# Intro

* Dagens program / formål
* I holder selv pauser efter behov
* Vi holder opsamling kl. 11 (ca.)

# Opgave 0 Rester fra sidste gang

***Arbejd sammen 2 og 2***

Besvar følgende spørgsmål:

* På side 55 splittes dataset’et i to: *train* og *test*. Hvad er formålet med det?
* Hvorfor bliver dataset’et først blandet inden det splittes?
* Hvorfor er det nu at det er en god ide at udføre en standardize via metoden *transform* på et dataset?
* Hvad er *overfitting*? Hvordan tror du det kan opstå?
* Hvorfor anvendes regularization?

# Opgave 1 EDA

***Arbejd sammen 2 og 2***

**I skal i dag udføre en (delvis) såkaldt EDA – exploratory data analysis!**

se et eksempel her: <https://www.analyticsvidhya.com/blog/2021/04/rapid-fire-eda-process-using-python-for-ml-implementation/> )

**Trinene er:**

Gå ind på colab.research.google.com eller lav en notebook i din egen editor fx vs code.

Hent csv-filen, som du har fået udleveret, ind i en pandas dataframe.

* Analysér data med div. Pandas-tricks
* Visualisér eventuelt data gennem forskellige plots, fx barchart for at få et bedre overblik over data.
* Find manglende værdier i datasættet
* Håndtér de manglende værdier
* Håndtér datokolonnen på en fornuftig måde
* Lav en ekstra kolonne, der viser, hvornår færgen har ligget i stille (i havn) baseret på trackspeed, og fjern derefter kolonnen trackspeed. Den nye kolonne er din target-variabel.
* Opdel datasættet i train og test
* Brug standardization på datasættet
* Træn en Logistic regression model på dine data, og udskriv dens accuracy. for at finde ud af, hvornår færgen ligger stille
  + afprøv forskellige resultater værdier for C eller andre hyperparameters, og notér hvad der fungerer bedst.
* Prøv at gøre modellen enklere ved at ændre på hvilke kolonner modellen har til rådighed, og se, om hvordan det påvirker resultatet. Hvad er den bedste score du kan opnå?

# OPSAMLING

Vi kigger på, hvad folk har nået, og hvilke resultater de er kommet frem til.

# Opgave 1b Refleksionsspørgsmål

* Lider din model af overfitting?
* Hvordan bekæmper man det?

# Opgave 2 Overfitting

Håndtér overfitting!

Opsamling / Outro (5 minutter)

* Var der noget vi ikke nåede eller ikke fik behandlet tilstrækkeligt?
* Var det svært / nemt?
* Lærte du tilstrækkeligt i dag?   
  Hvis nej: Hvad var årsagen?
* Andet og næste gang…