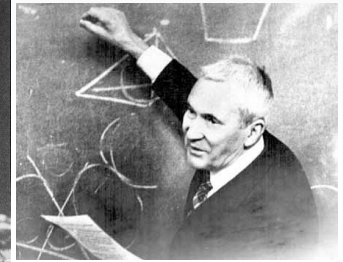
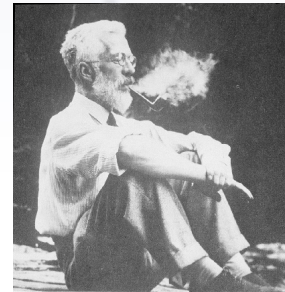
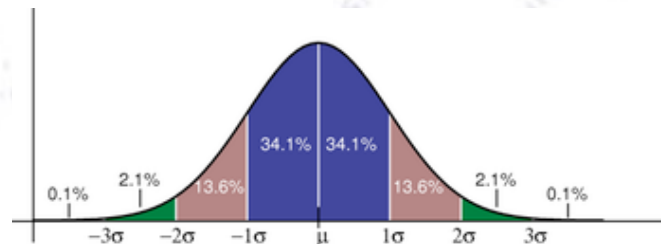


Applied ML

Potential Final Projects/Datasets



Troels C. Petersen (NBI)

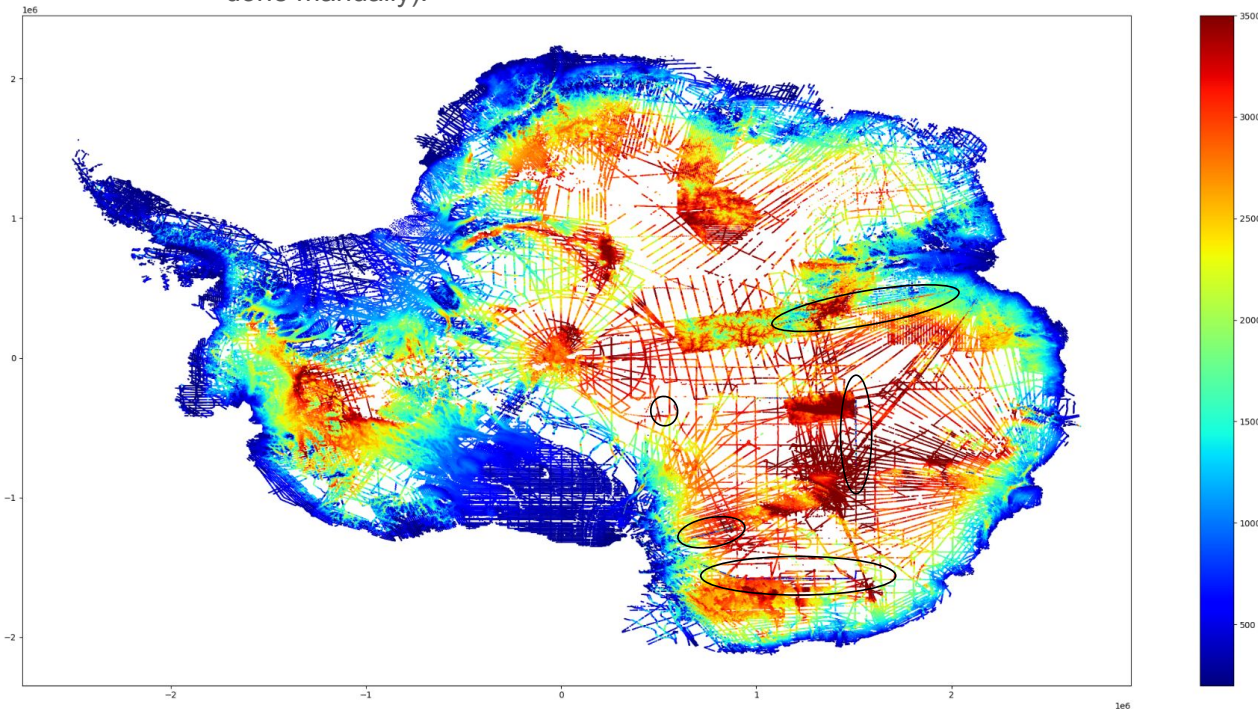


"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

Antarctic Ice Sheet Anomaly Detection

Out of a large number of measurements, some are clearly wrong, but which? And how to assess if this is the case? Several methods exist. Try them out.

Examples of outliers visible by eye. Data is ice thickness measurements from BedMap (<https://bedmap.scar.org/>). 80m points. The goal is to automate outlier detection (currently done manually).

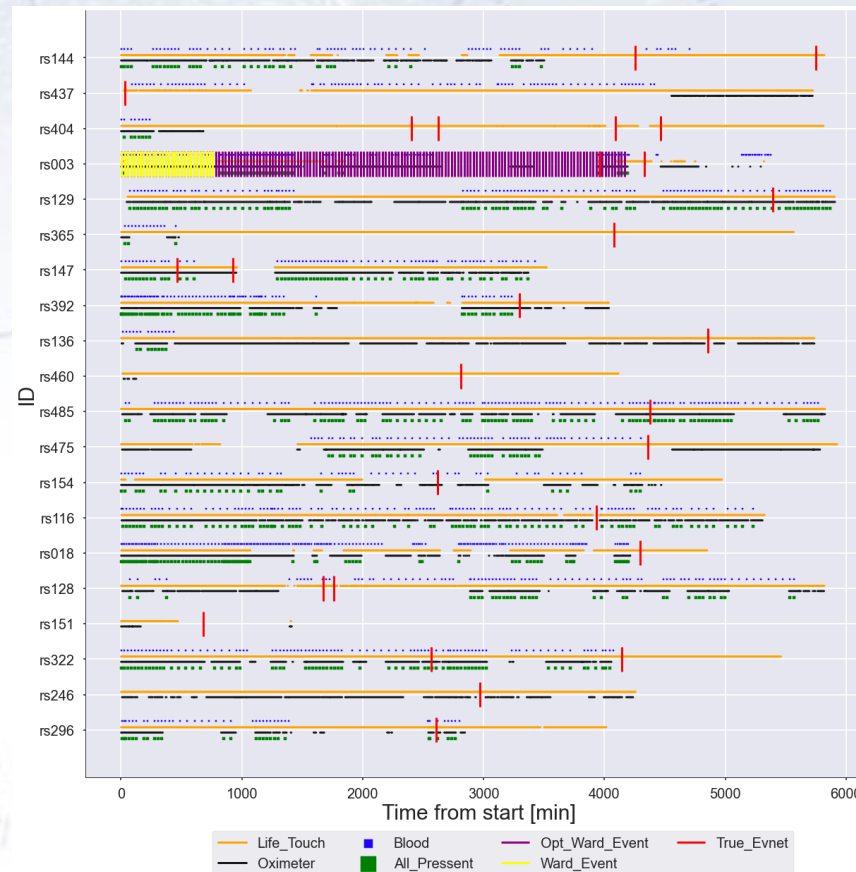


Data status: Ready here! Niccolo will be accessible on Zoom.

Data size: About 80.000.000 / 10.000 (?) measurements / labels, accessible format.

WARD data - Event Prediction

Based on 7 continuous patient outputs, determine if the patient is doing OK or...

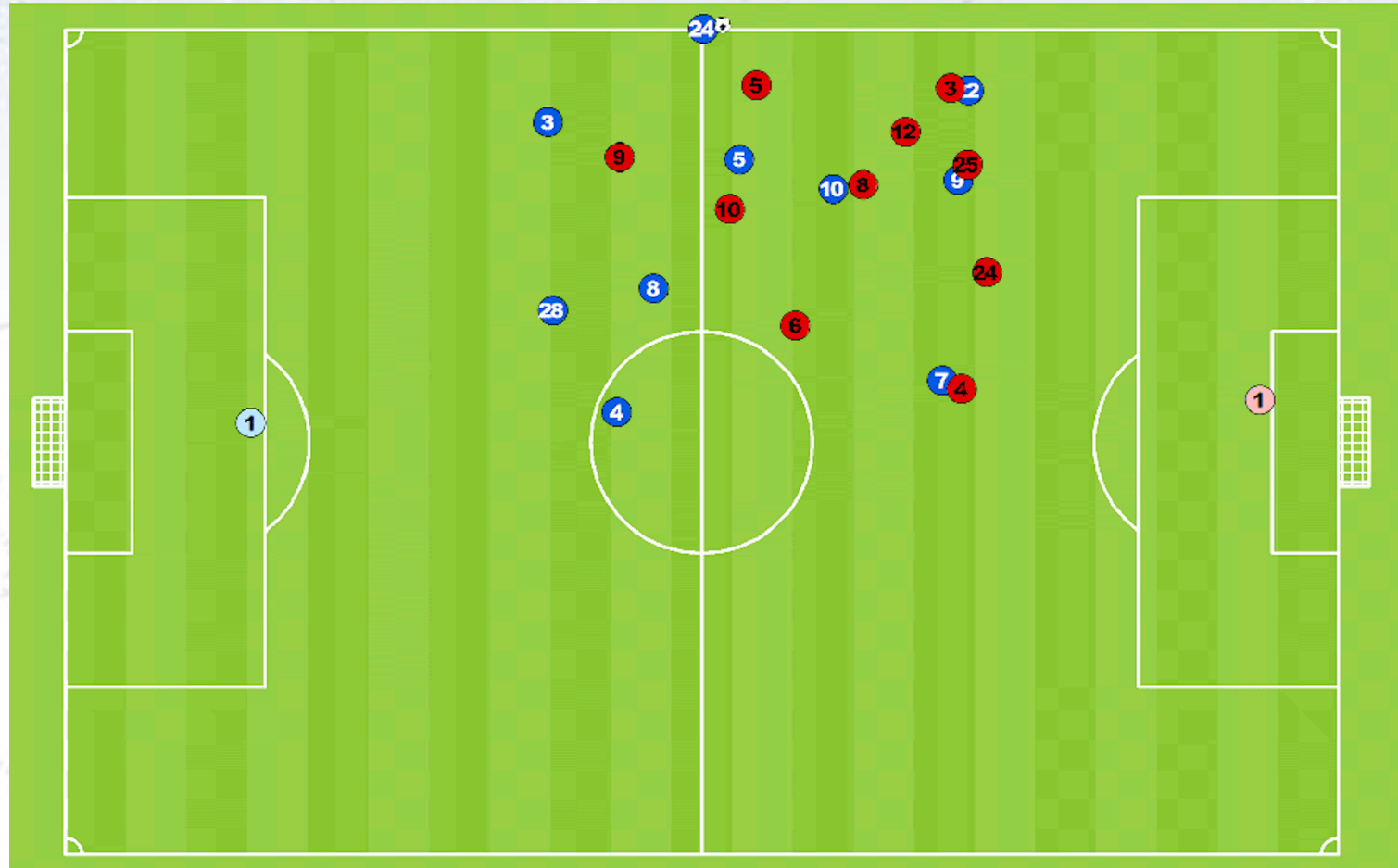


Data status: Sits on discs at NBI, i.e. ready (with Ph.D. Norman Pedersen).

Data size: 2600 patients with millions of measurements in accessible format.

Football data

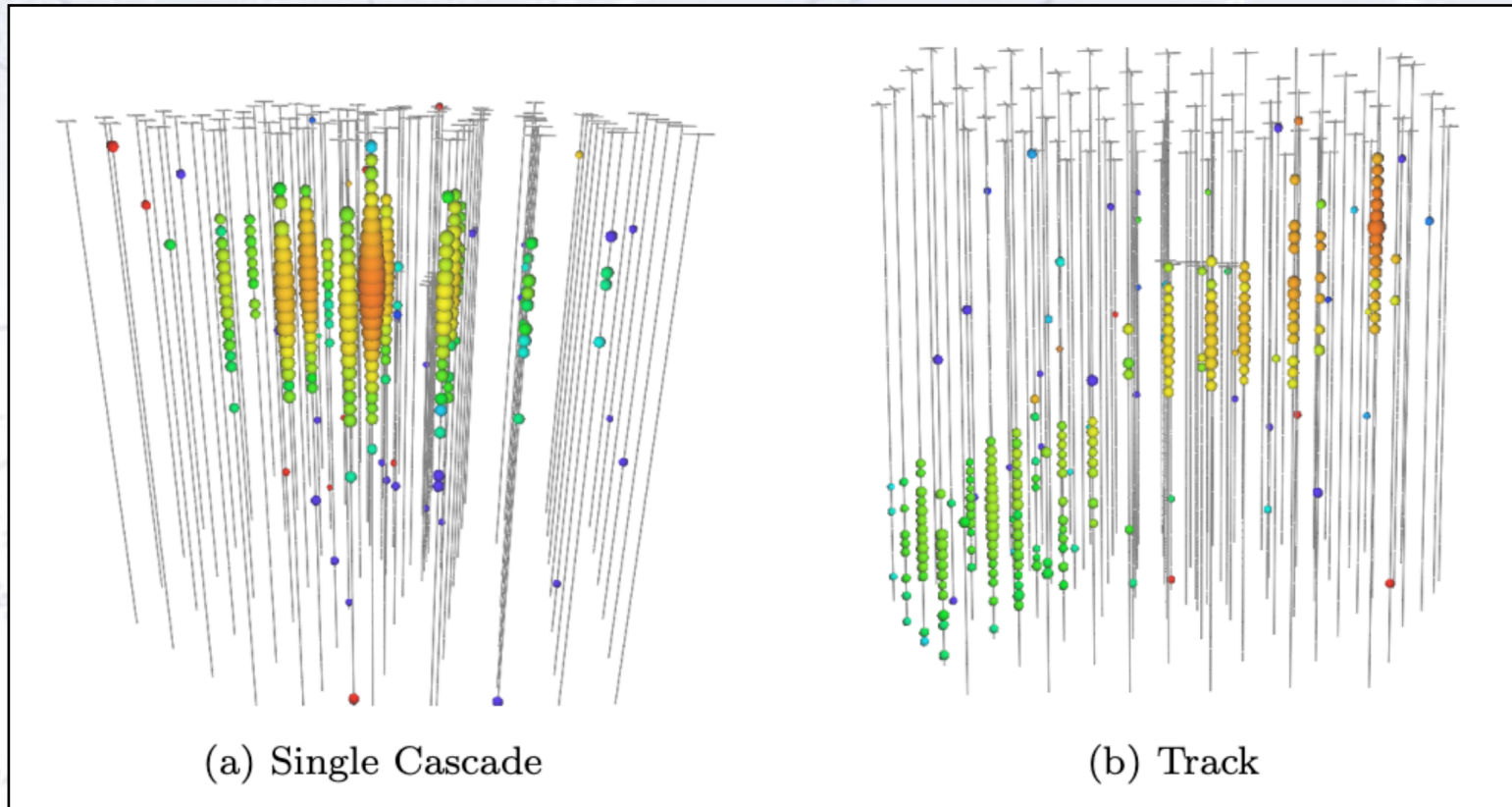
Positions of players and ball in xyz and timesteps along with annotations of what happens in the match. From (a few years in) Danish Super Liga.



The goals are to predict goal chances, best plays, player status, etc.

Particle Physics data

Based on detector signals (8-80000 pulses with 6 features each), determine the type of signal, its energy, direction, interaction vertex, etc.



Data status: Sits on discs at NBI, i.e. ready!

Data size: Real+Simulated millions of events in (semi-)accessible format.

Predicting turbulence for landing

Case: Turbulence Level in Airport Operations by FORCE Technology

Purpose: Predict aircraft turbulence level based on airports wind data

Description: Turbulence levels are important for comfort and safety during aircraft arrival and departure at airports. Some airports are located in landscapes where certain wind patterns cause turbulence. It is important to be able to predict the turbulence levels to be able to warn the aircraft about the turbulence levels. Aircraft turbulence is quantified/ categorized by the amount of vertical acceleration.

In this case the airport has two wind sensors (anemometers) continuously recording wind direction and wind speed at each end of the landing strip/runway. The task is to create an algorithm for predicting the turbulence level experienced by the aircraft during landing and take off within a height of 1500 foot range above the runway based on the readings from the anemometers.

Data:

Anemometer data are available in CSV formatted text files covering 24 hours per day. In total 521 files. Each file contains data for both anemometers logged at 5s intervals. Times are in UTC.

Flight recorder data are available as time series for 3140 arrivals and 3039 departures. Each timeseries is available in a single file containing position, horizontal and vertical speed, accelerations etc. File formats vary slightly with 2-4 different variations depending on the specific aircraft. Experienced turbulence level should be derived and characterized from the time series. The file names identify landing (ARR) or take-off (DEP) and heading defines runway (04 or 22). Times in UTC.

Overall requirements to project

There are hardly any formal requirements to the project.

The only really strict requirements are, that you use Machine Learning algorithms on some sizable data, and that you work in a group.

Groups should ideally consist of 3-5 persons. Exceptionally (and given good reasons), we also allow for smaller groups (1-2 persons), which will have the same deadline, but who will present their projects towards the end of the exam.

We highly encourage groups, as this is the ideal way of learning, and how most (all?) project work functions in both academia and industry.

A faded, circular magnetic chart, likely a compass rose or a magnetic field diagram. It features concentric contour lines with numerical labels such as 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 450, 480, 510, 540, 570, 600, 630, 660, 690, 720, 750, 780, 810, 840, 870, 900, 930, 960, 990, and 1020. The word "MAGNETIC" is visible at the top. A small crosshair is located near the center, with the text "VAR 10° 15' W" written next to it. The chart is overlaid with a grid of latitude and longitude lines. The text "THE BITTER END TACHT/CLUB" is visible in the upper right quadrant.

Bonus Slides



Let me tell you about this exam!