## Petfinder / Pawpularity

I får her et datasæt, som I skal anvende på forskellig vis gennem semesteret. Data stammer fra en konkurrence om at bygge et system som v.hj.a. ML kan give en “cuteness” score til et billede af et kæledyr. Se Bilag nedenfor (som indeholder link til flere kilder) og bemærk at der er mange valg som er overladt til udviklerne.

Brug jeres fantasi, eksperimenter og hjælp hinanden - også på tværs af grupper (og på tværs af geografi...). Og husk at det ikke altid er lige oplagt at anvende ugens teori/emne på dette datasæt.

## Projektarbejde

Hver uge kommer vi til at arbejde med ting, der er relevante for at løse opgaven, og derfor er projektet ugeopdelt, men I må tage det i jeres helt eget tempo, da vi ikke skal se det endelige resultat før eksamen, men det kan være en fordel at arbejde på projektet hver uge ang. ugens tema. Jeres eksamensprojekt skal ende ud i en præsentation af det I har arbejdet med gennem hele semestret.

Til selve eksamen skal I lave en præsentation af jeres forsøg og opdagelser, samt en demo af jeres applikation.

## Applikationen

Alle nedenstående krav skal kunne testes og vises i applikationen. Det kan være en webapplikation eller en desktopapplikation. Der er ingen forventninger til brugergrænsefladen eller funktionaliteten. Hav fokus på at applikationen skal anvendes til at kunne arbejde med data og hver uges tema samt at I skal kunne præsentere jeres arbejde til eksamen.

Det hele skal afleveres senest 30. maj i Wiseflow.

## Fejl = Succes (måske)

Det er **vigtigt** at understrege at I **ikke** behøver at komme i mål med nedenstående krav ift. ugernes emner. Løsningen af kravene er i bund og grund uvæsentlige, men jeres læring ang. ugernes emner er derimod ***supervæsentlige!***

Der kan være hundredvis af gode grunde til at I ikke kommer i mål med opgaven, men derfor kan I jo godt lære en masse – ofte lærer man mere af sine fejl end af sine succeser (og her taler jeg af erfaring... 🙄😬😲).

Hvis I kan påpege hvad der er årsagen til at det ikke lykkedes eller kan komme med en beskrivelse af hvad der skal til for at løse opgaven, er det til gengæld meget fint og kan sagtens høste høje karakterer!

Husk hele tiden at formålet med applikationen er at I kan demonstrere jeres erfaringer og viden indenfor de enkelte emner – og det behøver ikke at betyde at I lykkedes med at løse opgaven.

## Uge 10:

* Udforsk datasættet på forskellige måder.
* Udarbejd en (Python) applikation som kan indlæse og vise datasættet

## Uge 11:

**Kap 7:**

* Brug regression til at finde pawpularity-scoren for alle kæledyrene og test med forskellige inputs, om modellen regner rigtigt.
* Skriv en applikation, hvor man kan afprøve din model, så applikationen giver mulighed for at man kan indlæse data (for alle de eksisterende features) vedr. et billede ~~et nyt billede~~ og få vist pawpularity scoren for billedet.
* Brug supervised learning algoritmerne fra bogen til at finde ud af, om der er et menneske på billedet du viser. Hvis der er et menneske på, skal billedet fjernes fra visningen.

**Kap 8:**

* Anvend diverse metrics og vis et overblik over dem i din applikation, så man kan se, hvor gode dine modeller er, brug fx:
  + Performance measure
  + Cross validation
  + ROC curve
* Test med forskellige Hyper parameters
* Vis resultaterne i din applikation

## Uge 12:

**Kap 9:**

* ~~Beregn sandsynligheden for at der er et ”occlusion” på billedet med Bayes~~
* Lad os finde ud af om vi kan sige noget om scoren af Pawpularity givet kendte værdier af Eyes, Face og Occlusion v.hj.a. Bayes – dvs. modellen skal være en GaussianNB.
* Da der er 100 mulige værdier af Pawpularity bliver det ret svært at rame præcist og vi kan derfor ændre værdier til 0 (nul) hvis den oprindelige værdi er mellem 0 og 75 og 1 (en) hvis den oprindelige værdi er mellem 76 og 100.   
  Nedenfor vises hvordan det kan gøres:

df = pd.read\_csv('train.csv')

df['Pawpularity'] =   
 df['Pawpularity'].apply(lambda x: 0 if x >= 75 else 1)

X = df[['Eyes', 'Face', 'Occlusion' ]] # Features

y = df['Pawpularity'] # Målvariabel

X\_train, X\_test, y\_train, y\_test = train\_test\_split(... osv.

* Byg og træn modellen. Beregn og udskriv modellens accuracy.

## Uge 13:

**Kap 10:**

* Brug Bagging, Boosting og Stacking
* Kombinér Bayes resultat med resultatet fra din supervised algoritme til at afgøre, om der er mennesker og/eller occlusion på billedet, i så fald fjern dem fra visningen.
* Vis resultaterne i din applikation

## Uge 14:

**Kap 11:**

* Find mønstre og sammenhænge i data med unsupervised learning
* Vis resultaterne i din applikation

## Uge 15+18 (bemærk at ugerne 16-17 er påskeferie!):

**Kap 12+13:**

* Brug NN til at løse den opgave du tidligere løste med supervised learning
* Giv brugeren mulighed for at vælge mellem neuralt netværk og andre supervised learning algoritmer til opgaven.

## Uge 19:

**Kap 14:**

* Brug CNN til at afgøre om billedet indeholder en kat eller en hund
* Brug CNN til at beskrive alle de features der er i billedet
* Vis resultatet i applikationen

# **BILAG**



Et billede, der indeholder Små til mellemstore katte, Kattefamilien, pattedyr, Knurhår

Indhold genereret af kunstig intelligens kan være forkert.

A picture is worth a thousand words. But did you know a picture can save a thousand lives? Millions of stray animals suffer on the streets or are euthanized in shelters every day around the world. You might expect pets with attractive photos to generate more interest and be adopted faster. But what makes a good picture? With the help of data science, you may be able to accurately determine a pet photo’s appeal and even suggest improvements to give these rescue animals a higher chance of loving homes.

[PetFinder.my](https://petfinder.my/) is Malaysia’s leading animal welfare platform, featuring over 180,000 animals with 54,000 happily adopted. PetFinder collaborates closely with animal lovers, media, corporations, and global organizations to improve animal welfare.

Currently, PetFinder.my uses a basic [Cuteness Meter](https://petfinder.my/cutenessmeter) to rank pet photos. It analyzes picture composition and other factors compared to the performance of thousands of pet profiles. While this basic tool is helpful, it's still in an experimental stage and the algorithm could be improved.

In this competition, you’ll analyze raw images and metadata to predict the “Pawpularity” of pet photos. You'll train and test your model on PetFinder.my's thousands of pet profiles. Winning versions will offer accurate recommendations that will improve animal welfare.

If successful, your solution will be adapted into AI tools that will guide shelters and rescuers around the world to improve the appeal of their pet profiles, automatically enhancing photo quality and recommending composition improvements. As a result, stray dogs and cats can find their "furever" homes much faster. With a little assistance from the Kaggle community, many precious lives could be saved and more happy families created.

Top participants may be invited to collaborate on implementing their solutions and creatively improve global animal welfare with their AI skills.

Et billede, der indeholder Hunderace, kæledyr, hund, pattedyr

Indhold genereret af kunstig intelligens kan være forkert.

**Submission File**

For each Id in the test set, you must predict a probability for the target variable, Pawpularity. The file should contain a header and have the following format:

Id, Pawpularity

0008dbfb52aa1dc6ee51ee02adf13537, 99.24

0014a7b528f1682f0cf3b73a991c17a0, 61.71

0019c1388dfcd30ac8b112fb4250c251, 6.23

00307b779c82716b240a24f028b0031b, 9.43

00320c6dd5b4223c62a9670110d47911, 70.89

etc.