MSc Programme in Data Science
University of Helsinki

9 May 2017

Hannu Toivonen
hannu.toivonen@helsinki.fi
Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren’t seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone. So you just stand in the corner sipping your drink—
HOW TO PREVENT SHIPWRECKS WITH THE HELP OF BIG DATA?

Data science combines computer science and statistics to solve exciting data-intensive problems in industry and in many fields of science. As data is collected and analysed in all areas of society, demand for professional data scientists is high and will grow higher. This interdisciplinary Data
Data science

- Extraction of knowledge or insights from data
- About processes and systems
- An interdisciplinary field
- A continuation of some of the data analysis fields such as
  - statistics
  - data mining and machine learning
  - predictive analytics
- similar to Knowledge Discovery in Databases (KDD)
"The key word in ‘Data Science' is not Data, it is Science"

- Jeff Leek
of Data Science:

- **Algorithms** for data analysis
- **Architectures** for large/distributed data
- **Applications**
Data Science MSc Programme at the University of Helsinki

- A new multidisciplinary degree programme
- Trains new generations of data scientists for the industry, academia, and administration
- Offered jointly by
  - Department of Computer Science
  - Department of Mathematics and Statistics
  - Department of Physics
  - Helsinki Institute for Information Technology HIIT
  - Helsinki Institute for Physics HIP
- Based on their own research in data science
Additional offering in CS, math, stats

Other studies at UH/Industry
MSc degree content
Overview

1. Compulsory courses (35 cr)
2. Elective courses (at least 20 cr)
3. MSc thesis (30 cr)
4. Other studies (up to 35 cr)
MSc degree content
1. Compulsory courses (35 cr)

- Introduction to Data Science
- Introduction to Machine Learning
- Distributed Data Infrastructures
- Bayesian Inference
- Academic Skills for Data Science
- Data Science Project
- Data Science Seminar
MSc degree content

2. Elective courses (>= 20 cr)

At least four courses, in these broad areas:

- Machine learning
- Statistical data science
- Data science infrastructures
- Computers and cognition (AI)

(see following slides for courses)
MSc degree content

2. Elective courses

MACHINE LEARNING

• Advanced Course in Machine Learning
• Probabilistic Models
• Computational Statistics I
MSc degree content

2. Elective courses

STATISTICAL DATA SCIENCE

- Advanced Statistical Inference
- Advanced Course in Bayesian Statistics
- Computational Statistics I
- High Dimensional Statistics
- Spatial Modelling and Bayesian Inference
MSc degree content
2. Elective courses

DATA SCIENCE INFRASTRUCTURES
• Big Data Frameworks
• Introduction to Big Data Management
• Cloud and Edge Computing
• Tools for High Performance Computing
• Scientific Computing III
MSc degree content

2. Elective courses

COMPUTERS AND COGNITION

• Introduction to Artificial intelligence
• Computational Creativity
• Interactive Data Visualization
• Philosophy of Artificial Intelligence
• Perception Communication and Cognition
• Cognition and Brain Function
MSc degree content
2. Elective courses

ALGORITHMIC DATA SCIENCE

- Design and Analysis of Algorithms
- Data Compression Techniques
- String Processing Algorithms
Data science study packages for other programmes

- Basics of Data Science (15 or 25 cr)
- Basics of Artificial Intelligence (15 cr)
- Basics of Data Science and Artificial Intelligence (25 or 35 cr)

- Machine Learning (15 cr)
- Data Science Infrastructures (15 cr)

- All substance courses are open (but subject to prerequisites)
Admission criteria (formal)

- A first-cycle degree in computer science, statistics or mathematics, or
- A first-cycle degree in an applied field of data science or another applicable degree and a minimum of 25 credits in computer science, statistics and mathematics.
- In all cases, applicants are required to
  - be proficient in programming (>= 10 cr programming courses)
  - and have completed at least 10 credits in statistics and the probability calculus.
Admission criteria (in practice)

In practice, the bare minimum is

- at least 10 cr programming/computer science,
- at least 10 cr in statistics and probability calculus, and
- at least 5 cr mathematics

(but for successful studies more is better…)
Data Science MSc Programme

Some facts

- Teaching language: English
- First call: winter 2016/17
  - 84 external applications, admitted 23
  - Own students can transfer from Aug 2017 on
  - Admission quota: max 50 (soft)
- Tuition fee: 0 EUR (EU) / 15000 EUR (non-EU)
HiDATA: Helsinki Center for Data Science

HiDATA

- A new multidisciplinary center for data science
- To be established in 2017
- Joint effort between University of Helsinki and Aalto University
- A large international network with academia, research institutes and industry
Data science research at the University of Helsinki
A Sample of Research Groups Working in Data Science

- Adaptive Computing
- Collaborative and Interoperable Computing
- Collaborative Networking
- Content-centric structures and networking
- Computational Linguistics
- Combinatorial Pattern Matching
- Constraint Reasoning and Optimization
- Complex Systems Computation
- Discovery group: Data Mining and Computational Creativity
- Genome-scale Algorithmics
- Information, Complexity and Learning
- Machine Learning
- Multi-source Probabilistic Inference
- Neuroinformatics
- Probabilistic Mechanistic Models for Genomics
- Practical Algorithms and Data structures on Strings
- Secure Systems
- Sums of Products
- Ubiquitous Interaction
- Wireless Internet
International Top Research - Examples

- Breakthrough in scalable methods for strings (text, biology) in 1980s
- Pioneering work in the area of data mining since 1990s
- Data analysis applications in automated creativity, biology, ecology, linguistics, literature, medicine, …
- Pioneering contributions to early mobile computing between 1995-2005 for wireless data transport and performance enhancing proxies
- Pioneering contributions in the development of Information Centric Networking (ICN); winner of the EU Future Internet Award in 2013.
World-wide Interest in Media - Examples