Data Science Education
@ AAU
Presentation for Aftager Panel, Friday 18.11.16

Panagiotis Karras
Christian S. Jensen
Christian Thomsen
Bent Thomsen
Conclusions from 2015 Research Evaluation:

- Furthermore, society needs people that know how to handle data, so it is important to educate students in Data Science.
- The recommendation is to start a Data Science programme as soon as possible.
What is Data Science?

... a discipline that incorporates varying degrees of Data Engineering, Scientific Method, Math, Statistics, Advanced Computing, Visualization, Hacker mindset, and Domain Expertise.

Source: Wikibooks

... an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, which is a continuation of some of the data analysis fields such as statistics, machine learning, data mining, and predictive analytics, similar to Knowledge Discovery in Databases

Who needs Data Scientists?

According to Forbes there will be a need for 1,000,000 Data Scientist by 2018.

Data Scientist is the best job in America, according to Glassdoor, with median base salary $117K, and high career prospects.

Rambøl analysis for ITU note that in the Spring 2016 there were 270 job adds in DK for DS (which could not be covered by statisticians or CS).

I bogen “Data - virksomhedens nye grundstof” skriver Mikkel Holm Sørensen og Simon Bentholm: “Data er ikke længere blot et analytisk instrument, men i stigende grad selve grundstoffet i virksomhedens værdiskabelse på linje med teknologi, talenter og patenter.” (Gyldendal 2013)
How does a Data Science education differ from existing ones?

In comparison to Computer Science (DAT):

*More emphasis* on Statistics, Modeling, understanding data source domains.

*Less emphasis* on SW Engineering, Formal Methods, Prog. Lang. & Compilers

In comparison to Statistics (STAT):

*More emphasis* on Programming, Machine Learning, Algorithms, Systems

*Less emphasis* on proofs, theory, mathematical foundations
Data Science Educations elsewhere

USA:
- Stanford
  BSc in Mathematical and Computational Science
  [https://mcs.stanford.edu/academics/requirements/core-courses](https://mcs.stanford.edu/academics/requirements/core-courses)
  MSc in Statistics: Data Science
- NYU - MSc in Data Science
- U Washington - MSc in Data Science
  [https://www.datasciencemasters.uw.edu/](https://www.datasciencemasters.uw.edu/)
- Columbia - MSc in Data Science
- CMU - MSc in Computational Data Science
  [https://mcds.cs.cmu.edu/](https://mcds.cs.cmu.edu/)

UK:
- The Alan Turing Institute
  Cambridge, Oxford, Edinburgh, UCL, Warwick
  Offers MPhil/PhD degrees in DS
- Imperial College
  Data Science Institute
  [MSc in Statistics](#)
  [MSc in Computing (Machine Learning)](#)
  [MSc in Business Analytics](#)
  [MRes Biomedial Research (Data Science)](#)

Missing: Dedicated BSc+MSc program
Data Science Educations in DK

DTU: DS “profile” via 3 master programs

DIKU: DS “track” in CS masters program

ITU: bachelor program proposal (submitted 3.10.16)

- Want to establish first DS bachelor program in DK
- BSc in DS can go onto master in
  - SW or Business development at ITU
DTU master’s program profile

45 ECTS courses


15 ECTS M.Sc thesis in the area of data analysis

Master thesis has to cover at least two of the 4 main topics
(Data origins and collection, Data storage, Analytics, Consumers)
Overview of ITU BSc in DS

- Introduction to Data Science and Programming (15 ECTS)
- Applied Statistics (7.5 ECTS)
- Data Science in Research, Business and Society (7.5 ECTS)
- First Year Project (15 ECTS)
- Algorithms and Data Structures (7.5 ECTS)
- Linear Algebra and Optimization (7.5 ECTS)
- Machine Learning (15 ECTS)
- Data Management (7.5 ECTS)
- Network Analysis (7.5 ECTS)
- Second Year Project (15 ECTS)
- Data Visualization and Data-driven Decision Making (7.5 ECTS)
- Technical Communication (7.5 ECTS)
- Security and Privacy (7.5 ECTS)
- Reflections on Data Science (Videnskabsteori)(7.5 ECTS points)
- Bachelor Project (15 ECTS points)
Data Science activities @cs.AAU

Very strong research community (ranked 6th in Europe)
InfinIT interest group on Big Data & Business Intelligence
IT Vest Data Science og Big Data Fagpakke on offer in 2017

- Funding for person with data science background from IT Vest
Data Science activities elsewhere @AAU

Mathematics

Transport (CITS)

Energy

Finance

Health

Bio

Mediaology
### Existing Courses that could be used

<table>
<thead>
<tr>
<th>MAT</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete mathematics</td>
<td>Programming Paradigms</td>
</tr>
<tr>
<td>Linear Algebra</td>
<td>Data structures &amp; Algorithms</td>
</tr>
<tr>
<td>Probability Theory</td>
<td>OOP</td>
</tr>
<tr>
<td>Optimization</td>
<td>Database Systems (&amp; advanced)</td>
</tr>
<tr>
<td>Statistics</td>
<td>Machine learning (&amp; advanced)</td>
</tr>
<tr>
<td>Applied Statistics</td>
<td>Web intelligence</td>
</tr>
<tr>
<td>R programming</td>
<td></td>
</tr>
<tr>
<td>Data Mining</td>
<td></td>
</tr>
<tr>
<td>SuperComputing</td>
<td></td>
</tr>
</tbody>
</table>
New courses (or courses we need to find)

Data Analysis & Visualization

Data Privacy & Security

Ethics of Big Data (Technoanthropology degree?)

Application area courses:
- Finance
- Health
- Transport
- Energy
- Biomed

Network Analysis

Scalable Data Systems & Algorithms

Semantic Web

Functional Programming

High-dimensional statistical methods

Stochastic processes

Social Network Analysis
What students would a DS education attract?

1. Those inclined towards Science and Math, who might otherwise study in Aarhus.
2. Those who wish a Data Science degree, who might otherwise study in Copenhagen.
3. Those who seek an Applied Math degree.
Conclusions

We should develop a full BSc+MSc Data Science curriculum!

Should, as iDx, aim at attracting new students (not steal from CS or SW!)
Developing the programme should follow the exemplar process on iDx