1. **Name of Course:** Debate and Contradictions in Social Data Science

2. **Instructor:** Márton Karsai

3. **No. of Credits (US/ECTS):** 2 US credits / 4 ECTS

4. **Academic term:** Winter

5. **Course level:** Master

6. **Relationship with other courses:** This course has no pre-requisites

7. **Course type:** Mandatory or SDS MS students

8. **Specialization:** all specializations in SDS MS program

9. **Short description and the overall aim of the course:**

   Our digital era offers unique opportunities and poses a fundamental challenge to academic research, specifically in the social sciences, as technological development and data production on human behavior progresses at an ever-increasing pace. Every 1.2 years, more human-driven socioeconomic data are produced than during all preceding years of human history combined and most of human activities are traced by digital devices. These new technologies, massively relying on digital data, have changed our lives in many respects, like the way we form and maintain social ties, how we participate in social activities, how we consume news and form an opinion about it. We are in the position to follow the evolution of large real-world social systems and to detect the emergence of collective social behavior “in vivo”, for example to do real-time simulations and predictions about the spreading of deadly diseases or the outcome of political campaigns and elections.

   From a scientific point of view, clearly, several disciplines are actively engaged in this challenge, which call for a multidisciplinary effort in research and education. Social Data Science as a new field addresses the quantitative observations and understanding of social behavior of individuals, groups, and societies based on the analysis of large datasets as obtained from “digital footprints” or generated for the purpose of study.

   This seminar course provides a broader view on different fields where Social Data Science methods are applicable. The goal is to explain the actual questions that can be answered with a data science approach and the open debates about their broad applicability in various fields. This course will involve several guest lecturers from CEU departments and elsewhere, who will give lectures about their own disciplines. As a university wide course, it would provide an introduction to any participating students independent of their background to the actual topics of Social Data Science and to data...
driven research in general. It would help the students to develop a critical and reflexive view about the potentials and dangers of the applications of this field. Meanwhile, this course will serve as the first steppingstone for Social Data Science MS students towards their specialization in Applied Social Data Science, Economics, Environmental Science, Political Science and Policy.

**Schedule and thematic**

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Topic</th>
<th>Details</th>
<th>Readings</th>
</tr>
</thead>
</table>
| 1. Márton Karsai (DND) | Opportunities and risks in social data research | - Short introduction to Social Data Science  
- Human behavioral data collection  
| 2. Petra Kralj Novak (DNDS) | Machine Learning from Disagreement | - The nature of the disagreement and data annotation inconsistency  
- Perspectivist approach to data annotation, machine learning model training, and evaluation  
| 3. Balázs Vedres (DNDS/DSSA) | Data Science and the Algorithmic Transformation of the Public Sphere | - Diagnosing the impact of automation on civic publics  
- Bot detection and the coevolution of bots and their suppression  
| 4. Mariyana Angelova (DPSP) | Using automatic content analysis to study policy making and voting behavior | - Advancements in the field of quantitative text analysis  
- Research questions  

**Recommended readings**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 5. | Robyn Dora Radway (HD) | Cooks and Messengers: Data Science and Early Modern Social Networks | - Alba amicorum as sources and data extraction from historical sources  
- Problems with identification, standardization  
- Historical linked open data (like GND)  
- Network analysis and its applications for analyzing different social environments  
| 6. | Viktor Lagutov (DES) - TBC | Geospatial Data and Sustainable Development Goals | - Short introduction to Geospatial Data collection, analysis and visualization  
- Earth Observations: ex-situ vs in-situ data  
- Applications of geospatial methods and packages in monitoring and achieving various SDGs  
- Earth Observations for SDGs. Compendium of Earth Observation contributions to the SDG Targets and Indicators (European Space Agency, 2020)  
| 7. | Miklós Koren (DEB) | Discovering Discrimination with Data | Discrimination is often a covert practice that is hard to detect. This session shows how to use multiple regression analysis to detect and prove discriminatory practices. This requires the appropriate use and interpretation of statistical methods and a deep understanding of the many factors entering business decisions.  
| 8. | Elisa Omodei (DNDS) | Data Science for the Sustainable Development Goals | - Provide an overview on how digital data and data science/ML/AI methods are being used in the development and humanitarian sector to tackle and monitor the Sustainable Development Goals (SDGs)  
- Develop a critical understanding of the challenges and limitations of such approaches  
- Deep dive into two case studies: humanitarian aid targeting and food insecurity monitoring  
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>Mathias Czaika</td>
<td>Migration networks</td>
<td>Invited speaker</td>
<td>-</td>
</tr>
<tr>
<td>11.</td>
<td>Alessia Melegaro</td>
<td>Computational Epidemics</td>
<td>Invited speaker</td>
<td>-</td>
</tr>
<tr>
<td>12.</td>
<td>Márton Karsai (DNDS)</td>
<td>Project presentation and round table</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

10. The learning outcomes of the course:
   - Knowledge about relationship between the data science methods and their applications
   - Broad view on open questions for Social Data Science in various fields of Social Sciences
   - Reflexive view about the potentials and dangers of the applications of Social Data Science
   - Ability to identify relevant topics for specialization and capstone projects

11. Learning activities and teaching methods: This course is organized as a seminar course with several lecturers invited from various fields of Social Sciences. Lecturers will alternate between topics each week. Students will be confronted with the most important readings in the different fields, which will be discussed during the seminar lectures.

12. Assessment:

There are three goals that this seminar seeks to achieve:
   - At the most basic level, all participants are expected to develop an understanding of the key concepts in the application fields of Social Data Science.
   - Second, discussions of the assigned research articles should provide participants with tools to critically read contemporary research and to stimulate new ideas to help them to identify their specialization and capstone project topic.
   - Finally, the more practical assignments are meant to improve skills in presentation, writing and problem-solving.

To achieve these goals, students are graded on the following activities.
   - Active class participation is required (20% of grade)
   - Group project development (80% of the grade). This project is to be developed by teams of up to 6 students. Each team must analyze a controversial topic in Social Data
Science (e.g. from topics discussed by instructors) and suggest possible methods and solutions to address them in a form of a commonly written “white paper” that is presented in front of the class and exam committee during the last session.