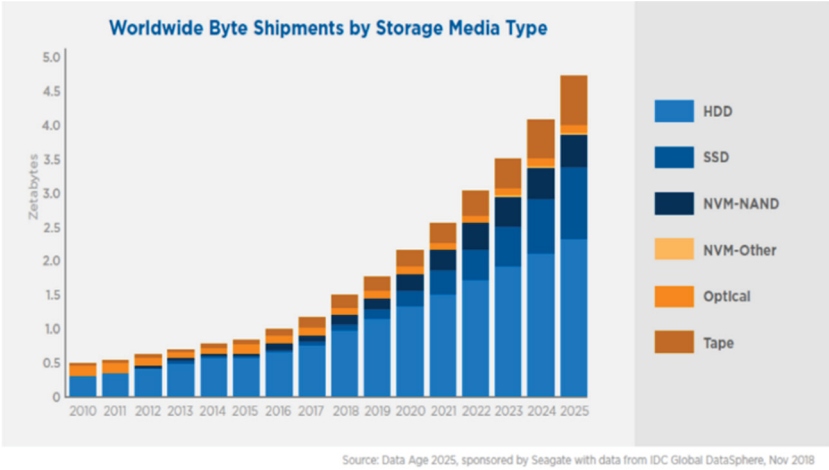

Figure. 2. Growth and digitalization of the global storage capacity to 2025 by type storage

2. General goal for the student training

The main goal of the training course is to form knowledge, skills, and competencies in the interdisciplinary field of data science and to enable students to gain practical experience with real data. Learning framework is based on IEEE guidelines for Big

Data in Machine Learning". The creation of the Big Data programmes is based on the most modern approaches already in practice and well developed in Poland, Bulgarian, Ukrainian and Serbian. The thus created program aims to develop the following aspects:

- digital skills - according to the EC: "the untapped potential of the digital economy and society for the Eastern Partnership was recognised in the Riga EaP Summit Declaration in May 2015 and the EU Foreign Affairs Council Conclusions in November 2016".
- using active, student-oriented aproach: inquiry-, project- and problem-based reconciling country standards and compatibility with the Bologna process. This requires complex interinstitutional work that is not very developed in the target countries.
- interdisciplinary approach as a cross-over between ICt and Economics. Teachers need ot be prepared for this new programme.
- contribution to the efforts of the Digital Single market - help secure some of the benefits of the EU's Digital Single Market (DSM) strategy to the EU's neighbouring countries. there are initiatives already in place: the policy frameworks of the European Neighbourhood Policy (ENP), the Eastern Partnership (EaP), and the Digital Community.
- the enhanced cooperation between participian countries will allow the imrpovement of academic competencies with each other and with more developed EU practices of the respective partners in analyzing the current situation in the region and possibilities for regional development and cooperation.

With this training will be Improve of quality and internationalization higher education of Poland, Bulgarian, Ukrainian and Serbian of the project in terms of education in IT and statistics/ maths and economics.

3. Minimum requirements

The training is suitable for undergraduate and postgraduate students.
Foundational knowledge of Big Data Domain, such as Big Data Strategy, Big Data Characteristics, Big Data Architecture, Big Data Processing, Big Data Value Chain.
Foundational knowledge of Data Mining, such as methods, algorithms, and tasks (classification, regression, clustering), data types (numeric, categorical, data, space), data measurements (nominal, ordinal, interval, ratio), data distribution.
Foundational knowledge of Statistics, such as mean, median, mode, variance, standard deviation, samples, correlations, distributions.
Foundational knowledge in Artificial Intelligence or/and Machine Learning/ Data Science can accelerate the learning process starting directly from Big Data Analytics Tools Overview and practical sessions

4. Course organisation

The organized training was in the form of a summer school - a seminar, where teachers and students met to exchange experience and new knowledge (Table 1, 2).

Table 1. Resource sheet

Resources	Document	Presentation	Video	Guide	Test Questions	Research task	Practical task	LS
Data Analytics Foundation I	x	x	-	-	x	x	-	1
Data Analytics Foundation II	x	x	-	-	x	x	-	1
Analytics Tools Overview I	x	x	-	-	x	x	-	1/2
Analytics Tools Overview II	x	x	-	-	x	x	-	1/2
Practical Session 1			x	x	x		x	1/2
Practical Session 2				x			x	1/2
Practical Session 3			x	x	x		x	1/2
Practical Session 4				x	x		x	1/2

Table 2. Timesheet by Activities

Activities	Lecture	Summary & Discussion	Questions & Feedback	Tasks	Overall
Data Analytics Foundation I	1 h	15 min	10 min	5 min	1h 30 min
Data Analytics Foundation II	1 h	15 min	10 min	5 min	1h 30 min
Analytics Tools Overview I	1 h	15 min	10 min	5 min	1h 30 min
Analytics Tools Overview II	45 min	15 min	10 min	5 min	1h 15 min
Total	3h 45 min	1 h	40 min	20 min	5h 45 min
Practical Session 1					1h 10 min
Practical Session 2					1h
Practical Session 3					1h 10 min
Practical session 4					50 min
Total					4h 10 min

Resources:

- Lectures,
- Presentations,
- Guidelines and Workflows,
- Datasets,
- Videos,
- Lists for additional reading and research.

Practical Sessions (Labs) (Table 3):
Applications of Tools in Analytics (Workflows),
- Session one & two: Orange 3,
- Session three & four: Tableau.

Table 3. Practical Sessions in detail by tasks

Session one:	Orange 3	Installation, Workspace, Data Exploration, and Preparation
Session two:	Orange 3	Regression, Classification, Clustering, Model Evaluation
Session three:	Tableau	Workspace, Visual Analysis and Techniques
Session four:	Tableau	Interactive Dashboards and Stories

LEARNING METHODS
- Learning-by-doing,
- Real-Case and Project-based Learning.

5. Data Analytics Life-Cycle

Life-cycle is a widely used term in all fields relying on a disciplinary approach.
In the current context, the term describes the major stages and activities needed for applying a complete analysis (Fig. 3).

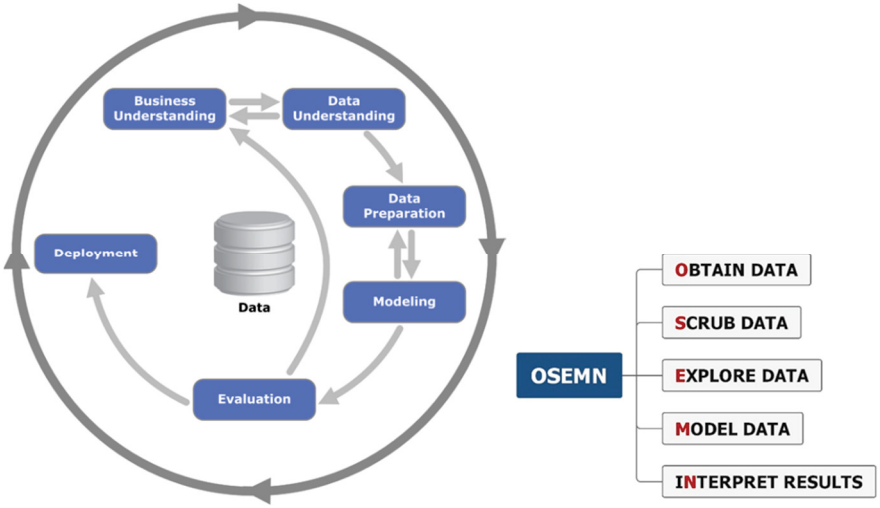

Figure 3. Cross Industry Standard Process for Data Mining (CRISP-DM) & OSEMN Framework

Even more. It is usually a stage of the overall data life-cycle or part of the workflow in all earlier discussed fields from the knowledge discovery paradigm (Big Data, Artificial Intelligence, Machine Learning, Deep Learning, Data Science, Data Mining). We have to get used to thinking of them as a whole scientific spectrum that explores, describe and benefit from Big Data through different perspectives

and convergence of methods to solve a complex data-driven problem. That explains why we can discover typical stages for all data-driven fields.
The following illustrations show and summarize some of the popular methodologies with which you are probably familiar (Fig. 4).

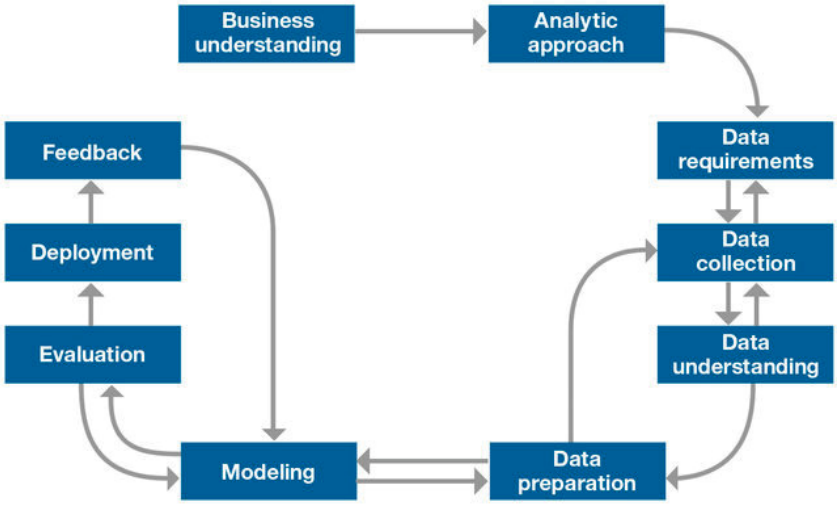

Figure 4. IBM Data Science Methodology

Working on Big Data projects is impossible without team collaboration, especially in the Analytics stage, as it interlinks all processes, depends on them, creates insights and responds for decision-making. Each activity is essential.
More or less, all of the discussed methodologies can be adapted for the purposes of different data-driven projects. Which of them to use depending on the particular analysis goals, knowledge of experts, V's characteristics of data available, and accessible technologies?
Based on the mini-course objective, what is needed to remember is the BDA Life-Cycle and data-intensive project's drivers.

6. Business needs and included analyses in the training

The organization of the training needs to consider four categories of analytics (depending on the workflow and the requirements of data analytics): descriptive, diagnostic, predictive, and prescriptive.
Together, these four types answer everything a business needs: from the current business maturity to what solutions are to be adopted to improve efficiency (Fig. 5).

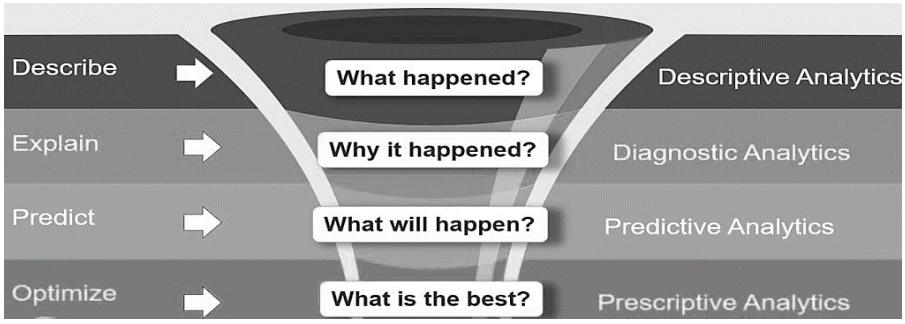

Figure 5. Types of Analytics

5.1. Descriptive analytics

Descriptive analytics is designed to get more information about what happened in the past: Who? What? When? Where? How many? It is the conventional form of analytics and focuses on the depiction of a summary view of facts.
It uses two primary techniques: data aggregation and data mining. Data aggregation is any process in which information is gathered, and expressed in a summary form for purposes such as statistical analysis. The example tools used in this phase are MS Excel, MATLAB, SPSS, STATA. A typical product of descriptive analytics is a company report.
Descriptive analytics is at the bottom of the Big Data value chain, but it can be valuable for uncovering patterns that offer insight. A simple case of descriptive analytics would be reviewing the number of people that visited the company's website over the past few months. Descriptive analytics can be helpful in the sales cycle, for example, to spot seasonal trends and adjust purchasing decisions accordingly.

5.2. Diagnostic analytics

Diagnostic analytics helps to identify why something happened in the past. It takes a deeper look at data to understand the root cause of events but has a limited ability to provide actionable insights. It provides an understanding of causal relationships and sequences while looking backward. Diagnostic is most beneficial for a person concerned with day-to-day operations. For example, to identify why a sales representative has sold considerably fewer items than usual-also, answering why the sales have decreased or increased in a specific year or so. In a social media marketing campaign, diagnostic analytics can determine why certain advertisements increase conversion rates.
Therefore, diagnostic analytics provide valuable insights for organizations because it reveals which decisions impact their performance. It's mostly done with techniques such as drill-down, data discovery, data mining, and correlations.

5.3. Predictive analytics

Predictive analytics is used to predict future outcomes in terms of the probability of an event occurring. It uses Big Data to identify past patterns to predict the future. Predictive algorithms calculate the likelihood of a specific event from trends or

patterns in existing data sets. A model is built on the preliminary descriptive analytics stage to derive the possibility of the outcomes.
Predictive analytics is found in sentiment analytics. All the opinions posted on social media are collected to predict a person's positive, negative, or neutral sentiment on a particular subject. Hence, this can identify which customers should be included in a promotional campaign to maximize response. Some companies apply predictive analytics for sales lead scoring, indicating which incoming sales leads will have the highest chance of converting into actual customers- other examples - weather forecast, plan-failure prediction, travel products recommendation system.
Predictive analytics is realized in machine learning algorithms and statistics. Usually, companies need trained data scientists and machine learning experts for building these models. The most popular tools include Python, R, Rapid Miner, KNIME, Watson.

5.4. Prescriptive analytics

Prescriptive analytics provides the solution for what is predicted to happen in the future. It creates and updates the relationship between action and outcome using a solid feedback system. It did not only help make optimal recommendations during the decision-making process but also helps in mitigating the risk for seeing based on predictive analytics available. Prescriptive analytics can suggest favorable solutions and is the final and most valuable frontier of advanced analytics. It gives a laser-like focus to answer specific questions. For example, in the health care industry, better management of the patient population is possible by using prescriptive analytics to measure the number of patients who are clinically obese, then add filters for factors like diabetes and LDL cholesterol levels to determine where to focus treatment. The same prescriptive model can be applied to almost any industry target group or problem.
Prescriptive analytics is at the nascent stage of implementation, and organizations have not used its full potential.

6. Tools

The following analytics tools are selected in order to accomplish the stated objectives while adhering to the logic of the learning path where the theoretical underpinnings of data mining have been covered.

6.1. Orange3:

Orange is a component-based visual programming open-source tool utilized for data mining, machine learning, data analysis and visualization.
Components of Orange range from basic operations such as data visualization, subset selection, and pre-processing to more complex tasks such as the evaluation of learning algorithms in practice and the development of predictive models. It supports bioinformatics, text, image, and signal processing add-ons, as well as advanced analytics features.

STEPS:
- Installing the software,
- Workspace (canvas) and components (widgets) familiarization,
- Creating a workflow,
- Work with built-in datasets,
- Basic Data Exploration with Orange,
- Feature Statistics,
- Data Preparation,
- Classification,
- Regression,
- Cluster Analysis.

Benefits	Features
- Free open-source - Visual programming - Both no-coding and coding (Python)	- Python 3 data mining library - Interactive Data Visualization - Add-ons Extended Functionality - Manual parameter optimization
Limitations	**Training**
- Not always reliable support - No automatic parameter optimization - Has error measurement but must rebuild model each time	▪ Blog, docs & online community support ▪ Classroom training ▪ Online tutorial and training videos

Learning Scenarios:
The lecture's information about the Orange product, benefits, features, limitations, conclusion can be provided at the beginning of the lab session. Otherwise, it can be as a part of the lecture: Big Data Analytics Tools Overview II. The next step is video watching, where the whole path of activities and tasks are briefly demonstrated: workspace, widgets, EDA, classification & regression (predictive analytics), hierarchal clustering, and image processing. Then, a discussion on demonstration and tasks, defining the objectives and activities for the first session. Following software installation and repeating the demonstration for EDA, building first workflows with built-in datasets and analyzing what and why is done. Comment what the results are, how to solve a particular task, what widgets fit the solution, how to pass through all activities for data preparation described in the lecture: Data Analytics Foundation II – measures of spread, shape, location, dealing with missing data and outliers, data dispersion and distribution, PCA, normalizing (scaling) the data. Finally, individuals complying the tasks with no built-in datasets

6.1. Tableau

Tableau is a data visualization platform that can perform Big Data analytics. Users can leverage well-known frameworks such as Apache Hadoop, Spark and NoSQL databases to meet their data needs. It simplifies the management, sorting and analysis of information through a single, digestible dashboard. Businesses can incorporate data from all sources and visualize it in a myriad of ways to acquire insights.

The vendor offers the following products: Tableau Online, Tableau Desktop, Tableau Prep, Tableau CRM and free of charge Tableau Public.

Tableau Lab Sessions:

- Installing the software – Tableau Public.
- Data workspace and loading data.
- Using limited preprocessing functionality.
- Familiarization with visualization and analysis workspaces.
- First visual analysis.
- Exploring different visualization techniques.
- Forecasting.
- Clustering.
- Dashboards.
- Story.

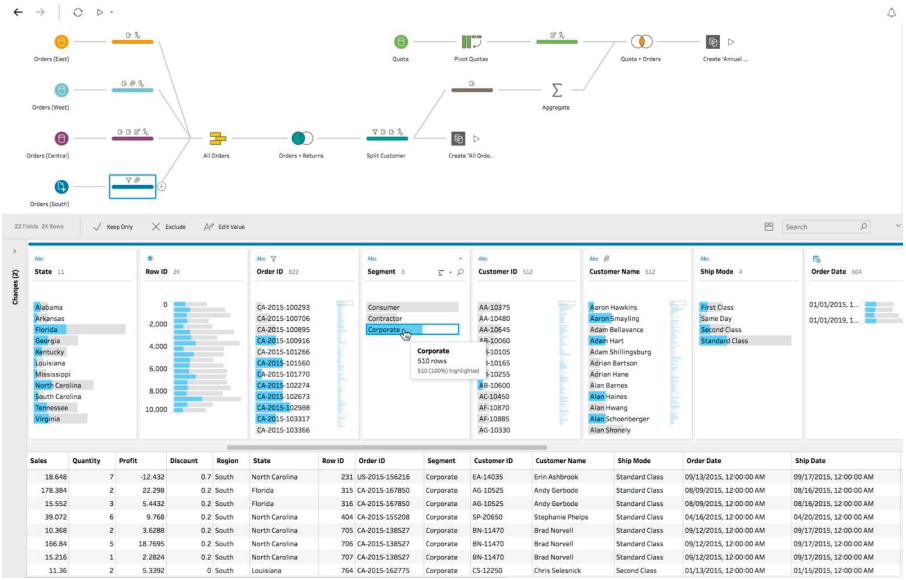

Figure 6. Tableau Prep View

Benefits	Features
- Responsive dashboards and reports - all without writing a code - Leverage a fast, in-memory processing engine - Combine and analyze large data sets	- One Data Interface - Big Data Integrations - VizQL (a visual query language for databases) - Data Catalog

Limitations	Training
- Organizations are dependent on Tableau to maintain servers - Challenges when interpreting complex business rules	▪ a library of free self-service training videos ▪ on-demand and live webinars ▪ e-learning and classroom training courses

Learning Scenarios:
The lecture's information about the Tableau product, benefits, features, limitations, a conclusion can be provided at the beginning of the lab session. Otherwise, it can be as a part of the lecture: Big Data Analytics Tools Overview II. The next step is video watching, where the whole path of activities and tasks are briefly demonstrated: data loading, workspace, visual analysis techniques, forecasting, and clustering. Then a discussion on demonstration and tasks, defining the objectives and activities for the first session. Following software installation and repeating the demonstration, building first visualizations and analyzing what and why is done. Comment on the results, how to solve a particular task, what techniques fit the solution, how to pass through all activities for data visualization described in the lecture: Data Analytics Foundation II –EDA through Plots. Finally, individuals complying the tasks with other datasets.

7. Course competences, knowledge and skills

Computers have fundamentally changed the way the world produces, manages, processes, and analyses data. Moreover, the amount of data produced keeps growing every day, and what we do with it and how we extract value from it have become crucial questions for our society. The course in Data Science provides high-quality education geared towards nurturing the next generation of data scientists. Research in the field of data science requires solid skills in managing and storing massive amounts of data, as well as the ability to develop efficient mathematical algorithms for data analysis. A description of the structure of the course and how it provides students with the necessary knowledge and skills follows.
The training is conducted according to the already established methodologies and teaching materials. The trainers are representatives of UNI teachers. Teachers from ULSIT, UBB, and TSNUK will attend the classes and assist in their conduct.
The format is tailored to the academic classes at the university.
Craft topics are related to the goals and objectives of the project:
- Database based on an advanced reference in the case of Big Data (covering 01 - A1.1. Data collection and A1.2. Analysis).
- How to analyze requirements with a bigger day and how to find the best solution to the problem (covering 02 and 03).
- How to best prepare Big Data professionals in the Data Like ecosystem (O3 and O4 outputs).
- How to help managers find the best way using their Big Data resources.
- Students from all HEIs participating in the consortium will join the activity either physically or virtually.

Certification of attendance will be provided to all participants.

Competences:
- Ability to select an efficient algorithm(s) for Big Data problem, which takes into consideration the scale.
- Ability to model, analyze, and evaluate an organization's business processes.
- Capability to choose the best sampling and filtering method(s) for a given Big Data analysis case.
- Effectively use a variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics).
- Apply quantitative techniques (statistics, time series analysis, optimization, and prediction).
- Using a wide range of Big Data analytics platforms.

Skills:
- Capable of quickly adapting activities to new technologies.
- Able to perform an objective analysis of a data-driven problem and take appropriate actions to solve it through analytics tools.
- Compare analytics tools and specify differences between them by purpose, features, application domain, limitations and training.
- Identify, compare, and apply open-source and automated machine learning data analytics tool(s).
- Select and apply the most appropriate analytics tool(s) for a specific data-driven problem.
- Critically assess the data source, usefulness, and potential problems associated with the data.
- Upload, edit, save, and export data using analytics tools.
- Assure data quality through analytics tools.
- Apply and fit ML techniques to the analytical problem using the appropriate tool (s).
- Apply adequate model evaluation metrics and accurately interpret the analytics output.
- Use analytics tools for data visualization to present concepts/ideas/phenomena from a new perspective to decision-makers.

9. Matrix of competencies

'Competence based learning course in Big Data " is a course whose ambition is to increase the competencies of learners in the field of Big Data and data science. The main goal of the training course is to form knowledge, skills, and competencies in the interdisciplinary field of data science and to enable students to gain practical experience with real data.

The total number of modules involved in the generalization of the course is 12. All modules follow a specially generated matrix of competencies (Table 1) in the course of research work on the project. Here

1 Ability to process large volumes of data using hierarchical storage, hashing and filtering.
2 Ability to select the efficient algorithm to Big Data, which takes under consideratio n its scale.

3 Ability to model, analyze, and evaluate organization' s business processes.
4 Ability to select appropriat e sampling and filtering method for given Big Data analysed case.
5 Ability to propose concepts, models, create and adapt methods and tools for professional activities using resources from others sciences.
6 Effectively use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics).
7 Design, build, operate relational and nonrelation al databases (SQL and NoSQL).
8 Apply quantitative techniques (statistics, time series analysis, optimization, and prediction).
9 Using wide range of Big Data analytics platforms.
10 Ability to tackle with concurrenc y / parallelism problems of Big Data scale.
11 Ability to process large volumes of data using hierarchica l storage, hashing and filtering.
12 Ability to select the efficient algorithm to Big Data, which takes under consideratio n its scale.
13 Ability to model, analyze, and evaluate organization' s business processes.
14 Ability to select appropriat e sampling and filtering method for given Big Data analysed case.
15 Ability to propose concepts, models, create and adapt methods and tools for profession al activities using resources from others sciences.
16 Effectivel y use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptiv e and Predictive Analytics).
17 Design, build, operate relational and nonrelation al databases (SQL and NoSQL).
18 Apply quantitative techniques (statistics, time series analysis, optimizatio n, and prediction).
19 Using wide range of Big Data analytics platforms.
20 Ability to tackle with concurrenc y / parallelism problems of Big Data scale.

Table 4. Matrix of competencies

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Introduction to Big Data Science, AI, ML and DL			x	x	x	x				x
Data collection and processing in business companies	x		x	x			x			
Data Mining	x	x	x	x	x	x				x
Technologies for collecting, storing and managing Big Data	x			x						x
Apache Hadoop Components - HDFS, Hadoop YARN, Hadoop MapReduce, HIVE, Apache Mahout	x			x		x			x	x
NoSQL and Oracle databases	x						x		x	
Tools for Big Data analysis – Watson,	x	x	x	x		x		x	x	x
Orange,	0	0	0	0	0	0	0	0	0	0
Knime etc	0	0	0	0	0	0	0	0	0	0
Data analysis. Workspace. Data acquisition	x	x								
Data Tidying and Cleaning, Data Visualization	x			x						

Exploratory Data, Data Analysis		x		x						
Classification methods, Neural Networks		x				x		x		
Statistical Methods		x				x		x		
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Introduction to Big Data Science, AI, ML and DL			x	x	x	x				x
Data collection and processing in business companies	x		x	x			x			
Data Mining	x	x	x	x	x	x				x
Technologies for collecting, storing and managing Big Data	x			x						x
Apache Hadoop Components - HDFS, Hadoop YARN, Hadoop MapReduce, HIVE, Apache Mahout	x			x		x			x	x
NoSQL and Oracle databases	x						x		x	
Tools for Big Data analysis – Watson, Orange, Knime etc	x	x	x	x		x		x	x	x
Data analysis. Workspace. Data Acquisition	x	x								
Data Tidying and Cleaning, Data Visualization	x			x						
Exploratory Data, Data Analysis		x		x						
Classification methods, Neural Networks		x				x		x		
Statistical Methods		x				x		x		

Conclusion

iBigWorld has a direct impact on the preparation of high ICT-skilled, empowered, motivated, and self-confident future professionals.

This student training aims to achieve two primary goals in the learning path of Big Data. Deepening the interdisciplinarity in the Big Data domain where Data Mining, Machine Learning, Data Science, and Advanced Analytics play a role as an approach palette to knowledge discovery. The lectures provide an overview of the Knowledge Discovery Paradigm based on Big Data, interdisciplinary links between fields, actors, and processes involved in Analytics, and potential applications, impact, and importance for business digital transformation, Industry 4.0, and Society 5.0.

Accelerating skill-building in Big Data Analytics by applying supervised and unsupervised approaches for regression, classification, clustering, and feature engineering through particular software tools (Orange, Tableau) following the learning-by-doing and project-based methods.

The training is competences oriented.

Being aware of the evolving employability requirements through the Smart Big Data Job Hub, students receive professional advice, support, and experience, they will be able to build a well-oriented, in-line market needs specialization accessing increased employability opportunities.

Students, jointly directed by academics and industrial partners, will have the opportunity to respond to the rapid technology changes, through the outputs and objectives of this project. Through the updated curricula, students and graduates

will be able to obtain state-of-the-art knowledge and skill background while accessing the Smart Big Data Job Hub will make job inquiring easier and more appropriate for their expertise.
Co-supervision of open seminars, lectures, skill-aided tutorials, boot camps, and open learning material will enforce their insight to perceive and realize the real needs of business demands, resulting in a significant added value to their professional profile.
As a whole, students will receive more strategic, integrated, and specialized BIG DATA education and training and this workload will be recognized through ECTS credits and EUROPASS.

REFERENCES

1. BERNDTSSON M., et al.: Thesis projects: a guide for students in computer science and information systems. Springer Science & Business Media, 2007.
2. DEY SOMDIP: A beginner's guide to computer science research. XRDS: Crossroads, The ACM Magazine for Students 20.4 (2014): 14-14.
3. BERNDTSSON M., et al.: Planning and Implementing your Final Year Project—with Success!: A Guide for Students in Computer Science and Information Systems. Springer Science & Business Media, 2002.
4. MEERBAUM-SALANT O., ARMONI M., BEN-ARI M.: Learning computer science concepts with scratch. Proceedings of the Sixth international workshop on Computing education research. 2010.
5. HOLZINGER A.: Process guide for students for interdisciplinary work in computer science/informatics. BoD–Books on Demand, 2010.
6. iBigWorld: Innovations for Big Data in a Real World (Erasmus+ project 2020-1-PL01-KA203-082197) documentation *https://ibigworld.ath.edu.pl/index.php/en/home-english/*