Introduction to the Master's Programme
Statistics and Machine Learning

Programme director: Bertil Wegmann, STIMA, IDA

Statistics and Machine Learning programme

Aims:

- To build advanced models for explaining complex real-life systems and predicting new events
- To extract, organize and explore large volumes of data
- To learn how to discover important information from large and complex data sets
- To get an in-depth knowledge of models and methods

Competences:

- Machine learning, statistical modeling, programming, big data, data mining, visualization methods, databases etc

The most in-demand tech jobs

Linkedin, most in-demand tech jobs for 2023

- AI and ML specialists will be needed to design and develop algorithms, create data models, and analyze large amounts of data. This field will likely see a rise in demand for roles such as AI engineers, data scientists, and ML developers.
- As data becomes more critical to business operations, the demand for professionals skilled in data analysis and data science will continue to grow.
- Data analysts and scientists will be responsible for analyzing large amounts of data, creating data models, and developing insights that can help drive business decisions. Roles in this field may include data analysts, data scientists, and business intelligence analysts.

Job opportunities

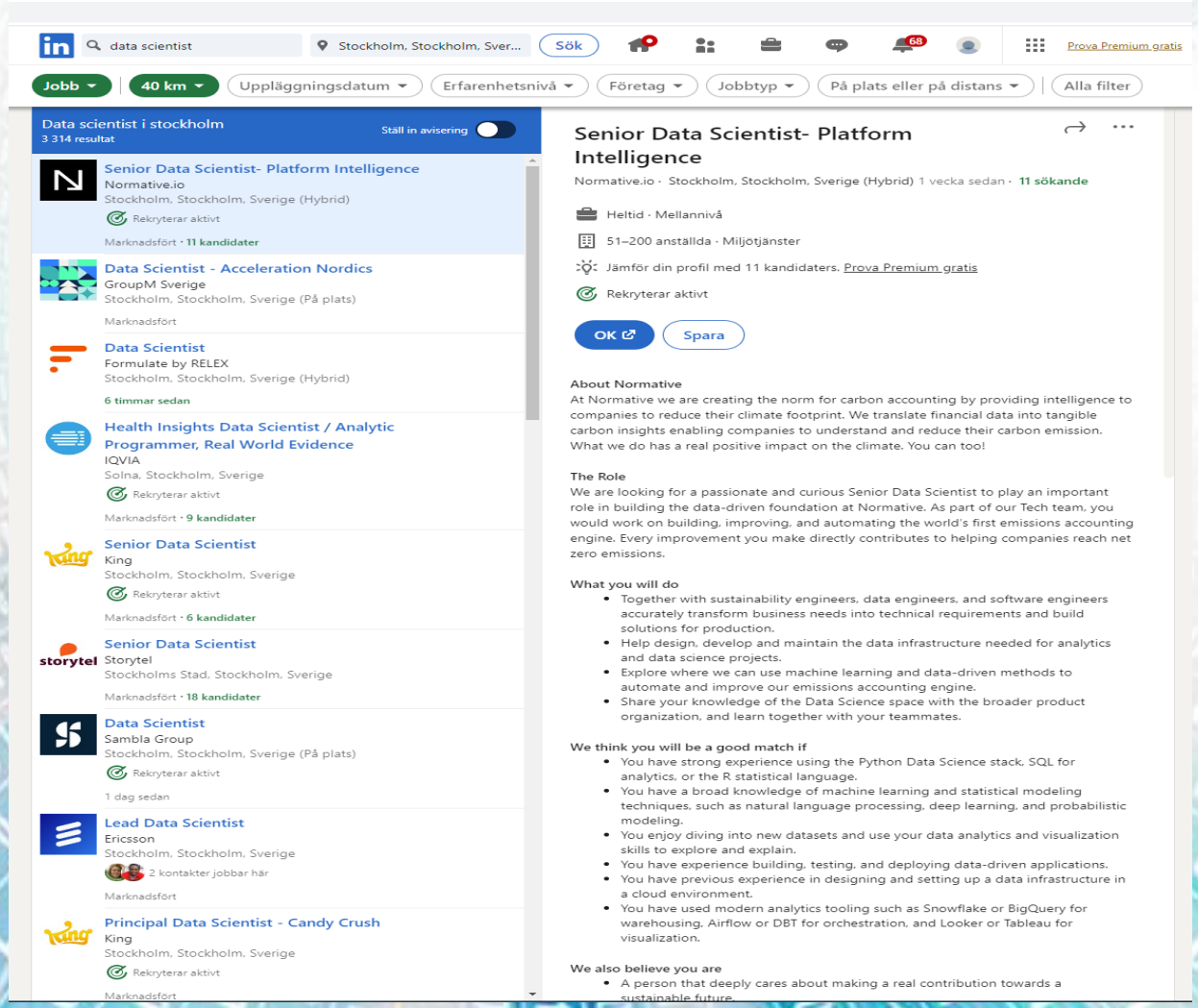
- **Plenty of jobs are awaiting you in Europe and USA**
- Master programme gives excellent background to search **jobs as analyst, data scientist, engineer, manager or consultant** in
  - Industry (IT, telecom, automotive)
  - Health
  - Business (bank, insurance)
  - Economics

…and many other areas where large or complex information systems are involved

Example jobs:

- Machine Learning Engineer, London
- Data Scientist, Amazon
- Data Scientist, IKEA

Many jobs in Sweden!



Master program overview

- **Master program = 120 ECTS credits (https://studieinfo.liu.se/en/program/F7MSL/5305)**
  - **Obligatory courses** (42 ECTS)
    - You must take and finish these courses to get a degree
  - **Introductory courses** (each 6 ECTS, you need to choose 2 out of 3)
    - **Advanced R programming**: recommended for all students who have no extensive programming background
    - **Statistical methods**: recommended to students with a minor background in Statistics, e.g. computer scientists or engineers (check syllabus if you are not sure)!
    - **Visualization:** profile course for static, interactive and dynamic graphics for data analysis
  - **Profile courses**
    - Courses in Statistics that you need in order to get a degree in Statistics.
  - **Complementary courses**
    - Courses in other subjects
  - **Master thesis** (30 ECTS)

- **In order to make a sufficient progress in studies**, you need to obtain 30 ECTS credits at each semester

Selection of introductory courses

- **Introductory courses** (each 6 ECTS, you need to choose 2 out of 3. **Deadline: August 24 at 23:59 (CET)**
  - **Statistical methods/Advanced R programming/Visualization**
- Course plan for each course:
  - https://studieinfo.liu.se/en/program/F7MSL/5305
  - The courses will be added to your space in Lisam
  - Register for these courses after that
- Course selection link

Advanced Academic Studies 732A60

- **Seminar 1**: Introduction to the master's programme.
  Seminar 2*: Systems and software: LISAM, R.
  Seminar 3*: Study advisor and stress management.
  Seminar 4: Writing reports: RMarkdown and LaTeX.
  Seminar 5: Scientific methods and data ethics.
  Seminar 6: Library session. Search and find scientific publications.
  Seminar 7: Summaries and critical reviews. Introduction to the project work.

- **WORKSHOP** at the end

Advanced Academic Studies

- Seminars – Attendance is obligatory
- Course end: October 2023
- Grading for this course: **Pass or Fail**
- Ouriginal is used → Plagiarism is forbidden! (discovered plagiarism implies a request to the disciplinary board)

- **Project work**: writing a paper on a topic, select among given topics (3-4 pages), in *Rmarkdown*

- Workshop with roundtable discussions

Advanced Academic Studies

- Course home page
  - https://www.ida.liu.se/~732A60/index.en.shtml
- Schedule is available on TimeEdit:
  - Timetable

Semester threshold requirements

- At least 6 ECTS credits from passed courses of the first semester, in order to get access to the second semester of the programme.
- At least 30 ECTS credits from passed courses of the first two semesters of the programme, including the course Machine Learning, 9 ECTS credits, in order to get access to the third semester of the programme.
- At least 60 ECTS credits of the programme, including 6 ECTS credits from semester 3 and the course Machine Learning, 9 ECTS credits, in order to get access to the fourth semester of the programme.

Master programme overview

<table>
<tr><th colspan="4">Year 1</th></tr>
<tr><th colspan="2">Semester 1</th><th colspan="2">Semester 2</th></tr>
<tr><th>Period 1</th><th>Period 2</th><th>Period 3</th><th>Period 4</th></tr>
<tr><td>Advanced Academic Studies
(732A60, 3 credits)</td><td rowspan="3">Machine Learning
(732A99, 9 credits)</td><td>Advanced Data Mining
(732A74, 6 credits)</td><td>Big Data Analytics
(732A54, 6 credits)</td></tr>
<tr><td>Visualization*
(732A98, 6 credits)</td><td>Introduction to Python
(732A75, 3 credits)</td><td>Deep Learning
(732A78,3 credits)</td></tr>
<tr><td rowspan="2">Advanced R programming*
(732A94, 6 credits)</td><td>Multivariate statistical methods*
(732A97, 6 credits)</td><td rowspan="2">Bayesian learning
(732A91, 6 credits)</td></tr>
<tr><td>Neural Networks and Learning Systems*
(732A55, 6 credits)</td></tr>
<tr><td>Statistical methods*
(732A93, 6 credits)</td><td>Computational statistics
(732A90, 6 credits)</td><td>Web programming*
(732A56, 6 credits)</td><td>Sports Analytics*
(753A01, 6 credits)</td></tr>
</table>

Master programme overview

<table>
<tr><th colspan="4">Year 2</th></tr>
<tr><th colspan="2">Semester 3</th><th colspan="2">Semester 4</th></tr>
<tr><th>Period 1</th><th>Period 2</th><th>Period 3</th><th>Period 4</th></tr>
<tr><td>Time Series and Sequence Learning*
(732A80, 6 credits)</td><td>Text Mining*
(732A81, 6 credits)</td><td colspan="2" rowspan="7">MASTER THESIS
(732A64, 30 credits)</td></tr>
<tr><td>Probability theory*
(732A63, 6 credits)</td><td>Bioinformatics*

(732A51, 6 credits)</td></tr>
<tr><td>Advanced Machine Learning*
(732A96, 6 credits)</td><td rowspan="2">Database Technology*

(732A57 , 6 credits)</td></tr>
<tr><td>Visualization*

(732A98, 6 credits)</td></tr>
<tr><td colspan="2">Research project* (732A76, 6 credits)</td></tr>
<tr><td colspan="2">Decision theory* (732A66, 6 credits)</td></tr>
<tr><td colspan="2">EXCHANGE STUDIES*</td></tr>
</table>

Obligatory courses

- Advanced Academic studies
- Computational statistics
  - Random number generation, MCMC
- Machine Learning
  - Predictive modelling: Ridge regression, Decision Trees, basic neural networks, support vector machines etc
- Advanced Data Mining
  - Clustering and association analysis, focus on algorithms
- Deep Learning
  - Deep NNs, Convolutional NNs, Autoencoders, GANs, Recurrent networks
- Bayesian learning
  - Using prior knowledge to make better decisions and inference
- Big Data Analytics
  - Hadoop, Spark, scaling up machine learning
- Introduction to Python
  - Python environment. Data structures. Basic Language elements

Profile courses

- Advanced Machine Learning
  - Bayesian networks, reinforcement learning, particle filtering
- Visualization
  - Static, interactive and dynamic graphics for data analysis
- Time Series and Sequence Learning
  - ARIMA models, state-space models, Neural Networks for sequences
- Probability theory
  - Multivariate random variables, transforms, order statistics, convergence. Necessary for PhD studies.
- Multivariate statistical methods
  - Principal components, factor analysis, canonical correlation
- Decision Theory
  - Bayesian hypothesis evaluation, Decision theoretic elements, Utility and loss functions, Graphical modelling, Sequential analysis
- Bioinformatics
  - sequence data, microarray data and trait data. Evolutionary tree reconstruction methods.

Complementary courses

- Neural networks and learning systems
  - Given by **Department of Biomedical Engineering** Advanced neural networks, kernel methods, reinforcement learning, genetic algorithms
- Web programming course
  - HTML, XML, PHP
- Research project
  - Implement and evaluate a statistical or machine learning task specified by a university researcher
- Text Mining
  - Extracting text data from different sources and analyze linguistically and with statistical tools
- Database technology
  - Relational databases, relational algebra, SQL, query optimization

Other information

Master program's homepage(schedule, courses, news…):
Home page on Lisam

Email to staff: *name.lastname@liu.se*

- *Example: bertil.wegmann@liu.se*

Webpages of courses: www.ida.liu.se/~course_code/

- This course: Courseinfo 732A60
- Search for courses: https://studieinfo.liu.se/en/

Research Seminars

- The LiU Seminar Series in Statistics and Mathematical Statistics
- IDA Machine Learning Seminars

Course registration

- To participate in an exam and get credits for a course, you must register for it.
- **Register for exactly 120 ECTS (Swedish language courses not included)**
- Registration is done in Lisam:

Exam registration

- If you have problems with registration, contact our administrator Erika Larsson (Erika.Larsson@liu.se)

LiU-Account and personal number

- It is necessary for you to get a LIU-account **as soon as possible** (Student house)
  - Access to Lisam
  - Course registration
  - Access to course materials
  - Access to department computers

- If you are not Swedish, **it is very important** to get a **Personal Number** at the Tax office:
  - Address: Kungsgatan 27, Linköping
  - Needed for medical help

Lectures, Labs, Seminars

- **Lectures**: normally presented in PowerPoint/PDF, later available either at the course page or LISAM. Attendance is typically not obligatory.
- **Labs:** typically computer exercises done individually or in groups of two. Attendance is typically not obligatory. A written report should normally be submitted.
- **Seminars:** Discussions of theory and labs, student presentations. Attendance typically obligatory.

Plagiarism in lab works

- **I have found some solutions to the lab on the Internet. May I use them in my report? Answer: NO**

- **The lab was very difficult and I managed to solve only some part of it before the deadline. What should I do?**
  - I will look at the solutions of my classmate(s) and will try to complete the lab in time **NO**
  - I will only look at the solutions of my classmate that I didn't manage to do myself and complete the lab in time **NO**
  - I will submit incomplete lab solutions **YES!**
  - I will talk with my group members about how the assignments can be solved **YES!**

- **I am not sure whether my lab answers are correct or not. What should I do?**
  - Compare my solutions with solutions of my classmates before I submit **NO**
  - I will submit my solution as it is **YES!**

Academic norms

- **Group works**
  - Every student should contribute equally to the lab work
    - Question: Can I write codes and my lab partner do analysis and interpretations? **NO**
    - Question: My group member works much faster/slower than me. How do we make a group report? **Try to find a pace that works for both.**
- **Don't share your solutions online or within larger groups**
  - Destroys a normal learning process for the students
  - Might lead to intentional/unintentional plagiarism

Exams and Credits

- **Exams**
  - Each course has 1 exam and 2 re-exams (exception: project course)
  - **You must register for the written or computer exam at least 10 days in advance.**
    - **If you forget, you will not be allowed to attend the exam.**
  - Exam results **may not** be improved → if you aim for a higher grade and feel that you perform bad during an exam → submit empty pages/files
  - Exam results should normally be available after **2 weeks**

- **Credits**
  - Most courses have separate credits for labs (or project) and for the exam
  - Credits for some courses can be obtained only after you are completely done with the course

Course evaluation

- **EvaLiUate**: course evaluation system at LiU
  - You evaluate the courses you have done
  - Sent via email
  - The surveys are anonymous!
  - Very important for improvements of courses – please answer these surveys!

- You can contact the study advisor to discuss your current studies and plan the coming studies.

Schedules of the courses

- Some schedules are on the course homepages
- Most schedules accessed via TimeEdit:
  - https://cloud.timeedit.net/liu/web/schema/
  - Type the course name and go

How to find a room

- LiU Maps

Useful links

- **Home page on Lisam**
- Practical Guide
- Welcome to LiU
- General information about the programme

Questions

- Questions related to the program?
  - Contact Bertil Wegmann https://liu.se/en/employee/berwe48

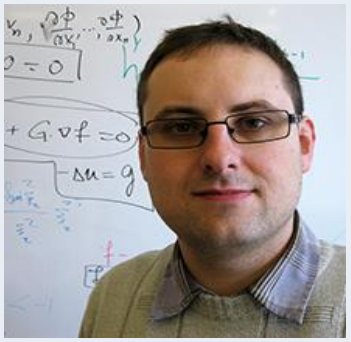
Personnel at Division of Statistics and Machine Learning Department of Computer and Information Science

	Name		Name
	**Bertil Wegmann** *Associate professor* *Programme Director "Statistics & Machine Learning"*		**Anders Eklund** *Senior associate professor*
	**Frank Miller** *Professor*		**Johan Alenlöv** *Lecturer*

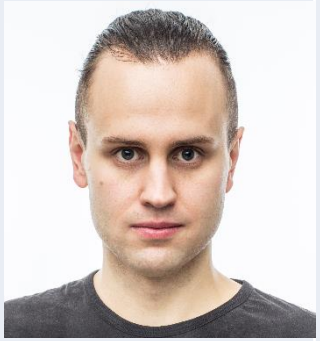
Personnel at Division of Statistics and Machine Learning Department of Computer and Information Science

	Name		Name
	**Annika Tillander** *Associate professor* *Responsible for the bachelor programme*		**Hector Rodriguez** *Lecturer*
	**Erika Larsson** *Administrator*		**Katarina Isotalo** *Study advisor*

Personnel at Division of Statistics and Machine Learning Department of Computer and Information Science

	Name		Name
	**Oleg Sysoev** *Senior associate professor*		**Fredrik Lindsten** *Senior associate professor* *Head of Division*
	**Jose Pena** *Senior associate professor*		**Krzysztof Bartoszek** *Senior associate professor*

Personnel at Division of Statistics and Machine Learning Department of Computer and Information Science

	Name		Name
	**Linda Wänström** *Associate professor*		**Anders Nordgaard** *Adjunct Senior Lecturer*
	**Jolanta Pielaszkiewicz** *Associate professor* *Director of Studies*		**Josef Wilzén** *Lecturer*

Other teachers

	Name		Name
	**Patrick Lambrix** *Professor, ADIT*		**Marco Kuhlmann** *Professor, HCS*
	**Mohammad Seidpisheh** *Lecturer, STIMA*		