Social Data Science (SDS)

General Semester Description

Recent developments in information and communication technology (ICT), growing data quantities (Big Data), and rapidly improving techniques to analyse it are fundamentally changing the context that businesses, public organisations and researchers are facing. Competencies to identify patterns and make sense of data as well as to inform the decisions of managers, policymakers and other actors are in high demand on the labour market. This semester is developed as an intensive training course in Data Science - a new term to describe the combination of data sourcing, management, analytics, visualisation and communication. Data Scientists can apply their skills to problems in various areas. Within business, they can contribute to extracting and combining knowledge from existing Enterprise Resource Planning (ERP) systems, data warehouses, and external sources, and use them to support data-driven strategic decision making. They are able to use sophisticated visualisation techniques such as dynamic dashboards to provide business intelligence and executive guidance. The most prominent examples in marketing include recommender systems in online commerce and entertainment platforms, advanced segmentation approaches or brand perception analysis. In (public) healthcare, they are often contributing to the development of predictive models to make processes more efficient or for more precise diagnostics. In fields such as sociology and political science, Data Science methods provide access to data sources that have traditionally been out of reach, such as the daily interaction of millions on social media, and the data generated by sensors (Bluetooth, GPS, camera etc.) in the emerging "Internet of Things".

The semester consists of three course modules (M1-3), addressing different aspects of Data Science and Big Data analytics plus an applied "industry capstone project" (M4) to give students an excellent opportunity to apply the acquired skills to a real-life setting. Prospect students are expected to have taken the equivalent of 10 ECTS in quantitative methods, statistics, econometrics or comparable (to be evaluated case by case) courses on bachelor or master level.

**M1 Applied Data Science and Machine Learning** is a condensed introduction to the “Data Science Pipeline”, taking students from data acquisition over pre-processing and modelling to evaluation and presentation.

**M2 Network Analysis and Natural Language Processing** will focus on unstructured network data. Students will learn how to explore and analyse natural language (text) as well as relational (network) data of various kinds.

**M3 Deep Learning and Artificial Intelligence for Analytics** focuses on the most recent developments in machine learning, which are deep learning and artificial intelligence (AI) applications, to analytics and particularly useful in Big Data settings. The module will provide a solid foundation to this exciting and rapidly developing field, and enable the students to on-demand update and acquire new knowledge.
The course modules will use online resources and e-learning tools such as podcasting, online tutorials, and mini-assignments, as integral parts of the teaching methodology in order to enhance student engagement outside the classroom. Physical face-to-face time will be centred around the tacit and interactive components of the problem-solving processes.

**Module 4 Applied Social Data Science Capstone Project** is a 15 ECTS project to be completed in collaboration with an external partner (optimally). External organisations can not only provide real-life datasets but also motivate students through real problems, and help to adjust the curriculum to make contents most relevant in terms of employability. The teaching staff will assist students with significant guidance and coordination to identify potential, mutually benefiting areas of project-collaboration with external partners.

Students can enrol in the whole semester or single courses (M1-M3). Enrolment in single courses requires pre-approval by the student’s original study board.

**Language of instruction:** Main instruction language for Modules 1-3 is English. Assignments in these modules may be delivered in English or Danish. Module 4 (semester project) can be written in English or Danish.
M1: Applied Data Science and Machine Learning

Aim: M1 intends to provide an opportunity to sample the core techniques of data science, understand their intuition and application cases. It also aims at showing best practice of how to select specific and appropriate methods for the particular data science project, as well as how to efficiently and autonomously acquire further knowledge of the rapidly evolving field. Insights and techniques learned in this module can be applied to real-world problems in, e.g. marketing (How do you classify customers who are likely to spend a lot?), management (How do you identify performance bottlenecks in the organisation?) or finance (Is this person likely to default on their mortgage?).

Content:
This module is an introduction to the main ideas behind (social) data science, and the essential principles and techniques in the data scientist's toolbox. It aims at providing a broad overview by taking a "bird's eye perspective" and presenting a range of topics briefly instead of focusing on a single topic in depth. The Introduction to Social Data Science will survey the foundational issues in data science, namely:

- Data Sourcing: Where and how to get the right data
- Data Manipulation
- Data Analysis with Statistics and Machine Learning
- Data Communication with Information Visualization
- Data at Scale - Working with Big Data
- Data at Scope - Working with non-traditional data-sources such as text, geographical data, relational data, and more
- Data at Mess - Working with incomplete, ill-structured, decentralised data

Credits (ECTS): 5

Instruction and learning forms:
Lectures will be complemented by online resources and e-learning tools such as podcasting, online tutorials, and mini-assignments, as integral parts of the teaching methodology to enhance student engagement outside the classroom. Physical face-to-face time will be centred around the tacit and interactive components of the problem-solving processes.

Scheduling:
3rd semester of the master programme, autumn.

Requirements for participation:
Completed course in introductory statistics or similar.
Learning Objectives:
Upon completion of the module students will have built a solid and expandable knowledge foundation in modern data science and will have acquired a broad range of skills enabling them to carry out own data analysis projects. Students will be capable of autonomously managing and evaluating complex projects and problems associated with data management, description, and analysis.

Knowledge:
- Understand and explain the main workflow routines and techniques how to obtain, store, manipulate, and analyse data.
- Identify the commonly used programming languages, software and other tools used in data science.
- Explain how to select and execute the most common data analysis techniques.
- Show an understanding of how to use a wide variety of visualisation techniques to explore and describe their data.
- Explain the differences and complementarities between the prediction focussed data science approach, and the causality seeking approach of traditional scientific statistics.
- Provide an overview over the current state-of-the-art in applied statistics and data science.

Skills:
- Install and use relevant software packages in data science.
- Read, import, export, and process data in most widely used data formats.
- Execute common data manipulation techniques such as data-merging, aggregation, pivoting, and treatment of missing values.
- Select and apply standard techniques from 'traditional' statistics and data science to solve empirical problems of data exploration, classification, optimisation, and forecasting.
- Evaluate model performance, fine-tune and optimize models.
- Understand, interpret, critically reflect upon, and explain the results of data analysis.

Competencies:
- Comprehend and participate in current professional and academic discussions in applied statistics and data science.
- Critically reflect possibilities and constraints related to the implementation and evolution of data-driven methods.
- Identify problems which can be wholly or partially solved by the use of data analytics.
- Apply a data-driven logic, structure, and workflow to problem-solving.
- Describe and communicate the results of data analysis in a precise, understandable and informative manner, using appropriate data description and visualisation techniques.
- Expand their knowledge in various data science topics of interest and relevance via self-learning.
**Assessment Criteria**
Module 1 is assessed according to the Danish 7-point grading scale. The grade 12 will be awarded to students who give an excellent performance and demonstrate that they have fulfilled the above objectives exhaustively or with few insignificant omissions. The grade 02 will be awarded to students who demonstrate that they have fulfilled the minimum acceptable level of the above learning objectives.

**Examination**
Portfolio exam:
60% obtained through various graded (and supervised peer-graded) problem sheets and mini-assignments throughout the module.
40% final internal evaluation seminar with oral presentation, peer-evaluation (opponent group), internal critique and discussion departing from the final assignment and presentation.
Module 2: Network Analysis and Natural Language Processing

Aim: M2 aims to give students insight into network and unstructured data types, as well as state-of-the-art approaches to map and analyse these data. Insights and techniques gained in this module will allow students to approach real-world problems in marketing (Who are the main influencers among our customers?), management (Can we identify new discourses in the communication within our organisation?), business economics (Can language patterns be used to understand R&D intensity across companies?), political science (How is a political candidate perceived by a certain demographic, based on their social network statements?), and sociology (How is a person’s behaviour and characteristics affected by their social network?).

Content:
With accelerating digitalisation of the modern world, we capture and store a growing amount of relational and unstructured (e.g. text) data. The former type of data encodes social, biological, physical and other complex systems as a collection of actual or potential relations between some entities. These can be users in an online social network, companies in a cluster, or research articles in a database linked via some association metric. Exploring such networks allows unveiling latent and general structural patterns, to understand how the interaction between elements reflects on their attributes, or how information flows through these systems. Indeed, envisioning and analysing complex systems such as national economies, natural ecosystems, or social interactions as networks have brought fresh wind to a broad range of academic disciplines and professional sectors alike. Working with relational data is not difficult, but it certainly requires some rethinking.

The other type of data, unstructured data, come in many varieties. The one that is arguably most attractive for social science analytics is text. Language encodes a vast range of meanings, entities, and relations. Natural language processing (NLP) has considerably advanced in the past years, making unstructured text suitable for machine learning.

The link between networks and unstructured data is given by the fact that unstructured data usually encode something that is closer to a depiction of reality than traditional structured data. Thus, it will typically contain information on some objects with their attributes as well as relational features linking the objects. Understanding the relational dimension is therefore essential to working with unstructured data.

Credits (ECTS): 5

Instruction and learning forms:
Lectures will be complemented by online resources and e-learning tools such as podcasting, online tutorials, and mini-assignments, as integral parts of the teaching methodology to enhance student engagement outside the classroom. Physical face-to-face time will be centred around the tacit and interactive components of the problem-solving processes.

Scheduling:
3rd semester of the master programme, autumn.

Requirements for participation:
Completed course in applied statistics or similar.

Learning objectives:
Upon completion, students will have built a solid knowledge foundation within network theory and analysis, computational linguistics and broader (unstructured) data processing. The module is application-focused, and thus students will gain a variety of skills to utilise relational and unstructured text data for analysis purposes.

Knowledge:
- Show insights in the conceptual particularities and explanatory power of relational and network data.
- Explain the interplay between network-theory concepts and real-world networks.
- Understand the theoretical foundations, core-algorithms and metrics in network analysis.
- Explain the concepts of multi-dimensional and multimodal networks and demonstrate comprehension of how they can be used for feature detection.
- Describe main approaches to using network data in more general machine learning settings.
- Explain main techniques used in data mining and structuration.
- Explain central concepts within computational linguistics and methods in natural language processing.
- Reflect upon the epistemology of language data.
- Explain how language data is integrated into analytical frameworks.

Skills:
- Source, store and pre-process network and text data.
- Calculate and interpret essential statistic metrics.
- Integrate network indicators into machine learning pipelines.
- Handle multiplex and multimodal networks.
- Visualise networks and interaction pattern.
- Perform grammar-based labelling and modifications on text data.
- Perform tasks such as automated summarisation and sentiment analysis.
- Extract entities from text.
- Identify topics within large collections of documents.
- Calculate semantic similarity.
- Train and use word embedding models.

Competencies:
- Represent any real-life complex systems as networks.
- Identify latent patterns, structures and interactions of entities in these systems.
- Explore the interplay between the structure of systems and their performance as well as particular features and behaviour of individual entities.
- Utilise natural language data for various types of mapping and analysis.
Assessment Criteria
Module 2 is assessed according to the Danish 7-point grading scale. The grade 12 will be awarded to students who give an excellent performance and demonstrate that they have fulfilled the above objectives exhaustively or with few insignificant omissions. The grade 02 will be awarded to students who demonstrate that they have fulfilled the minimum acceptable level of the above learning objectives.

Examination
Portfolio exam:
60% obtained through various graded (and supervised peer-graded) problem sheets and mini-assignments throughout the module.
40% final internal evaluation seminar with oral presentation, peer-evaluation (opponent group), internal critique and discussion departing from the final assignment and presentation.
Module 3: Deep Learning and Artificial Intelligence for Analytics

**Aim:** This module aims at providing insights into the most foundational architectures of deep learning algorithms within both supervised and unsupervised learning, thus building a strong foundation for further exploration of more specific and cutting-edge techniques. Real-world problems that are approached with the techniques covered in this module include the development of advanced recommender systems (marketing), computer vision models (healthcare, economics), powerful unsupervised pattern recognition systems (fraud detection or credit default prediction in finance) and (attempts of) stock market index prediction.

**Content:**
This module focuses on the most recent developments in the field of data science that build on deep learning and different architectures of artificial neural networks. While conceptually, these techniques were already conceived in the 70s and 80s, it was only recently that Big Data created a need and modern computers allowed to use them in practice. Today, deep learning algorithms are behind a variety of online and offline applications. They are enabling massive recommender systems in online retail and entertainment and powering artificial intelligence applications in medical diagnostics. Vast interest and investment in R&D within this area spurred progress of these techniques and made them more accessible. Only a few years ago deep learning and AI were barely known outside computer science departments. Today, these approaches are widely used in medicine, natural sciences and increasingly seen in social science as well as humanities.

While many of these techniques constitute compelling approaches, especially for predictive modelling, yet they do not make more traditional modelling approaches (e.g. techniques learned in M1) obsolete, but offers many synergies. Therefore, the module is structured in a way that makes it easy for students to see, where the analysis can make use of deep learning approaches as an alternative to more established techniques (e.g. regression analysis). Emphasis will be put on outlining the cases in which traditional (often leaner) methods are more suited.

**Credits (ECTS):** 5

**Instruction and learning forms:**
Lectures will be complemented by online resources and e-learning tools such as podcasting, online tutorials, and mini-assignments, as integral parts of the teaching methodology in order to enhance student engagement outside the classroom. Physical face-to-face time will be centred around the tacit and interactive components of the problem-solving processes.

**Scheduling:**
3rd semester of the master programme, autumn.

**Requirements for participation:**
Completed course in applied statistics or similar.
Learning objectives:
Upon completion, students will acquire theoretical and practical knowledge, enabling them to understand and explain central techniques and concepts of deep learning approaches as well as the fundamentals of artificial intelligence for analytics. They will be able to select and apply appropriate methods to real-world problems and critically reflect on them.

Knowledge:
- Explain the central concepts within deep learning.
- Define key elements of artificial neural networks and depict their functionality.
- Describe main architectures of supervised deep learning algorithms.
- Describe main architectures of unsupervised deep learning algorithms.
- Show insight into recent developments in deep learning and artificial intelligence.
- Reflect on ethical and societal problems concerning the use of artificial intelligence.

Skills:
- Install and deploy relevant software packages and cloud services for deep learning approaches.
- Select and prepare various types of data for use in deep learning environments.
- Select and construct different kinds of deep learning architectures (e.g. Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, Self-Organizing Maps, Restricted Boltzmann Machines).
- Implement “correct” training of selected models.
- Tune and optimise models.
- Utilise trained models for prediction tasks.
- Evaluate model performance.

Competencies:
- Use deep learning techniques to solve social science problems in Big Data contexts.
- Make informed decisions about the selection of algorithms (also where it is better not to use deep learning/AI techniques at all).
- Identify cases that require particular attention concerning ethical and social consequences of deep learning and AI application.

Assessment Criteria
Module 3 is assessed according to the Danish 7-point grading scale. The grade 12 will be awarded to students who give an excellent performance and demonstrate that they have fulfilled the above objectives exhaustively or with few insignificant omissions. The grade 02 will be awarded to students who demonstrate that they have fulfilled the minimum acceptable level of the above learning objectives.

Examination
Portfolio exam:
60% obtained through various graded (and supervised peer-graded) problem sheets and mini-assignments throughout the module.
40% final internal evaluation seminar with oral presentation, peer-evaluation (opponent group), internal critique and discussion departing from the assignment and presentation.
Module 4: Applied Social Data Science Capstone Project

**Aim:** Module 4 aims at providing the student with an opportunity to apply a set of data science methods – a combination of techniques covered in M1-3 as well as other relevant analytical approaches – to an existing empirical problem in an area, which is relevant to the student’s field of study.

**Content:**
Empirical semester project on a programme-relevant theme in collaboration with an external organisation (external partner collaboration is not required but highly recommended, and supported). The project departs from a real-life empirical problem and uses a suitable combination of methods covered throughout the semester (M1-3 and other relevant techniques) to address it. If possible, the analysis is based on real data provided by the collaborating institution, possibly combined with other sources.

In this module, students will – in part independently and partly under supervision – write an empirical semester project (in the optimal case) in collaboration with an external organisation. The length of the project report depends on the group size (maximum of 4 students), with a maximum of 25 normal pages (2400 characters incl. spaces, which equals to approx. 360 words) per student, including references, but excluding appendices.

The semester project can be written (and examined) in **Danish or English**.

**Supervision**
Students will have a main supervisor from their respective master programme, and complementary methods support by the Social Data Science teachers.

**Credits (ECTS):** 15

**Scheduling:**
3rd semester of the master programme, autumn.

**Requirements for participation:**
Successful Completion of M1 - M3.
Learning objectives:
After completion of the module, students are able to define an appropriate problem formulation within their line of study, identify a sophisticated data collection and analysis strategy, carry out the analysis and present their results using state-of-the-art data science approaches, as well as critically self-evaluate their findings. They can select the most suitable among the wide range of methods presented in the modules M1-3, and autonomously apply it to their specific problem.

Knowledge:
- Define relevant real-world empirical problems within organisations.
- Explain the limitations of quantitative analysis on different levels of sophistication.
- Demonstrate knowledge about the choice of ontological and epistemological positions.
- Explain the choice of the methodological implementation.
- Show insights in potential limitations of the undertaken analysis.

Skills:
- Identify and delineate a problem that can be analysed using data science approaches.
- Collect / extract / mine necessary appropriate data.
- Assess the reliability / validity / ethical and legal status / limitations of the data.
- Describe and explore the data.
- Identify and carry out appropriate data preparation and analysis.
- Visualise / communicate the results.
- Reflect on the robustness / limitations / ethical, legal, social consequences regarding the analysis and results.
- Present and discuss results written and orally at an appropriate academic level.

Competencies:
- Initialise, control and complete problem-oriented data science project work.
- Coordinate own resources for the solution of domain-specific related problems.
- Take responsibility for own professional learning and development.

Assessment Criteria
Module 4 is assessed according to the Danish 7-point grading scale. The grade 12 will be awarded to students who give an excellent performance and demonstrate that they have fulfilled the above objectives exhaustively or with few insignificant omissions. The grade 02 will be awarded to students who demonstrate that they have fulfilled the minimum acceptable level of the above learning objectives.

Examination
Oral group examination based on a group project or an individual project (duration depending on group size) with an external co-examiner.