Experiences from Social Data Science 2018-20
Knowing how to use the microwave,
...not how to build it

Roman Jurowetzki; roman@business.aau.dk; 10/3-2021

BUSINESS SCHOOL
AALBORG UNIVERSITY

About me

- BA Governance & Public Policy (Uni Passau)
- MSc Innovation Economics (AAU)
- PhD – Innovation Economics – Mapping the Development of the Danish Smart Grid

- Never been really good at mathematics
- Needed to solve a problem during my PhD
- Started with Python, NLP, Network analysis …

- Extended into ML/AI methods

- AI Denmark – aidenmark.dk
- Lead: AI Growth Lab

From catching up to industrial leadership: towards an integrated market-technology perspective. An application of semantic patent-to-patent similarity in the wind and EV sector

Daniel S Hain, Roman Jurowetzki ✉, Primoz Konda, Lars Oehler

The Privatization of AI Research(-ers): Causes and Potential Consequences

– From university-industry interaction to public research brain-drain? –

Roman Jurowetzki[φ], Daniel S. Hain[φ], Juan Mateos-Garcia[†], and Konstantinos Stathoulopoulos[†]

[φ]*Aalborg University Business School, DK*
[†]*Nesta, UK*

February 15, 2021

BUSINESS SCHOOL
AALBORG UNIVERSITY

Need for non-coders

Industry demand for SAMF students with strong quantitative / computational skill. Yet, current DS/ML teaching mostly for CS students.

Data & ML literacy becoming crucial across managerial positions → Transfer of ML to **business & policy opportunities.**

Need for data literacy

Need for stronger analytics

Need for stronger applied method training to support exciting master projects, transition into PhD, and advanced BI positions.

Students want to get challenged

Put into a situation where they have to solve problems that are **hard**, real, **measurable** but **fair** and **solvable**.
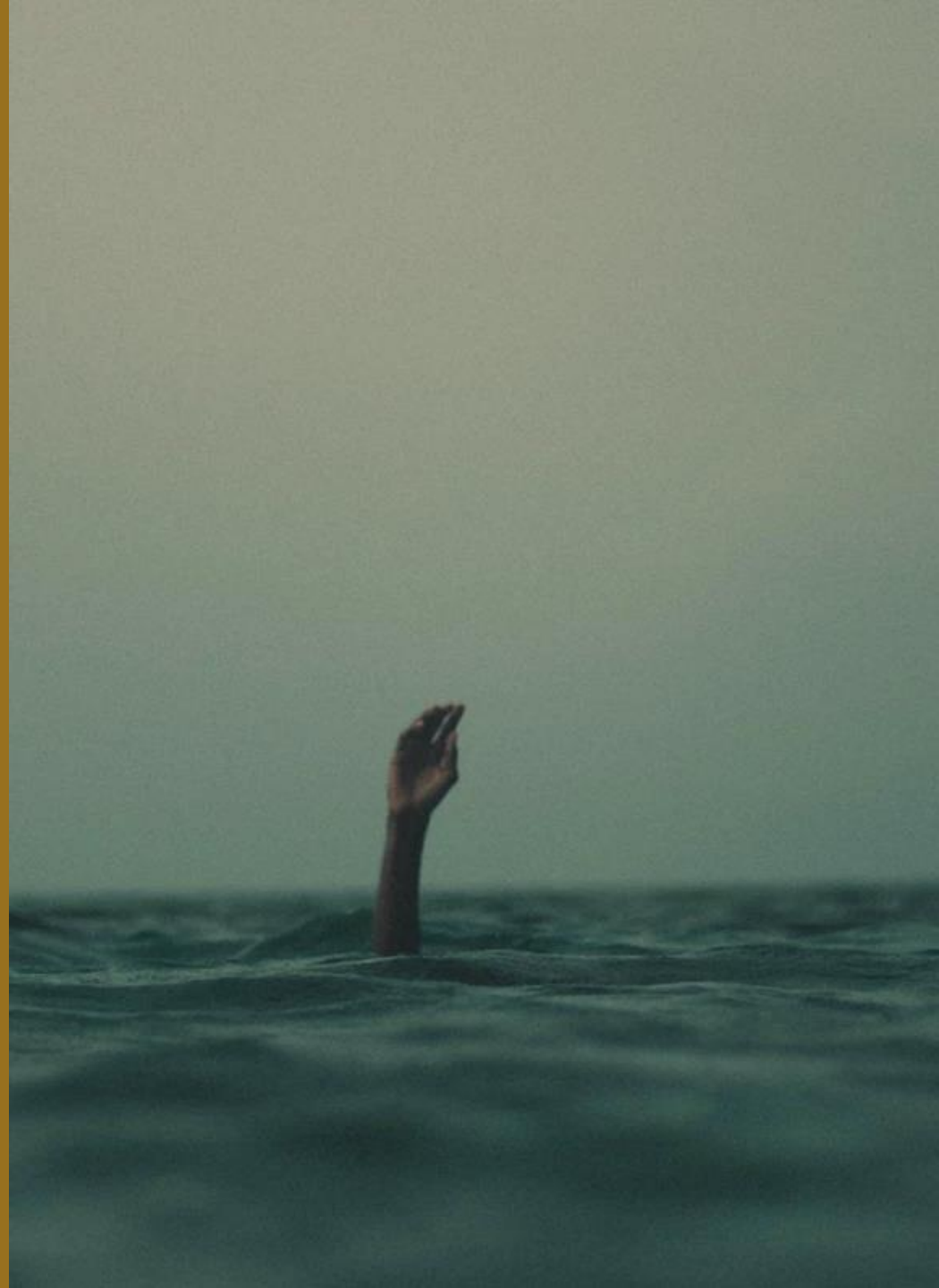
real PBL (my view…)

Too little PBL in SAMF/HUM

- Lack of tangible problems
- Lack of measurable goals
- Broad scope and (too) much freedom
- Very few quantitative projects
- Little problem-solving

BUSINESS SCHOOL
AALBORG UNIVERSITY

Disconnect of methods training and application

- Mathematics mainly thought in a theoretical vacuum

- Statistics and econometrics mostly thought with strong formal emphasis and

- main focus on inferential statistics (causality and variations of OLS) while predictive analytics runs the world

BUSINESS SCHOOL
AALBORG UNIVERSITY
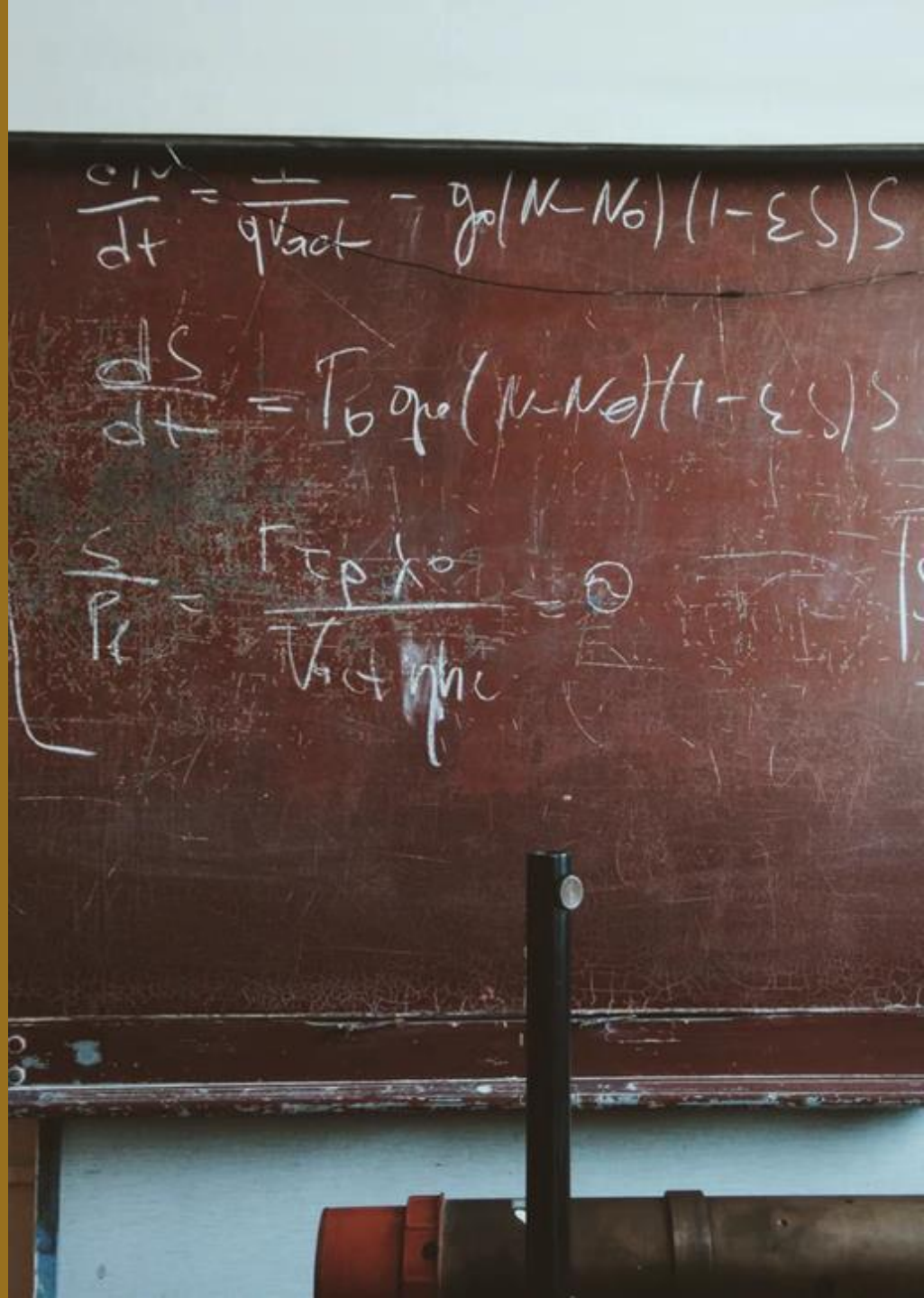
Lack of motivation and legitimacy
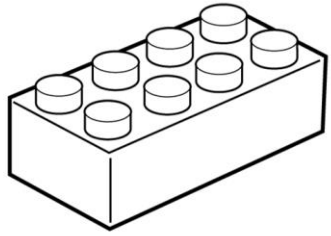
Bottom Up ☝   method first
vs.
Top Down 👇   **problem first**

- Are my students going to compete with mathematicians or ML engineers?

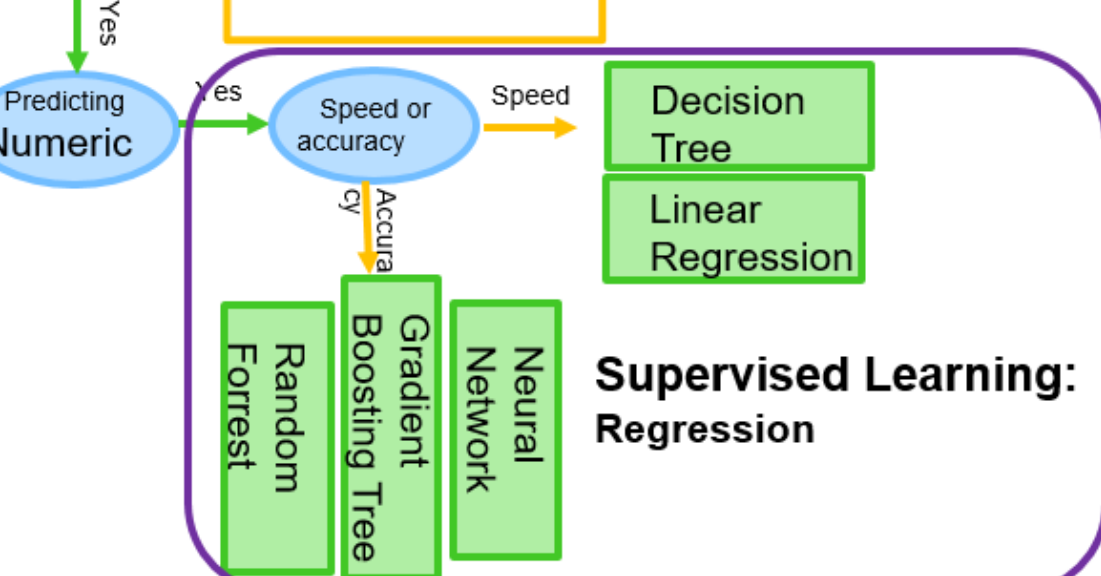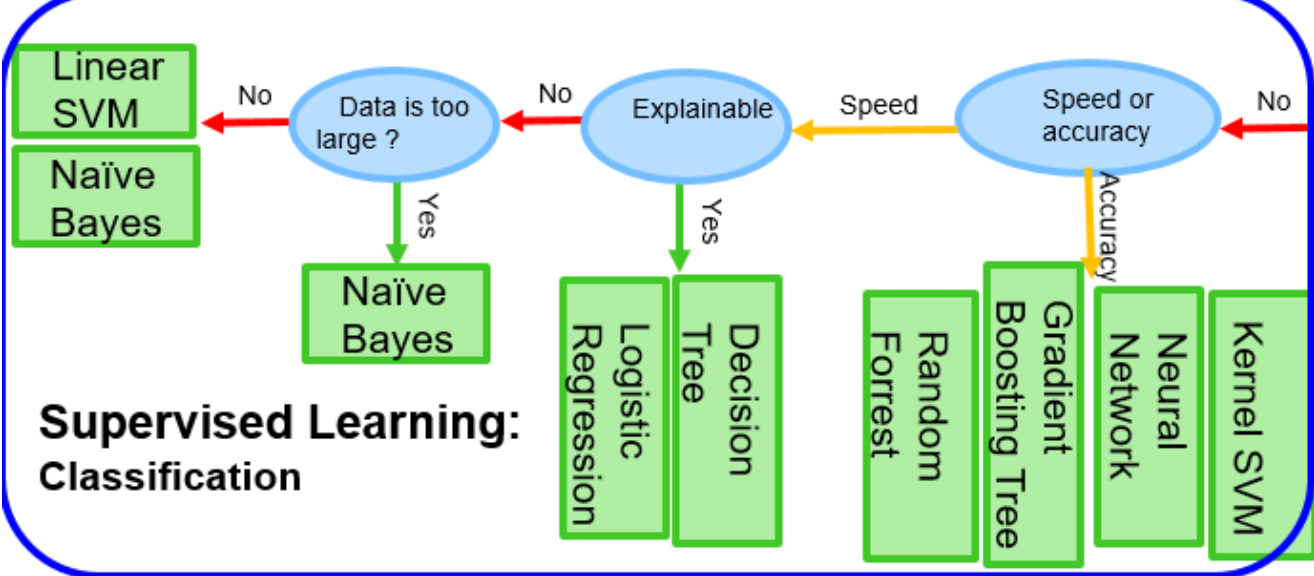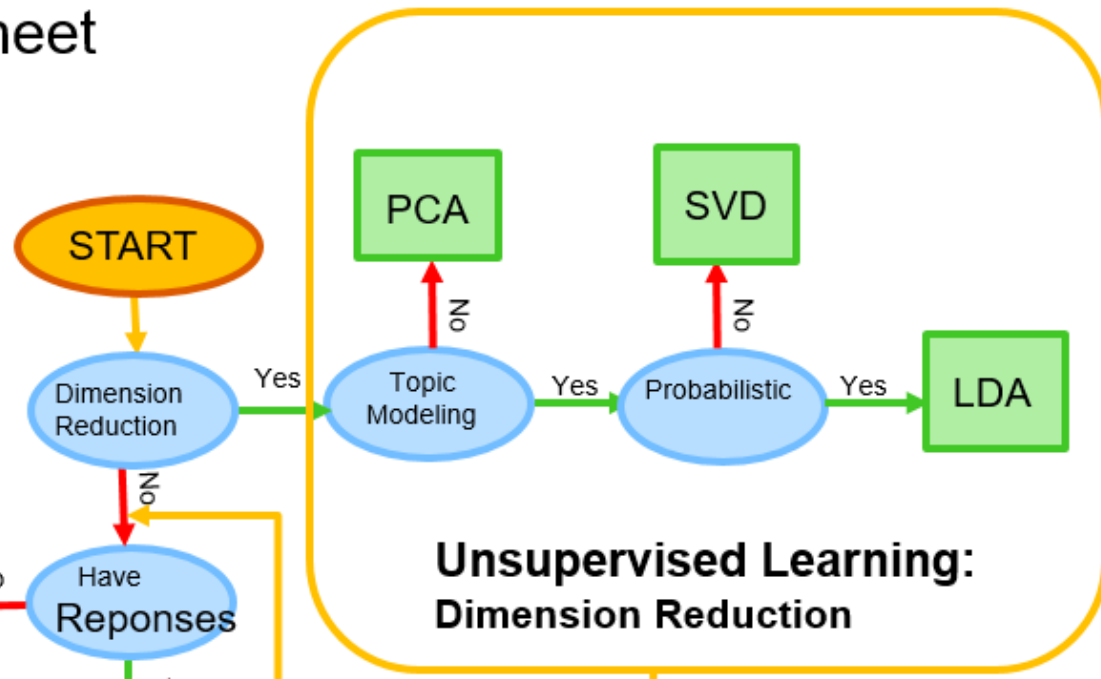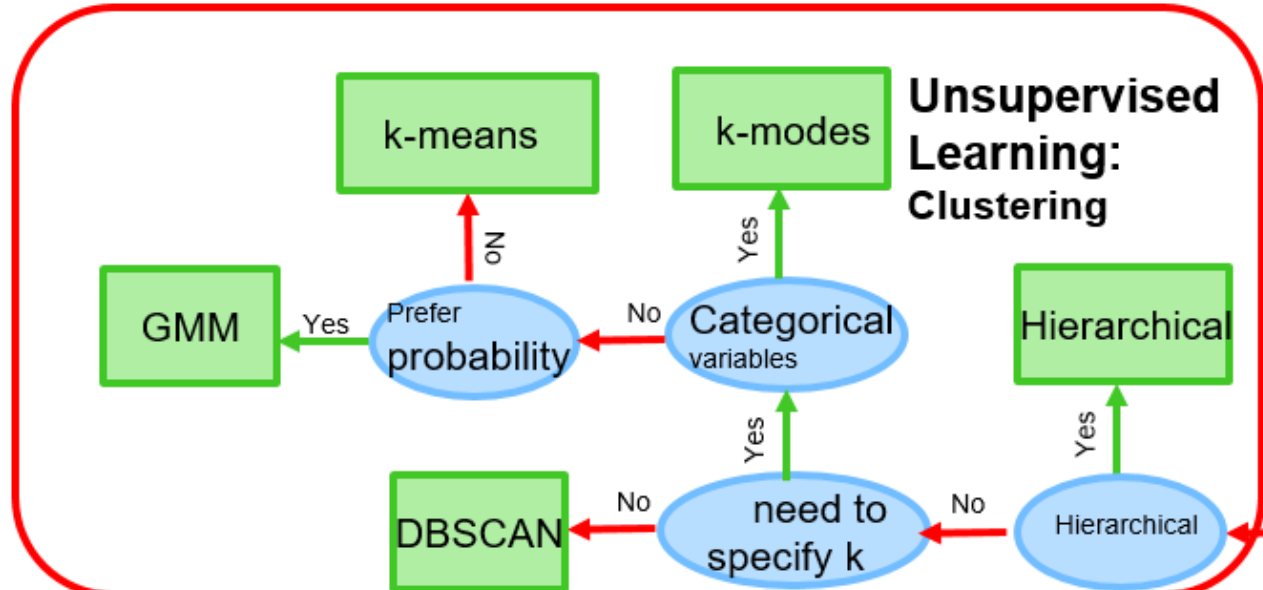- How many of my students are going to do a PhD?

"You do not need to have a PhD in electrical engineering to be allowed to use a microwave! "

*- Cassie Kozyrkov - Chief Decision Scientist - Google*

Machine Learning Algorithms Cheat-sheet

SEMESTER SCHEDULE

NLP and Networks (5ECTS)
Working with unstructured and relational data. Modelling semantic and relationships.

Capstone Project (15 ECTS)
Usually, work on a real-world data science problem in collaboration with companies and organizations.

M2 Oct.

M4 Nov-Jan

M1 Sept.

M3 Nov.
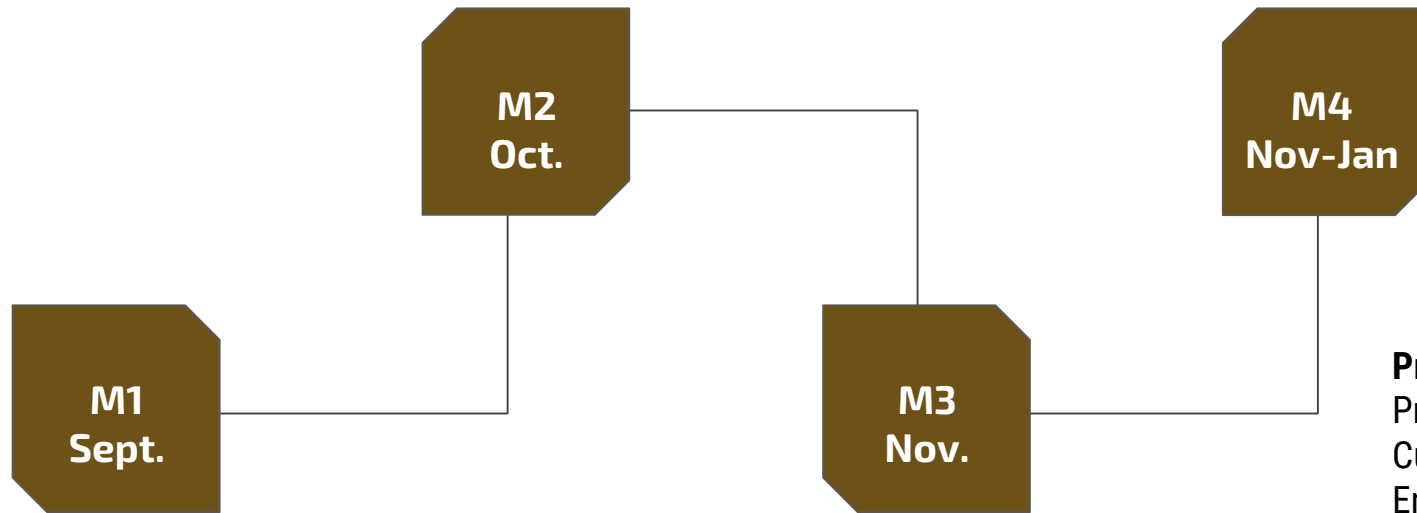
Intro to applied Machine Learning (5ECTS)
Working with data, "traditional" machine learning, model evaluation

Intro to Deep Learning and AI (5ECTS)
Designing artificial neural network architectures to get state-of-the-art results. Also: When not to use them.

Predicting
Production errors
Customer churn
Employee turnover
Energy prices
Street maintenance

spar nord

KKbox

centrica energy

SIEMENS Gamesa
RENEWABLE ENERGY

Aalborg Kommune

telenor

We teach R and Python

We host our material on Github and encourage students to upload their projects there

David Keyes
@dgkeyes

Teaching students to use software they won't have access to after graduation is immoral.

3:00 PM · May 1, 2020 · Buffer

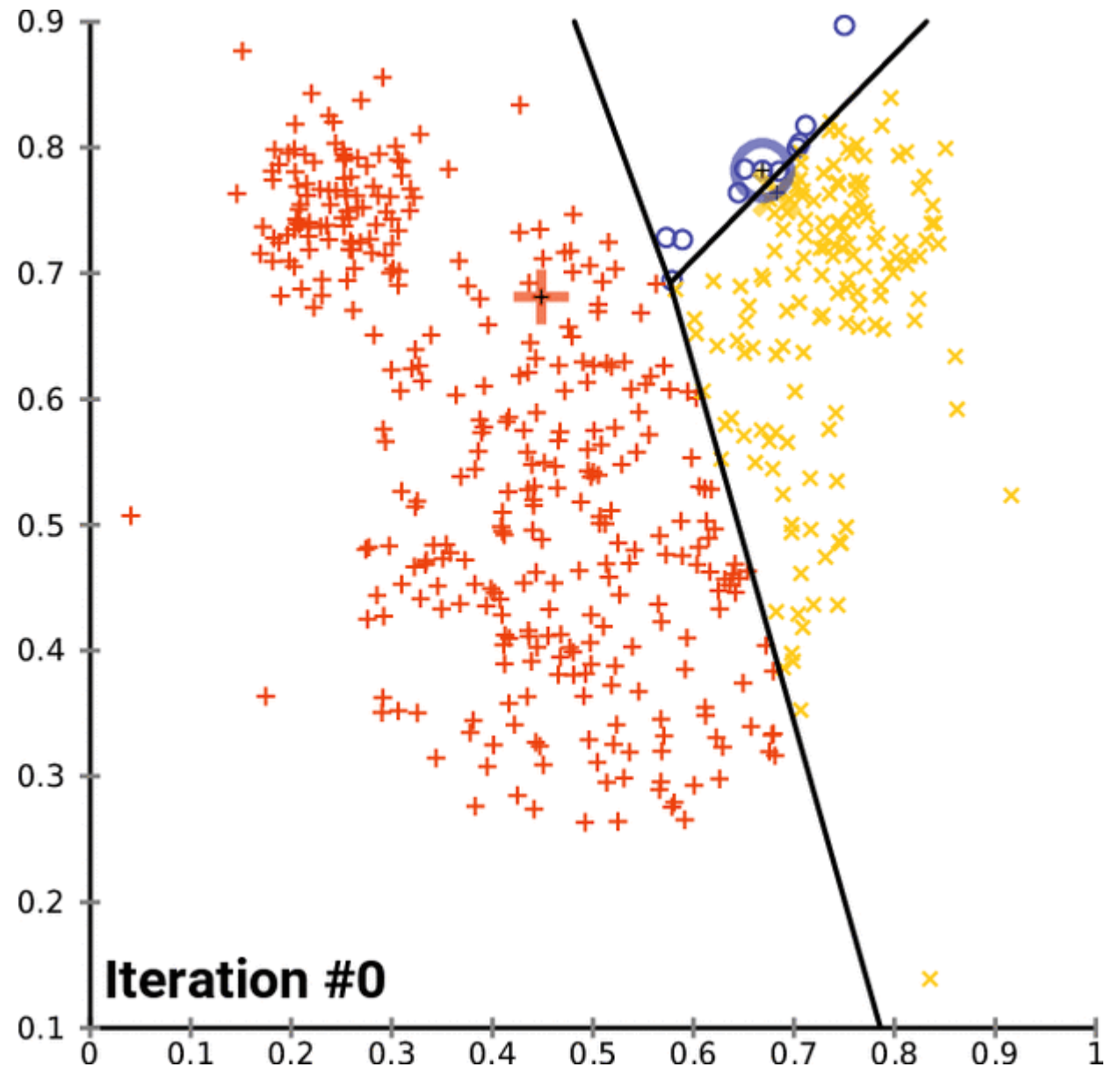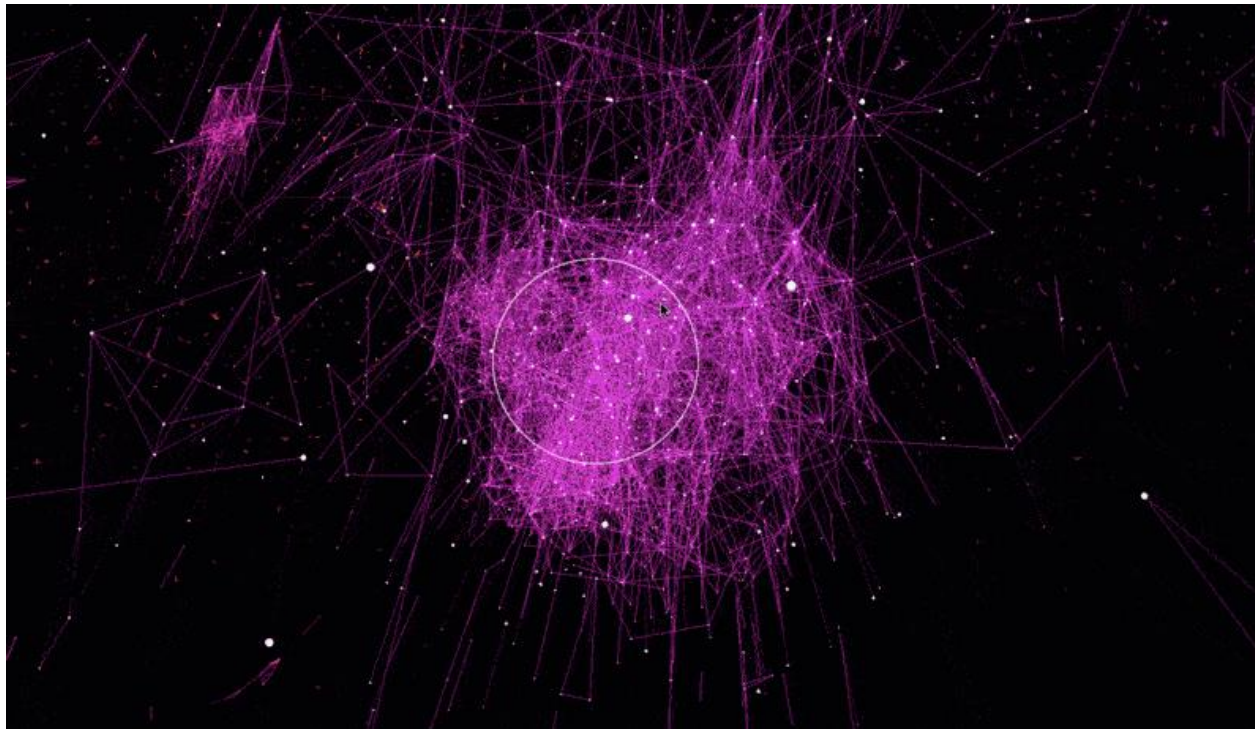236 Retweets   70 Quote Tweets   1,604 Likes

GitHub

Module 1

- Intro to machine learning
- Intro to "communities of practice"
- Data handling
- Descriptive statistics / Exploratory Data Analysis
- Data Visualization
- "Traditional" ML - supervised / unsupervised
- Evaluation
- Ethics, Algorithmic bias & other issues

Networks & NLP

- Understanding / handling relational data
- Network statistics / indicators / visualizations
- Complex networks (multiplex / edge color)

- Intro to natural language processing (working with text data)
- Bag-of-word models and derivatives
- Introduction to embedding
- Vectorization and use in ML pipelines

Deep Learning

- Intro to neural models and Keras-Tensorflow
- Common DL architectures
  - Feed forward nets
  - Convolutional nets
  - Recurrent nets
- Advanced architectures
  - Autoencoders
  - Multibranch models
  - Advanced work with embeddings
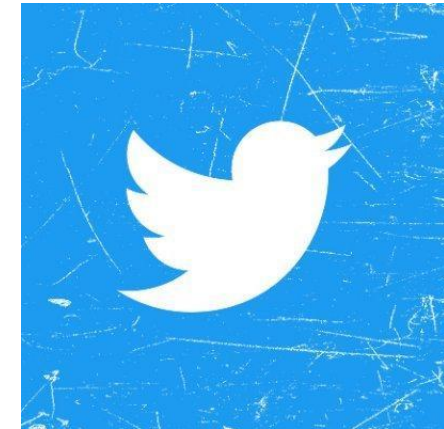 - Transformers
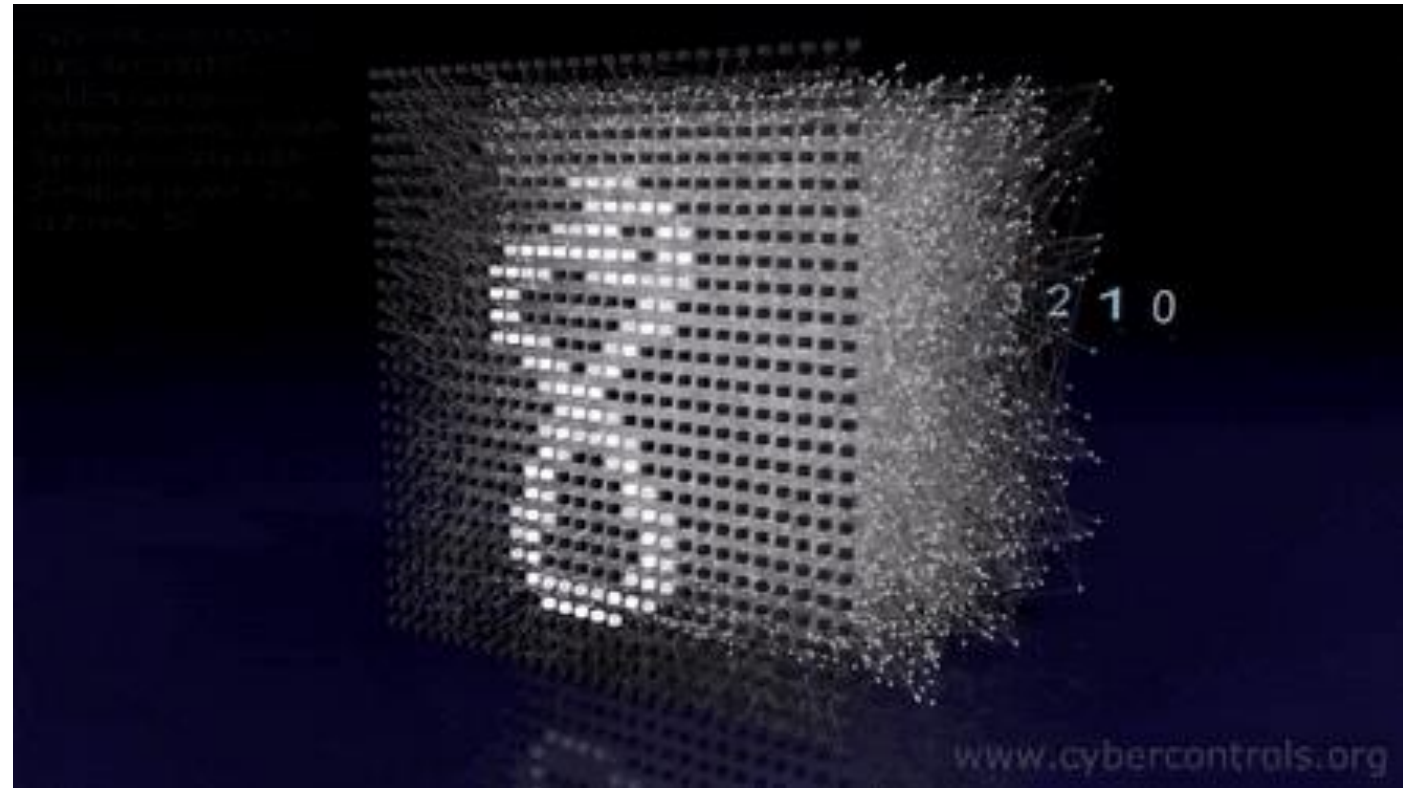
www.cybercontrols.org

Stock Movement Prediction and Trading Strategy based on Tweets: An Analysis on the Streaming Sector

Designing an algo-trading bot that uses
- Stock data
- Twitter sentiments ☺ ☹

To propose a trading strategy

Stock Movement Prediction and Trading Strategy based on Tweets:
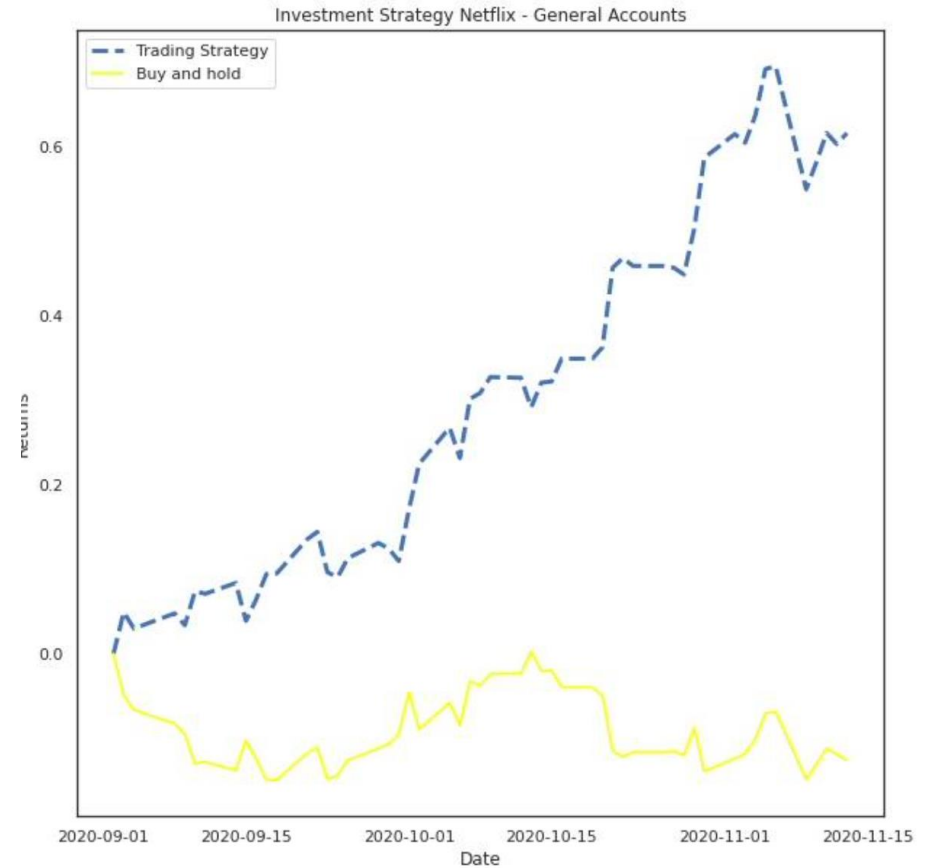# An Analysis on the Streaming Sector

Figure 15: Investment Strategy Netflix

- Source, clean and assemble data (financial & twitter)
- Identify networks of relevant twitter profiles (network analysis)
- Preprocess text data and run sentiment analysis
- Aggregate results and align with stock market data
- Train machine learning model (training, fine-tuning etc.)
- Define, assemble and run back-testing
- Document and communicate process and results

So what?

- Motivation first!
- Create more opportunities to learn from solving *real* problems
- Courage to confront SAMF/HUM students with real "engineering problems"
- Acknowledge that Github, Medium, tech. documentation and ArXiv are often better/timlier than journal articles or books
- Courage to leave some details to STEM

BUSINESS SCHOOL
AALBORG UNIVERSITY