**Opgave uge 11: Donér blod**

Opgaven (er taget fra en tidligere konkurrence, som nu er lukket):

Dataset:

Findes på itslearning

[~~https://archive.ics.uci.edu/ml/machine-learning-databases/blood-transfusion/~~](https://archive.ics.uci.edu/ml/machine-learning-databases/blood-transfusion/)

Din opgave er, at forudsige, om en bruger vil donere i marts måned 2007 alene baseret på hans historik (som ligger før marts 2007).

Følgende kolonner er indeholdt i datasættet:

* Months since Last Donation: this is the number of monthis since this donor's most recent donation.
* Number of Donations: this is the total number of donations that the donor has made.
* Total Volume Donated: this is the total amound of blood that the donor has donated in cubuc centimeters.
* Months since First Donation: this is the number of months since the donor's first donation.

Brug følgende metoder til at løse opgaven:

* Forward Stepwise variable selection
* Lav AUC train and test curves (hvor skal dit cutoff ligge?)

Se om du kan gøre det bedre end leaderboardet

Leaderboard:

A screenshot of a cell phone

Description automatically generated

Da konkurrencen er slut skal der ikke afleveres noget, men prøv alligevel at lave en fil i følgende format som dit output

Submission file:

Id,Made Donation in March 2007

659,0.5

276,0.5

263,0.5

303,0.5

...

Opgaven anvender log loss som evalueringsmetric, så det du skal skrive i filen er altså sandsynligheden for at en donor vil donere blod i marts 2007 (per donor).

Beskriv med egne ord:

Lift-curve

Cumulative gain curve

AUC-curves

Log loss (<http://wiki.fast.ai/index.php/Log_Loss>)

Husk at log loss især straffer predictions, som er sikre men forkerte!