

Obligatorisk innlevering 1 – DTE-2602 | H24

I denne oppgaven skal du lage et enkelt program for å klassifisere ulike dyrearter basert på utvalgte egenskaper.

Filen *data.py* inneholder en rekke observasjoner av ulike fysiske egenskaper hos forskjellige dyr. Observasjonene er organisert på følgende måte:

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	class_type
aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1	1	1
carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	4
catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1	0	1
cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2

...

wallaby	1	0	0	1	0	0	0	1	1	1	0	0	2	1	0	1	1
wasp	1	0	1	0	1	0	0	0	0	1	1	0	6	0	0	0	6
wolf	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
worm	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	7

De fleste egenskapene har verdier 0 (usant) eller 1 (sant) – for eksempel kan vi se at dyret «aardvark» har hår (kolonnen «hair» har verdien 1) men at den ikke legger egg («eggs» = 0) eller flyr («airborne» = 0). Kolonnen «legs» (antall bein) har imidlertid flere mulige verdier.

Class_type er et tall («label») som sier noe om hvilken type dyr det er (klasse):

1	2	3	4	5	6	7
Mammal	Bird	Reptile	Fish	Amphibian	Bug	Invertebrate

Observasjonene er delt opp i et treningssett (*data_train*) der vi vet hvilken klasse hvert dyr tilhører, og et testsett (*data_test*) der dette er ukjent.

Filen *classifier.py* har en tom metode *classify_animal(...)* som tar inn de ulike observasjonene som argument. Ta utgangspunkt i denne og skriv kode for å klassifisere dyrene som en av 7 mulige klasser. Du kan også bruke pytest og fila *test_classifier.py* for å teste koden din (denne koden brukes også på GitHub for å automatisk sjekke at koden din kjører). Hvordan du ellers løser oppgaven er helt opp til deg, og skal dokumenteres i form av en **kort vitenskapelig rapport i pdf-format (maks. 2 sider, ikke inkludert forside)**. Lagre rapporten som PDF sammen med koden i repo'et, og push alt til GitHub.

Hovedpoenget med denne oppgaven er at du skal bli kjent med verktøy som git og pytest (som brukes i alle innleveringer), og at du skal bli kjent med rapport-formatet som brukes i de karaktersatte oppgavene i faget. Det er på ingen måte forventet at du skal lage en perfekt klassifikator som det første du gjør i faget. Oppgaven er åpen, og du kan bruke hvilken fremgangsmåte du vil (et sett med if-setninger eller bruk av noen av de eksterne bibliotekene som brukes i faget, se *requirements.txt*). Det er viktigere at du bruker rapporten til å forklare hvordan du har tenkt enn å få høy nøyaktighet i klassifiseringen. Lykke til!