Welcome to
Applied Data Science!

prof. dr Arno Siebes

Algorithmic Data Analysis Group
Department of Information and Computing Sciences
Universiteit Utrecht

Welcome to Utrecht University

For those of you new to Utrecht

Welcome!

Utrecht University

- ▶ was founded in 1636
- ▶ has 6700 staff members and 30,000 students
- ▶ comprising 7 faculties
- ▶ with 49 Bachelor and 148 Master programs
- ▶ has a budget of 810 M€
- ▶ and a Shanghai ranking of 52
  - ▶ a fact that is mostly irrelevant to students
  - ▶ but makes the board very happy.

ADS is New

Among the 148 Master programs of Utrecht University

Applied Data Science is the youngest

The program was officially accredited on July 29, 2020

just in time for you to start as its first generation of students

We are all *pioneers*

both the students and the staff

New Means Problems

Don't worry, the program was carefully designed

▶ but given the difference between theory and practice
there will be mishaps

Sometimes we – the staff – will notice this first

▶ and we will try to remedy before you notice

Sometimes you'll be the first to notice

▶ tell us!

Together we can ensure that
pioneering does not mean hardship

Data science is fun
and so should be your program

ADS is Exciting

Data Science is an exciting new venture
data scientist has been called the sexiest job

In a world that is awash with data
data scientists – i.e., you! – are expected to lead the way

In all nooks and crannies of society
from commercial to idealistic, from andrology to zoonosis

(hence the accreditation)

Exciting Means Fast

Exciting areas are invariably
areas that develop fast

The program gives you solid foundations but we cannot cover everything
and there will always be new things that you'll need to learn on your own

If there is, however, something you would like to hear about
tell us and we'll see how we can cover it.

ADS is Multidisciplinary

Data science is inherently multidisciplinary

it builds on AI and computer science, domain sciences, and statistics

The problems you will solve

are seldomly, if ever, formulated as data science problems

You will have to:

transform a domain problem in a data science problem, find and construct the relevant data set, solve the data analysis problem on that data set, and translate the resulting analytical solution into a solution for the domain problem.

Multidisciplinary is Complex

There is no-one at Utrecht University (or any other university)

who is an expert on all the topics we cover

Hence, the program was designed by and is now governed by a program council

with experts from the 5 collaborating faculties

Together we cover the necessary expertise.

But this does not mean that the program is perfect

if you spot a hole, let us know.

ADS is Tough!

Its broad scope makes Applied Data Science

a tough master program

Most of you have a bachelor degree in a domain science

but not all applications will be from your domain science

All of you know some statistics and have some programming experience

but you will have to master a plethora of tools, from algorithms to z-transforms.

No formal class and assignments done means

you play with data

Practice Makes Perfect

Practical sessions to help you master techniques
are sprinkled throughout the program

But playing with a new technique once
is not enough to make you certified

While the 10,000 hours arguments are mostly nonsense
playing with data often really helps

As our eastern neighbours say it:
Übung Macht den Meister

ADS is Unique

Applied Data Science is a unique program

it is an educational collaboration of 5 faculties that have a research collaboration one of Utrecht University's focus areas: Applied Data Science

First a common core (Faculties of Science and Social and Behavioural Sciences)

then elective courses on data science in context (Faculties of Humanities, Geo Sciences, Health, and Social and Behavioural Sciences)

Its focus on data science in context
is unique world wide

Unique is Useful

Most Data Science programs

train people to become experts in algorithms and/or statistics

Domain experts are a dime a dozen

turned out in huge quantities by a plethora of bachelor and master programs

But these two communities

do not have a common language

At the end of the program

You will speak both languages

ADS is Collaboration

Data scientist
almost never work alone

You'll learn to speak two languages because you will often speak to people from both communities
that is the only way to ensure that a technical solution solves a practical problem

Collaboration is integral to the program
you all have a unique background, share this with your fellow students!

And then there was a Virus

Our aim was to give you a big room throughout the first block
so that you could work together, discuss, and help each other to master your common core

Unfortunately, SARS-CoV-2
threw a spanner in those works

Your safety and well being as well as ours
allows only for very limited possibilities to meet in real life

I hope you will all use
the online equivalents we offer you extensively

The Program

I already briefly discussed large parts of the program
- ▶ and you will learn more details later today

But there a few components that I would like to talk to you about
- ▶ the thesis project
- ▶ the practice colloquium
- ▶ the ethics colloquium

Thesis Project

The program ends with a thesis project

your first(?) full data science project

Such a project always aims to solve a *domain problem*

and it may involve more than one of the contexts you have seen before (e.g., epidemiology often has a geographic component) or a completely new context (e.g., on government)

Projects may involve one or more students, but you are always graded on *your* contribution

we want to make sure that you can have a successful career as an applied data scientist

Data Science Practice Colloquium

I already alluded to the fact that

the world of applied data science is much bigger than we can cover in one year

After graduating

you may never work in one of the contexts in which you learned data science in the program!

To help you choose

the data science practice colloquium series is there to show you even more examples of data science in practice.

Ethics

In many applications,

your solution will have an impact on people

Is that impact reasonable, legal, ethical?

to help you think abut such issues, we have the

Data Science Ethics Colloquium

There is no algorithm that decides whether or not something is ethical

but by discussing cases you'll develop your own antenna

ADS is Fun!

Trust me

- ▶ I've done my first big data science project in 1984
    - ▶ and being completely untrained in the area (but, hey, I'm a mathematician) I made every mistake in the book
- ▶ I've been doing research in the area since 1994
- ▶ I co-founded a data science company in 1995
    - ▶ which by way of SPSS became part of IBM

While my expertise is in developing new algorithms and foundations

applications are what makes Data Science Fun!

So

Have Fun!

Two Very Important People

Finally, I'll introduce two very important people to you:

Dr. Mel Chekol
- ▶ Mel is the coordinator of the program
- ▶ coordinator.ads@uu.nl

Yvonne Tromp
- ▶ Yvonne is the study advisor for the program
- ▶ and she will introduce herself next.