EPSILON

Info & Networking EVENT

Harz University of Applied Science

DECEMBER 18, 2024

Hochschule Harz
Harz University of Applied Sciences

EPSILON

Funded by
the European Union

AGENDA

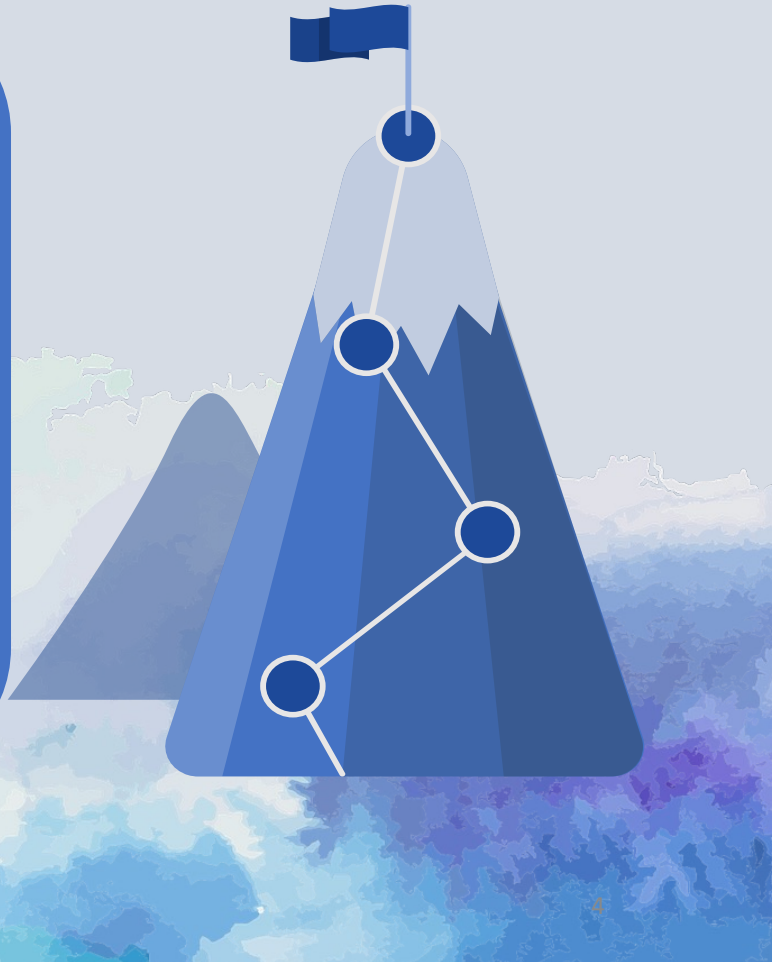
1. Data Science for Good
2. Introduction to the EPSILON Project
3. Project Result 1: European Data for Good Needs Analysis
4. Project Result 2: Knowledge Platform and Business Intelligence Toolkit
5. Project Result 3: New Data for Good Initiative in Lithuania
6. Project Result 4: Teaching & Training Material
7. Presentation CorrelAid
8. Discussion and Conclusion

What is Data Science for Good?

- Use of Data Science Methodologies to address and solve Societal Challenges
- **Key Areas of Application:**
  - Public Health & Social Services
  - Education
  - Environment
- **Benefits:**
  - Empowers organizations with evidence-based decision-making
  - Promotes transparency and accountability in addressing social issues
  - Fosters collaborations between data scientists, NGOs, and governments
- **Challenges:**
  - Ensuring data privacy and ethical use of data.
  - Overcoming resource constraints in non-profit sectors.
  - Bridging the gap between technical experts and domain-specific stakeholders.

Introduction to the EPSILON Project

The transnational project EPSILON with partners from Germany, Portugal, Cyprus and Lithuania addresses the needs of European Data for Good initiatives and higher education institutions with degrees in Data Science as well as NGO´s/NPO´s.



Hochschule Harz
Harz University of Applied Sciences

EPSILON

Funded by the European Union

Nova School of Business & Economics
Project Result 1: Overview & Objectives

- **Title:** Needs Assessment and Best Practices for Data for Good Initiatives in Europe
- **Overview:**

- Assessment of the needs and challenges of stakeholders in the Data for Good field across Europe, focusing on creating resources and establishing best practices

- **Objectives:**

- Compilation of a database of current European Data for Good initiatives
- Workflows for best practices in volunteer organizations
- Analysis of target groups to identify essential processes for Data for Good associations

Funded by
the European Union

Project Result 1: Methodology and Key Tasks

- **Task 1:** State of the Art
  - literature research on existing European initiatives and projects
- **Task 2:** Stakeholder Interviews
  - Assessment of the challenges and needs
- **Task 3:** Focus Groups
  - insights and opinions on best practices and needs
- **Task 4:** Workflow Creation
  - standardized workflows for volunteer, sponsor, and institution management, as well as for project development and communication

Funded by
the European Union

Project Result 1: Impact and Transferability

- **Innovation:** Bridges the gap in European resources on best practices for Data for Good associations through collaboration with educational and social impact entities
- **Impact:** Supports further project outputs, including the prototyping of BI tools, launching new initiatives, and developing educational materials
- **Transferability:** Enables application of identified processes and solutions in future project phases and external initiatives, promoting broader adaptation across various regions and organizations

University of Cyprus
Project Result 2: Project Overview

- **Title:** Creating a Digital Environment for Data Science Volunteers
- **Objective:**

- Develop a digital environment to support the work of data science volunteers & social organizations

- **Key Deliverables:**

- An open-access, sustainable, interactive European Knowledge Platform
- A Business Intelligence (BI) Toolkit for data storage, synthesis, and presentation.

Project Result 2: Needs Analysis and Innovation

- **Needs Analysis:**
- Conducted during the first project phase
- Requirement gathering from stakeholders via questionnaires as needed
- **Target Groups:** Higher education students, staff, and alumni
- **Innovation:**
- Developing new functionalities based on unmet needs
- Implementation Process: 1. Gather user requirements. 2. Elaborate specifications, including GDPR privacy requirements. 3. Design system architecture and components. 4. Develop the platform and BI toolkit.

Project Result 2: Impact and Transferability

- **Expected Impact:**
  - Integration of project results meeting stakeholder needs
  - Accessible platform supporting "Data for Good"
  - Sustainable resource offering free, open access information and training material
- **Transferability Potential:**
  - Usable by volunteer data scientists across various contexts
  - Accessible materials and tools available anytime, anywhere
- **Evaluation:**
  - Testing and feedback process with consortium members and external stakeholders

Funded by
the European Union

Project Result 3: New Data for Good Initiative in Vilnius University, Lithuania

- Core team:
- Dalia Krikščiūnienė
- Virgilijus Sakalauskas
- Giedrius Romeika

Vilnius university

Kaunas Faculty

LITHUANIA

Project Result 3: Action Plan and Sustainability Measures

- **Action Plan:**

1. Develop a regional initiative action plan
2. Engage regional stakeholders
3. Recruit volunteers - Utilizing VU's network and other HE institutions.
4. Training: Using resources and materials from previous project results.
5. Initial Project Launch: Select and support an institution.
6. **Sustainability:** - Develop a sustainability action plan
7. Integrate into the European 'Data for Good' ecosystem.

Project Result 3: Data science laboratory

The Data science laboratory is established as a hub for engaging data science professionals, students, experts to join volunteer network :

Date of establishment: 02/02/2024

Contact: VU Kaunas Faculty - Institute of Social Sciences and Applied Informatics

VILNIUS UNIVERSITY

KAUNAS FACULTY BOARD

RESOLUTION

ON THE ESTABLISHMENT OF THE SCIENTIFIC AND APPLIED RESEARCH LABORATORY "DATA SCIENCE LABORATORY" (LITH. TITLE: „DUOMENŲ MOKSLO LABORATORIJA“) AT THE INSTITUTE OF SOCIAL SCIENCES AND APPLIED INFORMATICS OF THE KAUNAS FACULTY OF THE VILNIUS UNIVERSITY

In accordance with the regulations of the Kaunas Faculty of Vilnius University, approved by the resolution of the Senate of Vilnius University in October 18, 2016. No. S-2016-9-2 "Regarding ratification of Kaunas faculty regulations" (Amended edition of resolution No. S-2018-5-8 of Vilnius University Senate dated May 22, 2018), clause 22.11, Vilnius University Kaunas Faculty Board makes a resolution:

1. To establish a scientific and applied research laboratory "Data Science Laboratory" at the Institute of Social Sciences and Applied Informatics of the Kaunas Faculty of Vilnius University, which does not have the rights of an administrative unit;

2. Assign prof. Dr. Dalia Krikščiūnienė a supervisor of the "Data Science Laboratory".

Chair of the Faculty Board Assoc.prof.dr. Vytautas Evaldas Rudžionis

Project Result 3: Engage regional stakeholders

The Data science laboratory is linked to the expert network of professional organizations:

In Lithuania:

- Lithuanian computer society
- VU Institute of Data Science and Digital Technologies
- Artificial Intelligence Association of Lithuania
- BCCS (Blockchain CyberSecurity Compliance Solutions) cluster

International organizations:

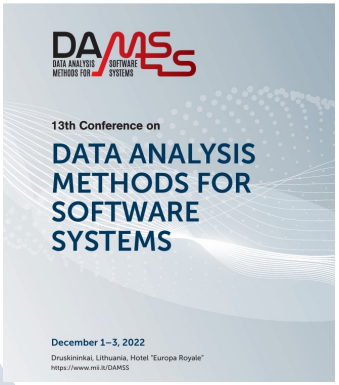
- **EPSILON network - Data for Good organization map**
- **Affiliated with the Z-inspection® initiative.**

Project Result 3: Engage regional stakeholders and data science experts

EPSILON was presented at DAMSS 2022 (proceedings Vol.31,2022)

Lithuanian Computer Society conference :, September 28-29, 2023 (moderated by Giedrius Romeika, dean of Kaunas Faculty, VU)

The DAMSS 2023 was used for inviting volunteers,launch of PR3 activities and tasks (Nov.30-Dec.02, 2023 (EPSILON meetup moderated by Dalia Kriksciuniene and Virgilijus Sakalauskas



Conceptual Framework of Data Science for Good Squad

Dalia Kriksciuniene, Virgilijus Sakalauskas, Giedrius Romeika

Vilnius University

dalia.kriksciuniene@knf.vu.lt

The idea of the research is based on the EU initiative «Data science for good» (DS4G), which aims to effectively use data collected by public organisations due to digital transformation of society. These data are related to health, education, legal environment and security, development of labour resources, energy, transport, sustainability, climate change problems, and other solutions.

The DS4G concept tackles the problem of how to help public organisations not only to collect but also to extract useful information from data, allowing them to make decisions useful to the society. The use of data is complicated not only due to the abundance of collected data but also due to their different sources, collection formats, insufficient data «cleanliness», and the lack of specialists in IT, data science and analytics.

The efforts to bring together specialist communities and solve the data analytics problems raised by the DS4G initiative are already being made. Leading companies in the field of technology, such as Facebook, Google, and Amazon are becoming more and more actively involved in solving these issues.

The objective of the research is to propose the conceptual framework to create an innovative data science hub model (Data for Good squad), which would ensure the creation of effective workflows of operational processes, preparation of training materials, methodologies for identifying company problems and needs, management of risk and sustainability factors, and a building a relevant network of data science specialists. To implement this idea, volunteer teams of specialists in the field of data science and analytics are being assembled.

The research is inspired by the international EU project EPSILON «European Platform for Data Science: Incubation, Learning, Operations and Network». Project coordinator – Harz University (Germany), Nova SBE Institute of Science, Business and Economics (Portugal), University of Cyprus (Cyprus), Vilnius University (Lithuania).

Project Result 3: Initial Project Launch- Select and support an institution

The list of social organizations in the State Tax Inspection was analysed: Paramos statistika - VMI **- (The total number 22504 registered non-profit organizations in LT)**

The social organizations with established interest for partnership DS4G initiative at the Data science laboratory:

1.Public institution Viešoji įstaiga VŠĮ „PENKTA KOJA“ (registration code 301536790)

2.Public institution Viešoji įstaiga Mixed voice choir LELIUMAI (registration code 193262535)

3.Public institution Prienai-Birstonas diabetes club VERSME (Viešoji įstaiga Prienų-Birštono diabeto klubas „Versmė“) (registration code 190161189)

Project Result 3: Initial Project Launch
Pilot organization

Public institution „PENKTA KOJA“ was selected as the PILOT project

It was contacted via alumni and social partner network of Vilnius:

web: Šunų prieglauda - Penkta koja

Project Result 3: Pilot activity plan

DS4G project 1: Identification of the Data science project scope of 'Penkta koja' non-profit organization, project task definition, volunteer teams, roles

DS4G project 2 : The processing, cleansing procedures of the dog data and its analysis and visualization

DS4G project 3: Online advertising (Google Ads) project: creating advertising content, ads for paid advertising, based on the analysis of the data during project 1.

DS4G project 4: Image recognition project by applying machine learning methods in Python environment

DS4G project 5: Dog data transformation for data analysis, Statistica-Tibco tools

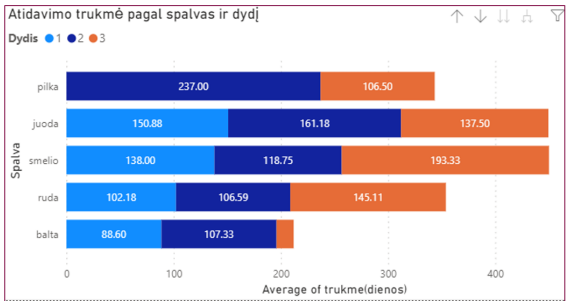
Project Result 3: Pilot activity
DS4G project 2: The analysis of the dog data

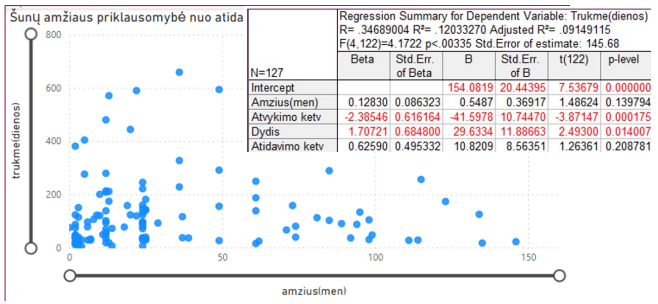
- Data source, processing, analysis

We have added 8 attributes to our initial data:

- Days in private animal shelter
- Dog age in month
- Arrival month
- Arrival quarter
- Release month
- Release quarter
- Dog color
- Dog size

Nr	Reg.Nr	Vardas	Atvykimo data	Atidavimo data	Lytis	Išvaizda	Amžius (apytikslė gim.data)	Trukme(dienos)	Amzius(men)	Atvykimo menuo	Atvykimo ketvirtis	Spalva	Dydis	Atidavimo menuo	Atidavi ketvirti
1	20-349	nida	19/08/2020	07/03/2021	m	smėlinė balta	04/11/2019	200	10	8	3	smelio	2	3	
2	20-386	nordas	02/10/2020	09/03/2021	v	ruda	01/03/2019	158	19	10	4	ruda	3	3	
3	21-94	keksas	15/02/2021	09/03/2021	v	juoda ruda	17/02/2009	22	146	2	1	juoda	2	3	
4	21-99	briutas	16/02/2021	16/03/2021	v	juoda balta	01/09/2020	28	6	2	1	juoda	2	3	
5	20-473	fiona mirta	24/11/2020	16/03/2021	m	balta pilka	20/03/2014	112	81	11	4	balta	1	3	
6	20-388	doveris	05/10/2020	31/12/2020	v	balta	01/01/2013	87	94	10	4	balta	1	12	
7	20-369	karalis	28/08/2020	15/01/2021	v	juoda ruda	18/08/2018	140	25	8	3	juoda	2	1	
8	20-453	hera	13/11/2020	18/01/2021	m	juoda ruda	01/01/2015	66	71	11	4	juoda	1	1	
9	20-418	čikas	24/10/2020	05/02/2021	v	juoda ruda	06/10/2012	104	98	10	4	juoda	2	2	
10	21-877	sirena	29/01/2022	04/11/2022	m	juoda	29/01/2021	279	12	1	1	juoda	3	11	
11	22-094	bonė	11/05/2022	01/11/2022	m	juoda balta	01/05/2021	174	13	5	2	juoda	1	11	
12	22-404	semas	02/10/2022	30/10/2022	v	ruda balta	05/10/2020	28	24	10	4	ruda	3	10	
13	22-347	ananasas	12/09/2022	30/10/2022	v	ruda juoda balt	18/07/2022	48	2	9	3	ruda	1	10	
14	22-351	vaivas	12/09/2022	18/10/2022	v	ruda juoda balt	18/07/2022	36	2	9	3	ruda	1	10	
15	21-248	dodo	24/07/2022	22/10/2022	v	kreminė	23/07/2020	90	24	7	3	juoda	1	10	





Registracijos numeris	20 343
Vardas	PEDRO / NAMO 2021 02 12
Atvykimo data	2020 10 15
Priežastis	+ tipo Juodis 5-6m iš kiaulių fermos UAB švaistūnas teritorijos
Lytis	Patinas
Išvaizda	Juodas
Amžius	
Ypatingos charakterio savybės	
Sterilizacija/kastracija	2020 11 04 Kastracija 2020 10 16
Skiepai	
Vaistai nuo erkių/blusų	
Vaistai nuo kirminų	
Gydymas	
Kitos pastabos	2020 11 04 pasalinti pridetimai pirštai

Project Result 3: Pilot activity
DS4G project 3: Online advertising analysis

1) The project teams defined (total 7 student teams)

2) The registration in the Google Ads environment, certifications

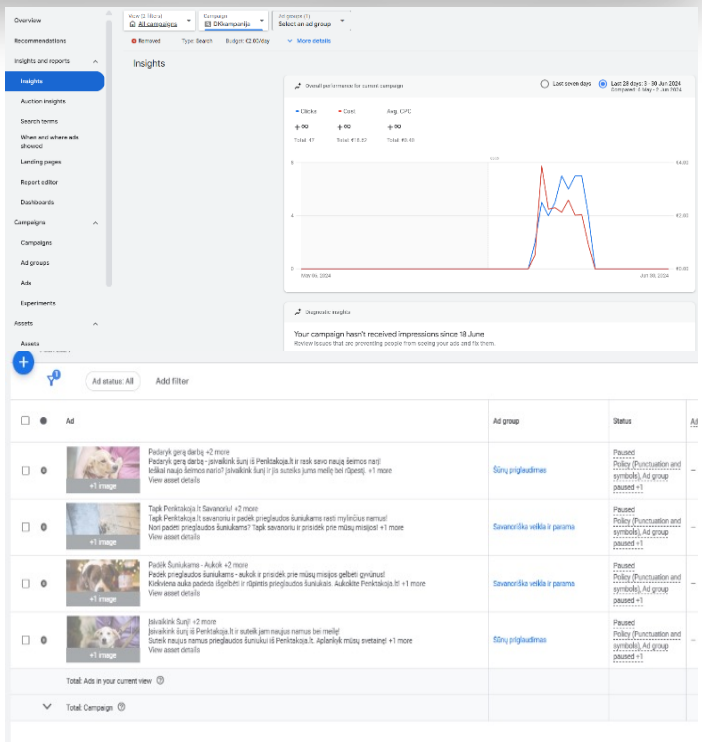
3) The keyword and search term databases created, analysed

4) Google Ads campaigns and the ads database created, indicators analysed

5) The campaigns and ads quality evaluated

6) The online campaign result analysis

7) Presentation and report to the content marketing and data analysis experts



Project Result 3: Pilot activity
DS4G project 4: Image recognition by ML

Dog Recognition System

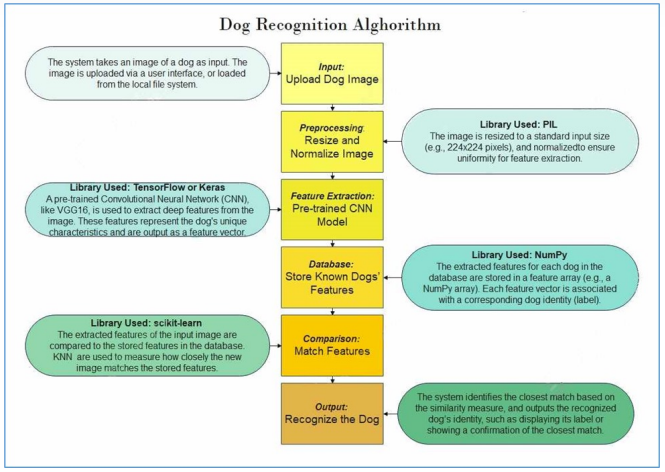
The aim of this project is to identify and take care of homeless or lost dogs who may have been hosted by dog shelter "Pekta koja" and are photographed and registered there. The idea of the project is to try to identify a lost dog based on a database of dog photos in “Penkta koja” shelter.

Goal

To identify a lost dog based on a database of dog photos.

Method

Leverage computer vision and machine learning to create a model that can compare a new dog’s photo with the photos in the existing database and recognize the dog.



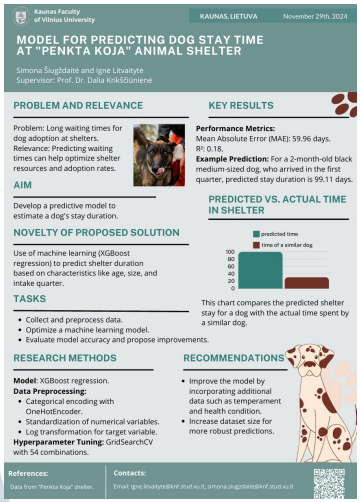
Python code (fragment)

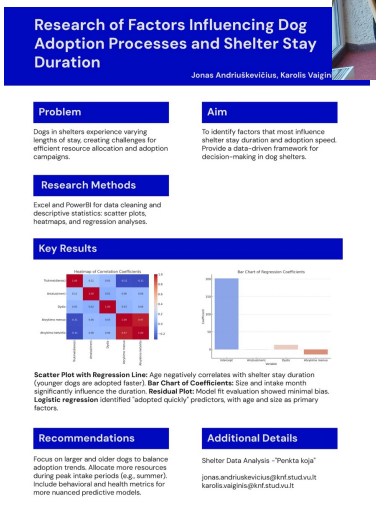
```python
# Extract features from the database images and get the labels and image paths
database_features, labels, image_paths = load_database(database_path)

# Train a KNN classifier using the extracted features
knn = KNeighborsClassifier(n_neighbors=1)
knn.fit(database_features, labels)

distance_threshold = 500
# Function to recognize a new dog from an image
def recognize_dog(new_image_path):
    # Extract features from the new image
    new_features = extract_features(new_image_path)

    # Predict the closest dog from the database
    recognized_dog = knn.predict([new_features])

    # Find the corresponding image path for the recognized dog
    # index = knn.kneighbors([new_features], n_neighbors=1, return_distance=False)[0][0]
    distances, indices = knn.kneighbors([new_features], n_neighbors=1, return_distance=True)
    nearest_distance = distances[0][0]

    if nearest_distance < distance_threshold:
        # If the nearest distance is below the threshold, recognize the dog
        recognized_dog = knn.predict([new_features])[0]
        recognized_dog_image_path = image_paths[indices[0][0]]
        return recognized_dog, recognized_dog_image_path, nearest_distance
    else:
        # If the nearest distance exceeds the threshold, return "Unknown Dog"
        return "Unknown Dog", None, nearest_distance

# Function to show the recognized dog's image
def show_image(image_path, title):
    if image_path:
        img = Image.open(image_path)
        plt.imshow(img)
        plt.title(title)
        plt.axis('off')  # Hide the axes
        plt.show()
    else:
        print(title)

# Test the recognition system with a new dog image
recognized_dog, recognized_dog_image_path, nearest_distance = recognize_dog(file_selected)

if recognized_dog == "Unknown Dog":
    print(f"Recognized Dog: {recognized_dog} (Distance: {nearest_distance:.4f})")
    show_image(None, "No matching dog found")
else:
    print(f"Recognized Dog: {recognized_dog} (Distance: {nearest_distance:.4f})")
    show_image(recognized_dog_image_path, f"Recognized Dog: {recognized_dog}")
```

Demo video was created

Project Result 3: Pilot activity
DS4G project 5, involvement of students

Project 5 results are presented in the poster session of [VU Kaunas Faculty - Prof. Vladas Gronskas International Scientific Conference](), November 29, 2024





Summaries and posters are published in 19th Prof. Vladas Gronskas International Scientific Conference Abstract Book (2024), ISSN 26690233, https://doi.org/10.15388/VGISC.2024.II , Vilnius University, 2024

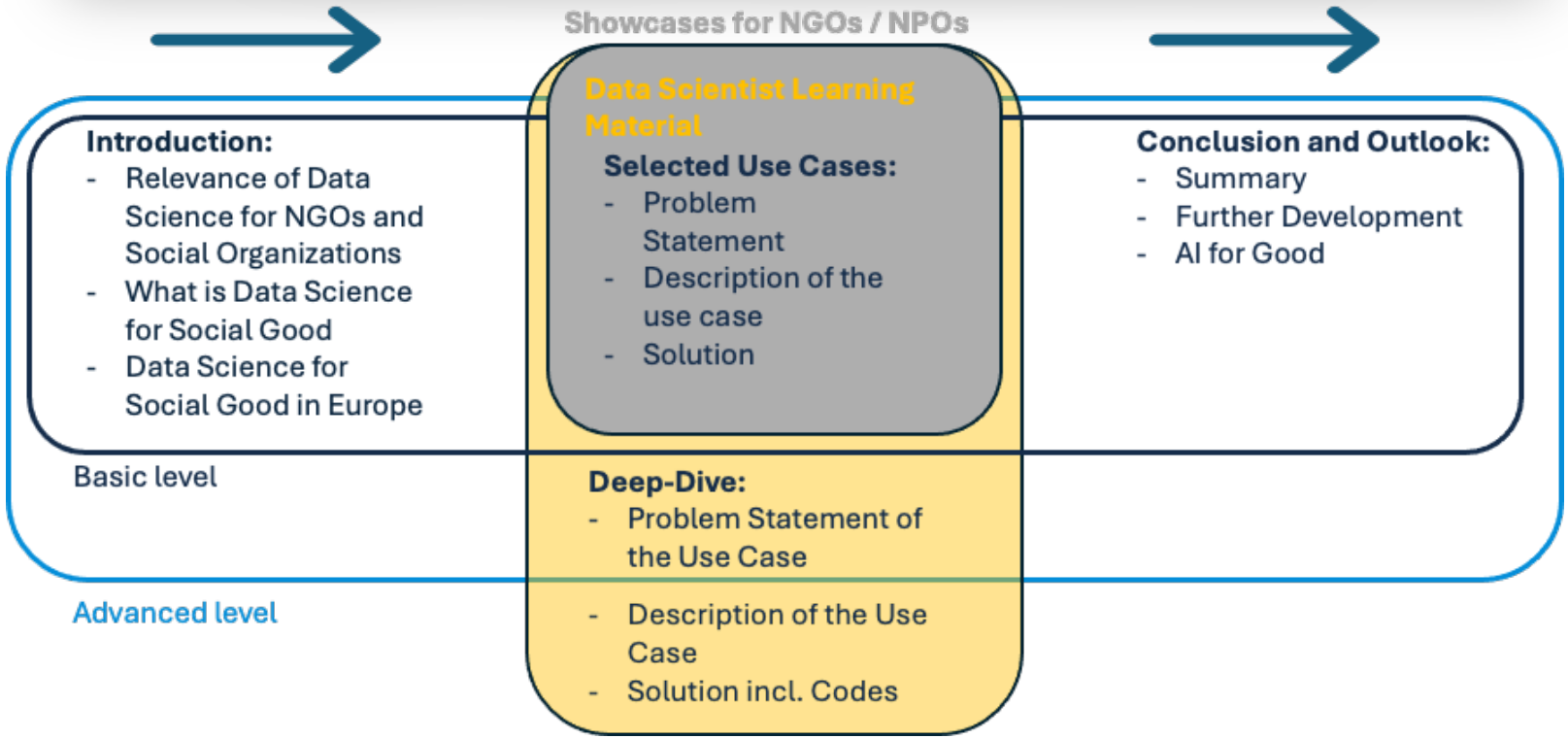
Hochschule Harz
Harz University of Applied Sciences
EPSILON
Funded by the European Union

Harz University of Applied Sciences
Project Result 4: Teaching & Training Material

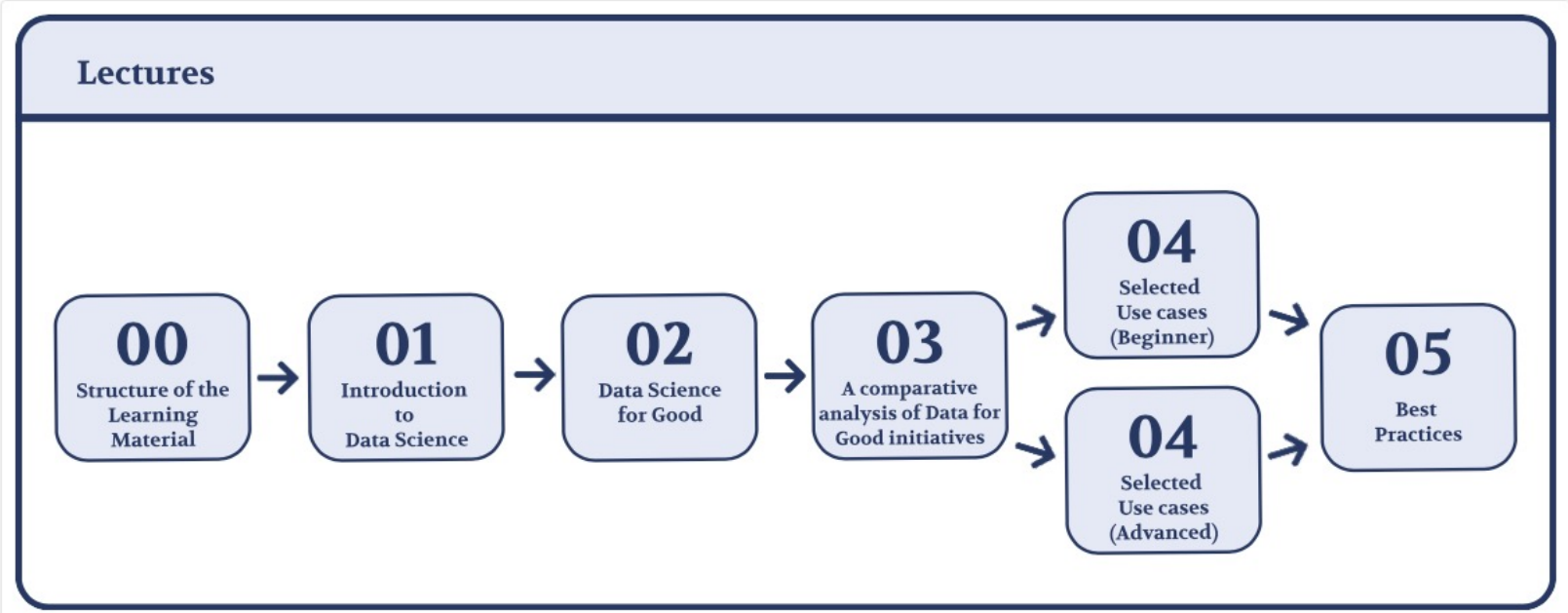
- **Project Outcome:**
  - Development of learning and training materials accessible via Knowledge Platform
  - Documents and Videos (OER)
- **Needs Analysis:**
  - Identification of challenges and use cases relevant to different target groups
  - Knowledge from previous project results transformed into specific learning materials
  - Focus areas: recruiting, project scoping, and measuring project impact
- **Target Groups:**
  - 1. Higher Education Institutions: Learning material at different levels
  - 2. Data Enthusiasts: Best practices for efficient organization of Data Science teams
  - 3. Social Organizations: Insight into possible applications of Data Science

Funded by
the European Union

Project Result 4: Overview and Levels



Project Result 4: Structure of the Learning Material



https://epsilon.cs.ucy.ac.cy/index.php/lectures/
https://www.hs-harz.de/forschung/ausgewaehlte-forschungsprojekte/epsilon/training-materials

Project Result 4: Innovation and Impact

- **Elements of Innovation:**
  - target-group-specific learning materials in Social Data Science, applicable across various disciplines
  - addresses recruitment, project identification, and management in Data Science
  - Tailored materials for Data for Good projects, merging different academic and professional backgrounds
- **Expected Impact:**
  - access to relevant Data Science knowledge for higher education institutions
  - Addresses need for expert knowledge
  - filling the gap in tailored learning content for different disciplines and academic levels

Funded by
the European Union

Project Result 4: Process and Transferability

Process - Agile Project Management Framework

1. **Understand:** Analyze challenges and needs
2. **Create:** drafts of teaching material and didactical concepts
3. **Test:** Test drafts with German data scientists and students to gather feedback
4. **Amend:** Revise materials based on feedback
5. **Apply:** Deploy revised materials at LLTA2 event
6. **Finalize:** Refine materials using feedback from the application phase

Transferability Potential: Learning materials applicable to various higher education programs in fields linked to Data Science (e.g., social science, mathematics, computer science, design)

Funded by
the European Union

Fragen / Questions

Fragen / Questions bzgl. Data Science for Good / the Project EPSILON / CorrelAid

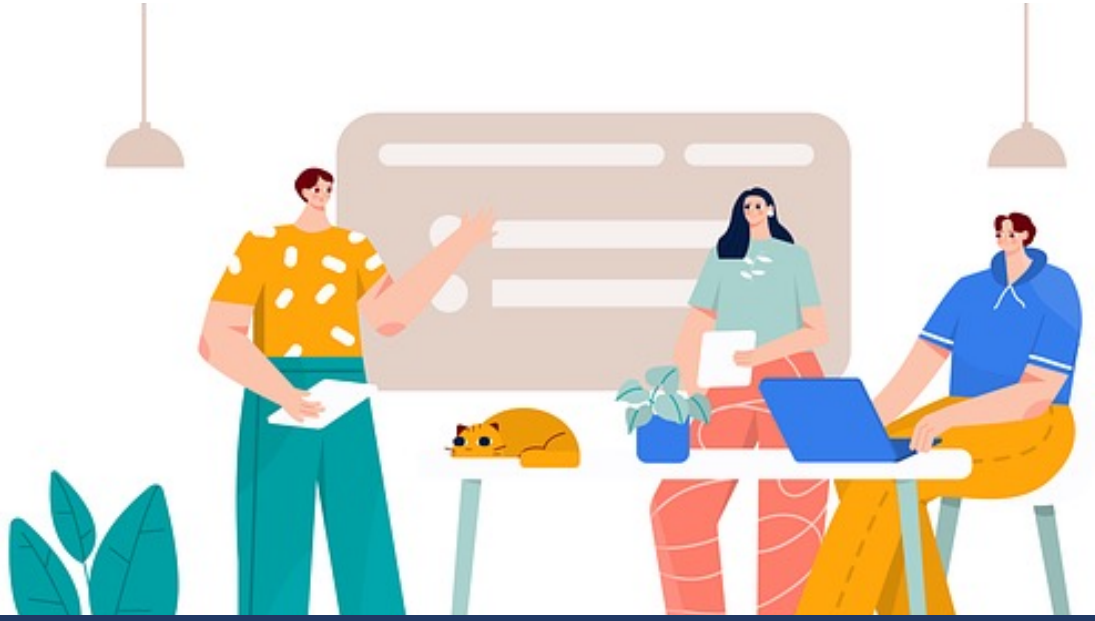
Funded by
the European Union

Interaction

Open Space

1. Wo sehen Sie Anwendungsmöglichkeiten der vorgestellten Inhalte in Ihrem Tätigkeitsbereich? / Where do you see potential applications for the content presented in your area of activity?
2. Was können wir noch tun, Wie müssten die Projektergebnisse aufbereitet sein, damit Sie sie in Zukunft nachhaltig nutzen? How should the project results be prepared so that you can use them sustainably in the future?

Discussion and Conclusion



Thank you!

▲ Hochschule Harz
Harz University of Applied Sciences

EPSILON

Funded by
the European Union

Contact us

Harz University of Applied Science
Friedrichsstraße 57 - 59
38855 Wernigerode

Tel.: 3943 659 872
Mail: gritlehmann@hs-harz.de

Hochschule Harz
Harz University of Applied Sciences

EPSILON

Funded by
the European Union