

A decorative graphic on the left side of the slide, consisting of a network of white lines and circles on a blue gradient background, resembling a circuit board or data flow diagram.

DATA SCIENCE – WHAT IS IT AND WHAT'S IN IT FOR PROGRAMMERS

Andrzej Prokop
PhUSE SDE 2018-10-12

Presentation plan

- Data Science and Machine Learning: Principles
- The boom of AI in the industry
- What's there for Programmers?
- Demo
- How can you become a Data Scientist?

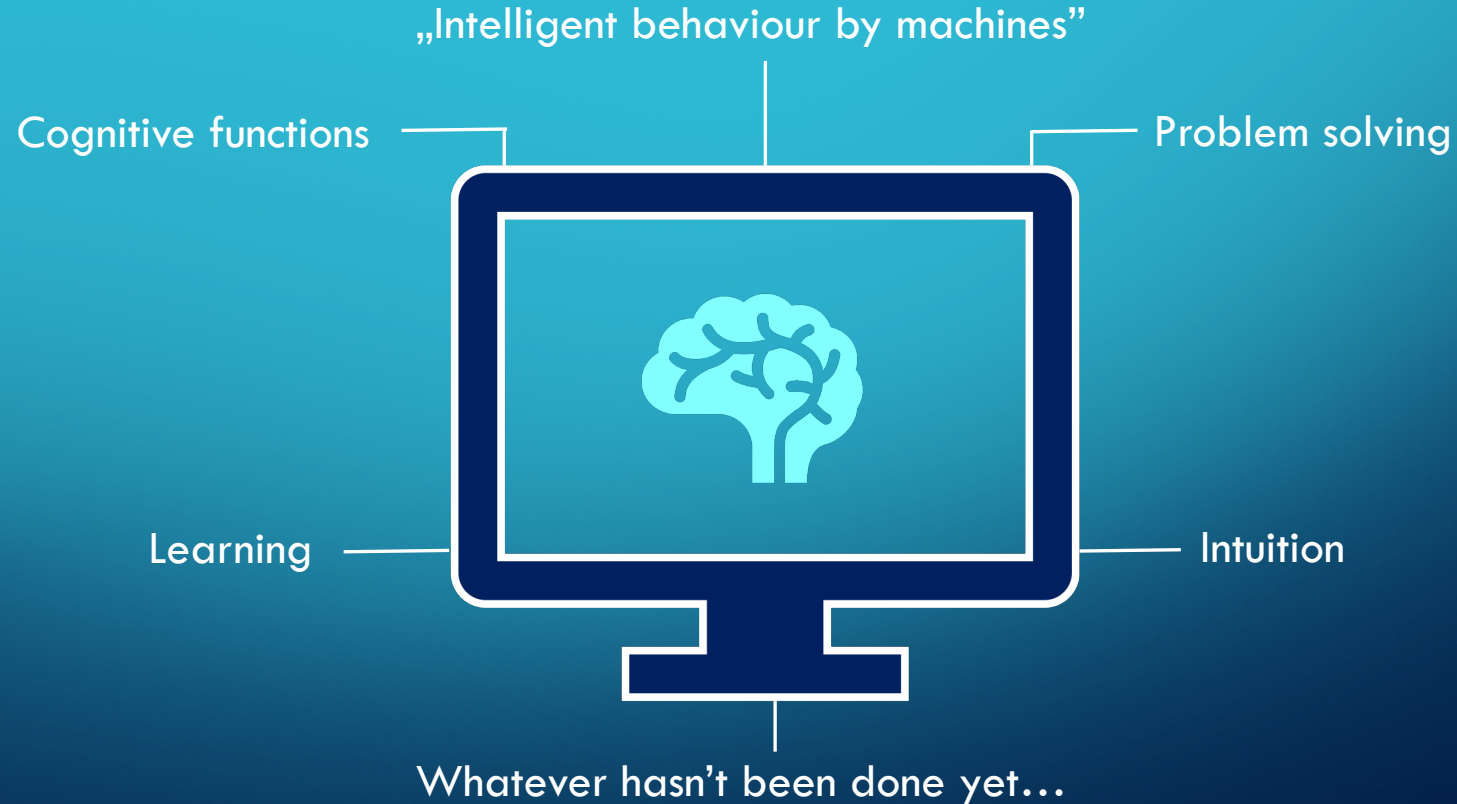
WHO IS A DATA SCIENTIST?

*„A programmer fascinated by statistics
or
a statistician who loves programming”*

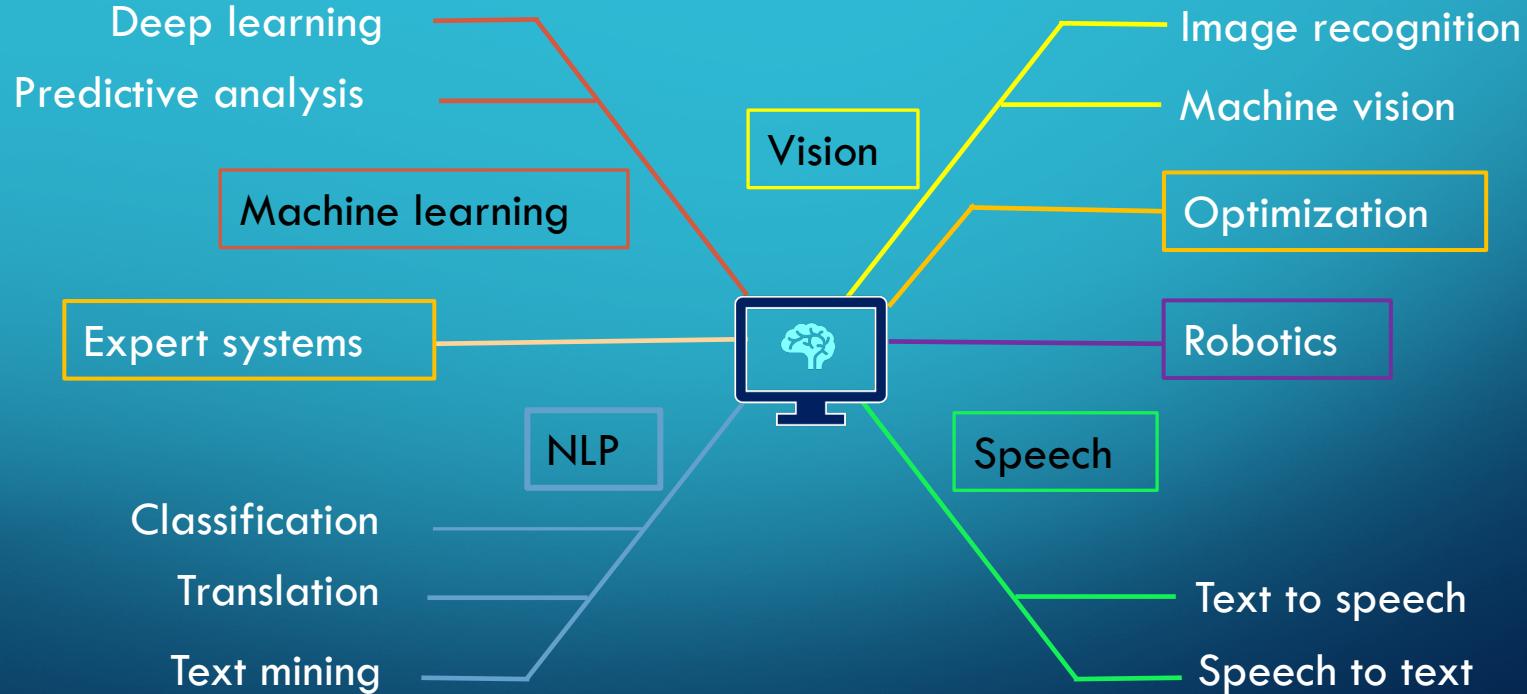
~ The Internet

- Data Science and Machine learning: Principles

AI – WHAT IS IT?



AI – WHAT IS IT?

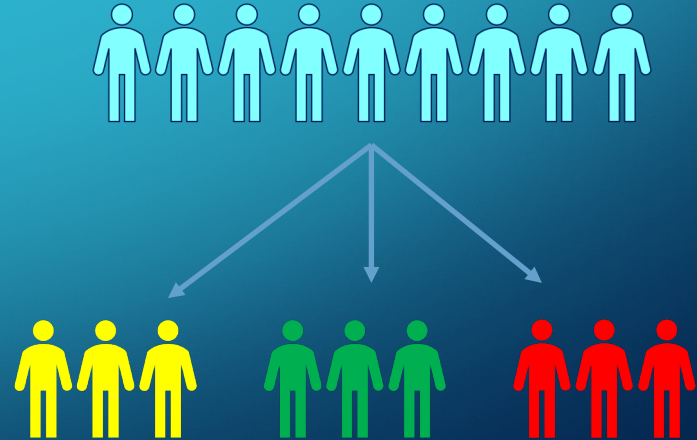


TRAINING AN ALGORITHM

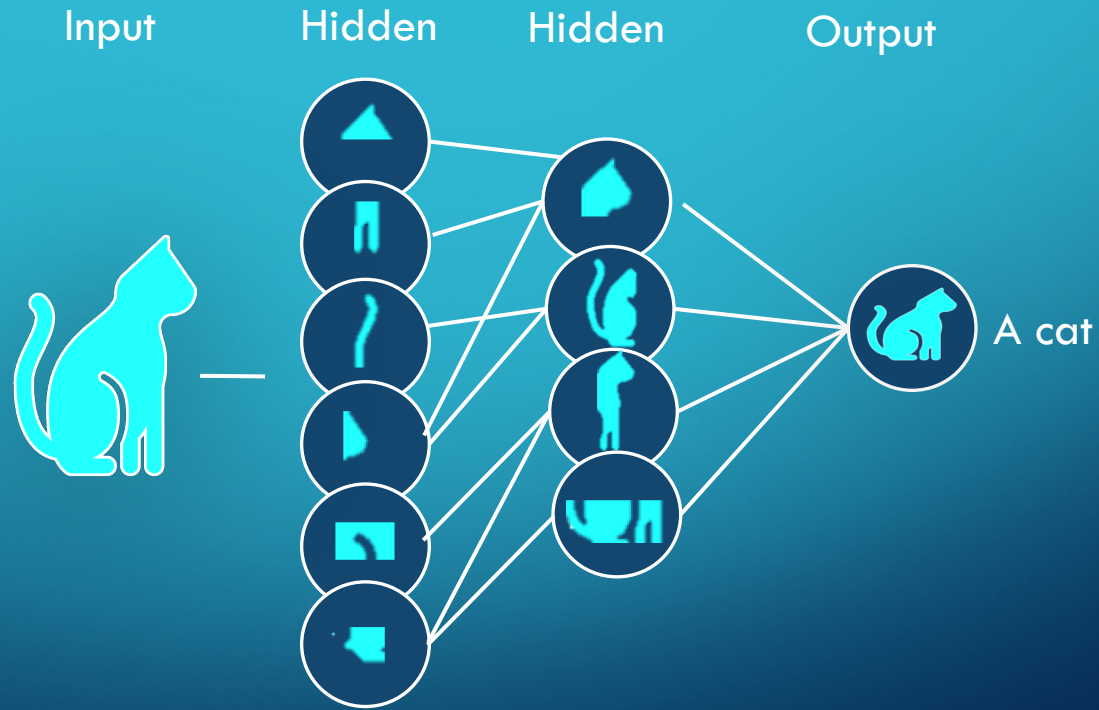
SUPERVISED



UNSUPERVISED

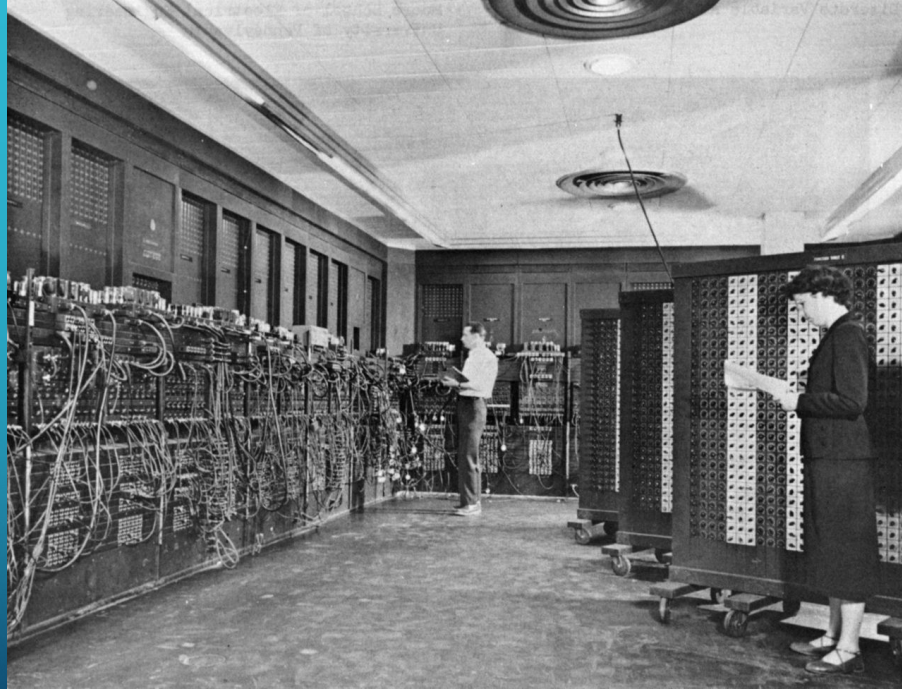


ARTIFICIAL NEURON



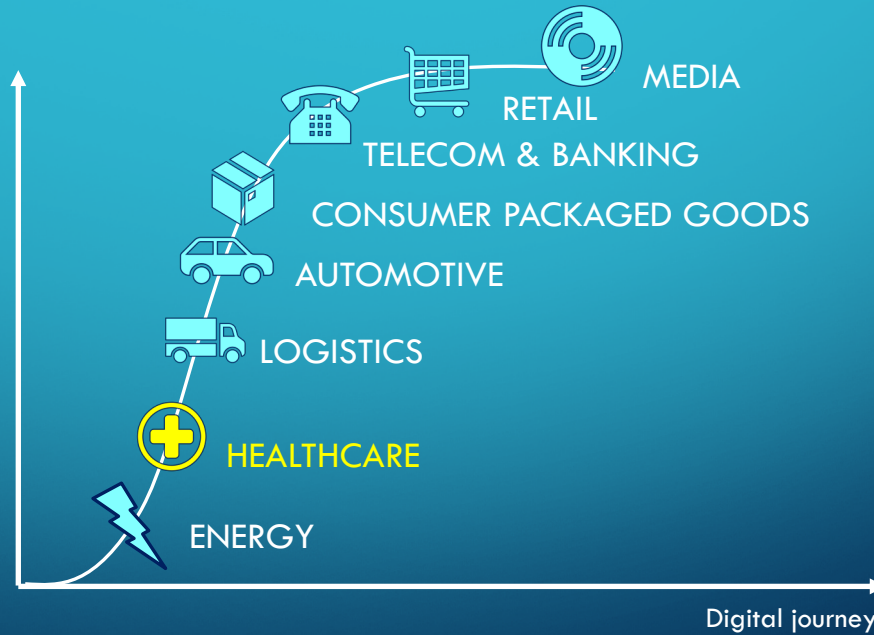
- The boom of Data Science in the industry

WHY NOW?



Source: wikipedia

DIGITIZATION ACROSS INDUSTRIES



- What's there for Programmers

POTENTIAL IN HEALTHCARE

Drug development

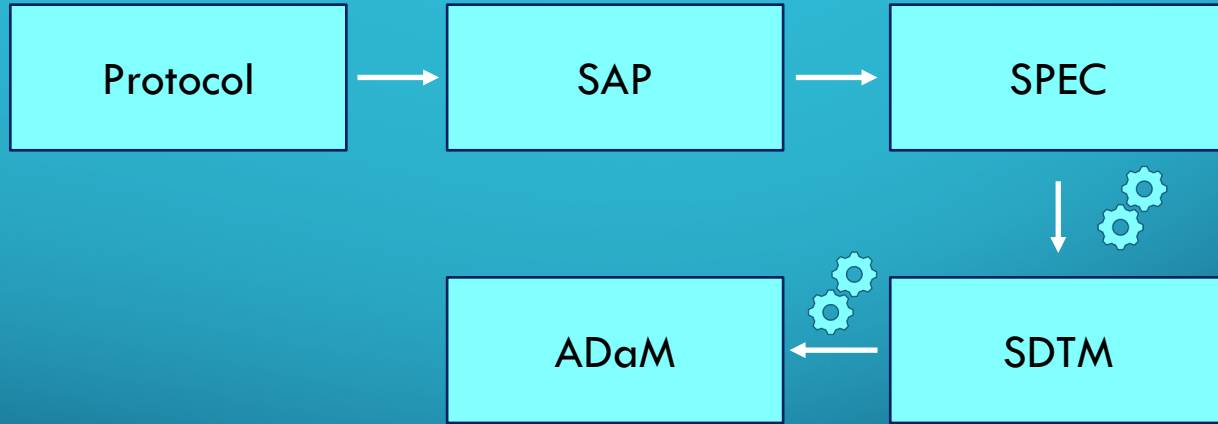
Inclusion-exclusion criteria



Exploratory analyses

Tools for automation

SCENARIOS OF AUTOMATION IN PROGRAMMING



Typical areas for automation

SCENARIOS OF AUTOMATION IN PROGRAMMING

Corporate/Standard

Tools & Macros



SDTM



ADaM

SPEC



Study-specific

Manual pgm writing

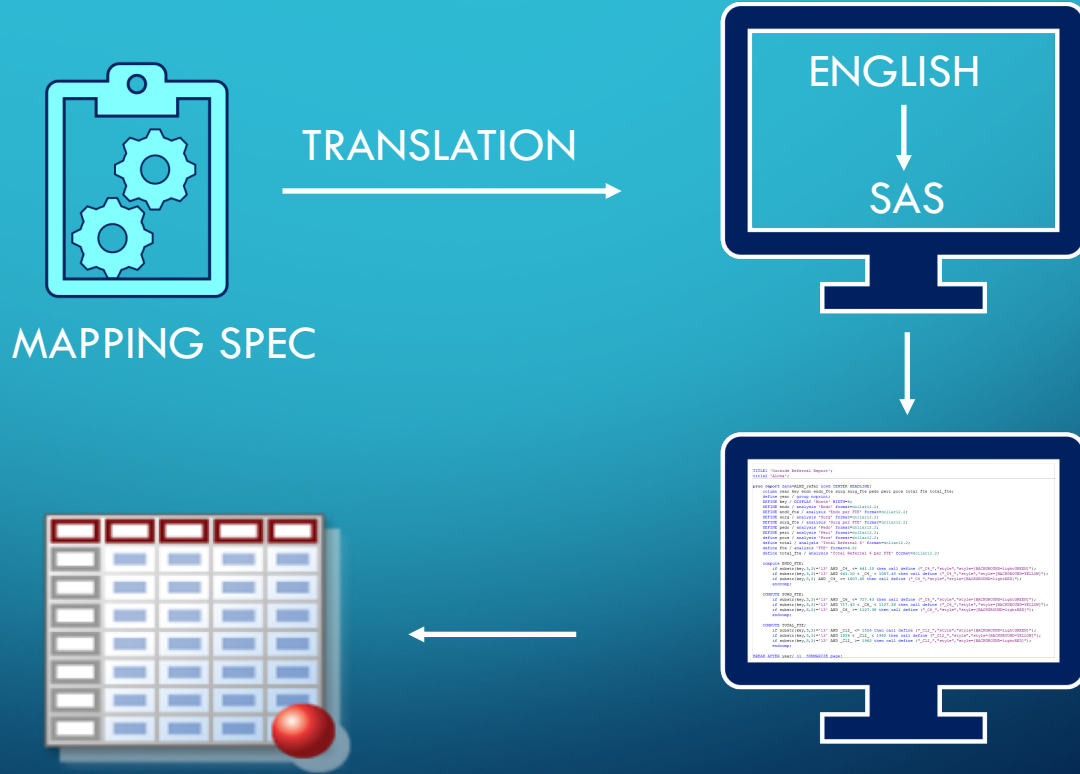


SDTM



ADaM

SCENARIOS OF AUTOMATION IN PROGRAMMING



- DEMO

WHAT HAVE WE DONE?



SAS CODE WITHOUT SAS



NO MACROS



ENGLISH TEXT AS INPUT



FLEXIBLE INTERPRETATION

STANDARD MACROS VS AI-BASED TOOLS

SAS macros



Great for standard mapping



Bullet-proof when combined with CT



The more customization,
the more manual work

AI-based tools



Incredible flexibility



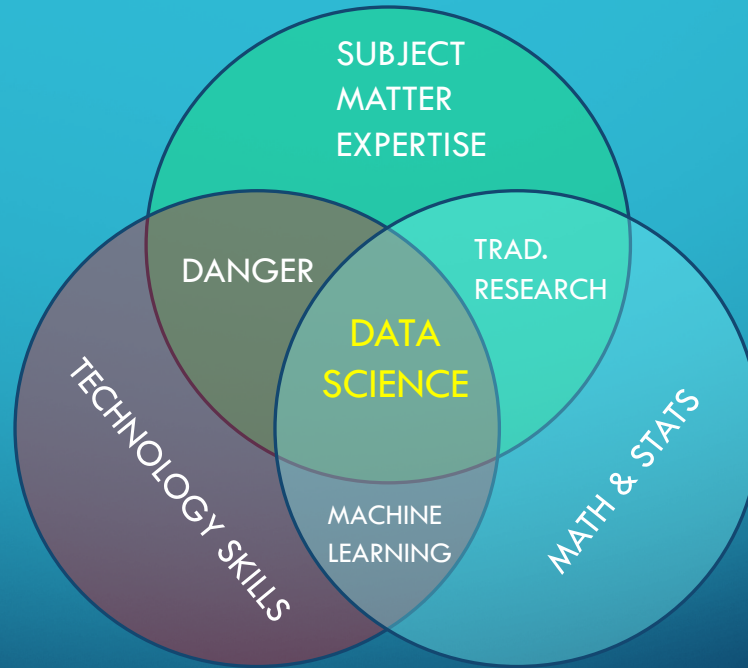
Great with non-standard



Can and will make mistakes,
depending on training of
the algorithm! Resulting SAS code
and dataset will require a review!

- How can you become a Data Scientist?

YOU ALREADY HAVE MOST OF THE SKILLS



SAS CLINICAL PROGRAMMERS VS DATA SCIENTISTS

SAS
Programmer

Data Scientist



Data manipulation techniques

Statistics

Handling of missing data

Machine learning

Work on Big Data

Data visualization

Many different data types and formats



TECHNOLOGIES



ANACONDA

SKLearn

Theano

TensorFlow

KERAS

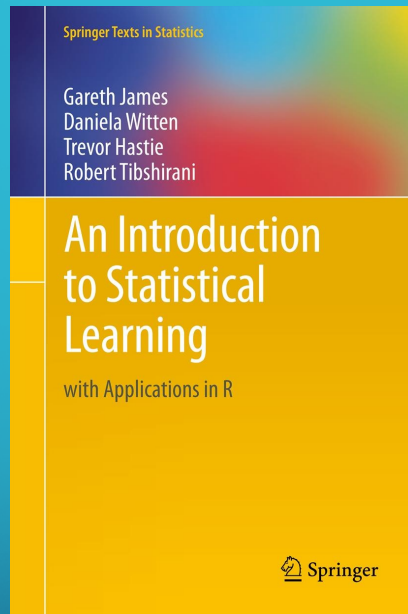


ML packages on
CRAN



Proc neural
(separate
license might be
needed) SAS
Enterprise Miner
for exploratory
analyses, SAS
VIYA

TO PICK UP BASICS



<http://www-bcf.usc.edu/~gareth/ISL/>

<http://www.deeplearningbook.org/>