

Master in Data Science

Departments of Informatics & Mathematics

Programme board

Geir Storvik (Chair, MI), geirs@math.uio.no

Thordis Thorarinsdottir(MI), thordist@math.uio.no

Ole Chr. Lingjærde (IFI), ole@ifi.uio.no

Martin Giese (IFI), martingi@ifi.uio.no

(Riccardo de Bin (MI))

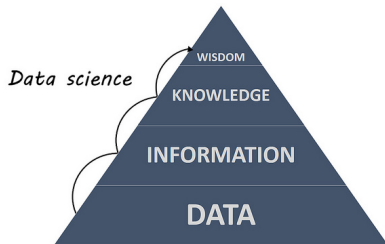


universitetet i Oslo

Master in Data Science at UiO

- ▶ What is Data Science?

The science of extracting knowledge or insight from various types of data.



- ▶ This program: Combines **mathematics**, **statistics** and **informatics** making you ready for meeting the Data revolution.
- ▶ Goal of the program:

Data Science is the science of extracting knowledge or insight from various types of data. The study of Data Science combines mathematics, statistics and informatics making you ready for meeting the Data revolution.



UiO • Universitetet i Oslo

Background

- ▶ Why such a program?
 - ▶ Data Science methodology partly included in many other programmes
 - ▶ Lack of studies focusing on **methodology** in Data Science
- ▶ Why at UoO?
 - ▶ Two centers for research-based innovation
 - ▶ **BigInsight: Statistics**
 - ▶ **Sirius: Informatics**
 - ▶ Both based on problems related to **Big Data**
 - ▶ Both financed by
 - ▶ The Norwegian Research Council (excellent researchers)
 - ▶ The industry (useful research)
 - ▶ A new **Center of excellence: INTEGRAT**: Knowledge-based machine learning
 - ▶ Also many other excellent researchers/groups within topics related to Data Science
 - ▶ **dScience**- Centre for Computational and Data Science



Programme structure

An interdisciplinary background in **statistics** (probability theory, inference, machine learning) and **computer science** (algorithms, visualization, database) in addition to a solid foundation in general **mathematics**.

▶ 1. Semester:

▶ Two **compulsory** courses

STK-IN4300 - Statistical learning methods in Data Science

IN-STK5000 - Adaptive Methods for Data-Based Decision Making

▶ You need to **choose** one more extra course

▶ Many possible courses in both statistics and informatics, depend a bit on specialisation!

▶ First deadline: **August 17!**

▶ Master thesis/supervisor

4. semester	Thesis		
3. semester	Thesis		
2. semester	Course	Course	Course
1. semester	Course	Course	Course



Programme structure

An interdisciplinary background in **statistics** (probability theory, inference, machine learning) and **computer science** (algorithms, visualization, database) in addition to a solid foundation in general **mathematics**.

▶ 1. Semester:

▶ Two **compulsory** courses

STK-IN4300 - Statistical learning methods in Data Science

IN-STK5000 - Adaptive Methods for Data-Based Decision Making

▶ You need to **choose** one more extra course

▶ Many possible courses in both statistics and informatics, depend a bit on specialisation!

▶ First deadline: **August 17!**

▶ Master thesis/supervisor

4. semester	Thesis		
3. semester	Thesis		Course
2. semester	Thesis	Course	Course
1. semester	Course	Course	Course



Choice of specialization

- ▶ Common
 - ▶ Use of **data**
 - ▶ Data **analysis**
 - ▶ Both **machine learning** and **statistical methods**
- ▶ Generic methodology
 - ▶ **Statistics and Machine Learning**
 - ▶ **Database Integration and Semantic Web**
- ▶ Specific types of data
 - ▶ **Data Science and Life Science**
 - ▶ **Language Technology**
 - ▶ **Digital Image Processing**



Database Integration and Semantic Web

- ▶ Focus on **structured** data, **reasoning** and the relation between **data** and the **real world**
 - ▶ Structure of data elements and the relationships among them
 - ▶ combining and preparing data for analysis
 - ▶ machine learning on graph data
 - ▶ relational databases and knowledge graphs
- ▶ Central courses
 - ▶ ***IN4070 - Logic**
 - ▶ Background should be ok for DS students!
 - ▶ **IN4060 - Semantic Technologies**
 - ▶ **IN5040 - Advanced Database Systems for Big Data - 3. semester**
- ▶ Examples of projects
 - ▶ Extract "logical" models of the real world from language models
 - ▶ Can neural networks learn to reason?
 - ▶ Use machine learning to simplify logical models
 - ▶ Use logical models to help machine learning



Language Technology

- ▶ Topics related to Natural Language Processing (NLP), a subfield of Artificial Intelligence enabling computers to ‘make sense’ of human language.
- ▶ Central courses
 - ▶ [*IN4080 - Natural Language Processing](#)
 - ▶ [IN5550 - Neural methods in Natural Language Processing](#)
- ▶ Examples of projects
 - ▶ Create training data and machine-learned models for Sentiment Analysis for Norwegian Text.
 - ▶ Machine learning methods to automatically anonymise text documents containing personal data



Data Science and Life Science

- ▶ Increase the understanding of biological mechanisms while developing better theories and methods in computer science and statistics.
- ▶ Computational and statistical applications in the medical and biological sciences.
- ▶ Combined, the two research areas bioinformatics and statistical genomics cover most of our activities.
- ▶ Central courses
 - ▶ [IN4030 - Introduction to bioinformatics](#)
 - ▶ [STK4051 - Computational statistics](#)
- ▶ Possible 1. semester courses
 - ▶ [*IN4080 - Natural Language Processing](#)
 - ▶ [*STK4021 - Applied Bayesian Analysis](#)
- ▶ Example of projects
 - ▶ Cancer genomics
 - ▶ Machine learning for Immune receptor repertoire analysis + Genome statistics and visualisation



Digital Image Processing

- ▶ Focus on automatic methods for extracting information from images.
- ▶ The main application areas are medical imaging and remote sensing.
- ▶ Deep learning is becoming more and more important and several of our projects include that.
- ▶ Central courses
 - ▶ ~~IN5520 — Digital Image Analysis~~
 - ▶ [IN5400 – Machine Learning for Image Analysis](#)
 - ▶ [*STK4021 - Applied Bayesian Analysis](#)
- ▶ Examples of projects
 - ▶ Robust and secure deep learning
 - ▶ How much should we overparameterize a NN?



Statistics and Machine learning

- ▶ Focus on **statistical** aspects related to data science and machine learning methods
 - ▶ Implementation/testing of a new idea
 - ▶ Evaluation of newly proposed methods
 - ▶ Can lead to new ideas!
 - ▶ Both algorithmic based and model based approaches
 - ▶ Modification of existing methods to big data
 - ▶ Modification of existing methods to new applications
- ▶ Central courses
 - ▶ *STK4021 - Applied Bayesian Analysis
 - ▶ STK4051 - Computational statistics
- ▶ Example of projects
 - ▶ Bayesian Neural networks
 - ▶ Computational efficient algorithms for big data settings
 - ▶ Statistical methods for graphical data



Events

- ▶ Many interesting seminars, see email notifications
- ▶ Special events for Data Science students:
 - ▶ Pizza evening ++
 - ▶ Lunch with waffles
 - ▶ ????

