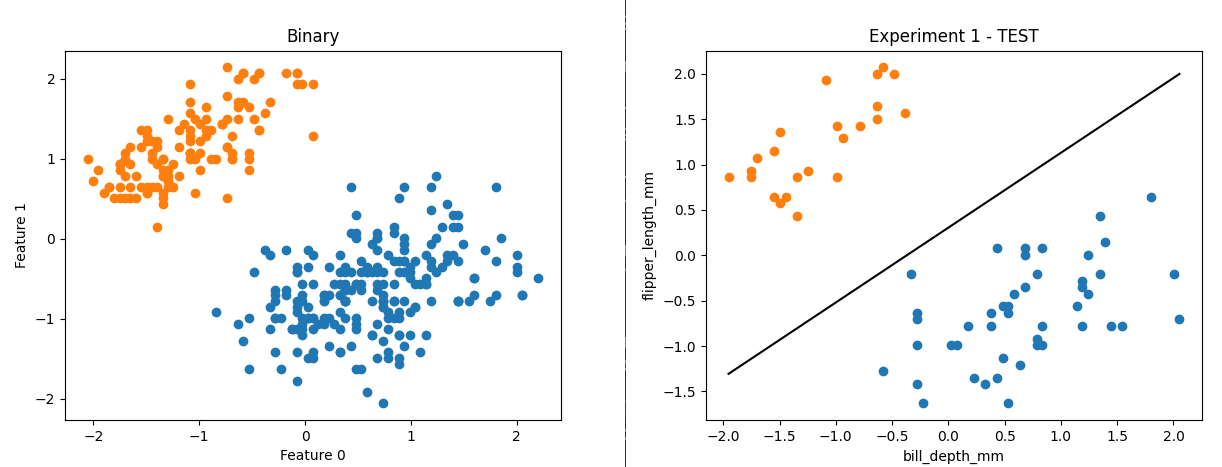
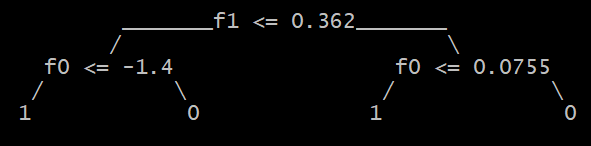
14.11.2024

Eirik Tennøfjord

Student



Supervised Learning

Karaktersatt oppgave 2 DTE2602 – Introduksjon Maskinlæring og AI

Innhold

[1. Introduksjon 1](#_Toc183305098)

[2. Teori 2](#_Toc183305099)

[3. Metode 2](#_Toc183305100)

[4. Resultat 2](#_Toc183305101)

[5. Diskusjon og konklusjon 5](#_Toc183305102)

[6. Kilder 6](#_Toc183305103)

# Introduksjon

I dag er hører vi stadig om at AI er i ferd med å ta over mange forskjellige jobber. Vinklingen på hvordan dette vil slå ut samfunnsmessig er ofte varierende. Det er ingen hemmelighet at de store automatiserte omveltningene gjør utslag på hvordan samfunnet utvikler seg. Bar innenfor programmering ser vi at tendensen øker når det gjelder å lage kode [1].

Spesielt med den utviklingen som har vært innenfor vision-system, så har en del av produktiviteten innenfor deler av industrien økt. Et eksempel på dette kan være sortering av fisk. I disse tilfellen bruker man et «kamera», som da kan sortere basert på hvilken art det er, samt anslå størrelse på fisken [2].

I denne rapporten skal vi se på noen metoder man kan benytte innenfor supervised learning, som går på klassifisering. Når vi snakker om supervised learning tenker vi på en metode innenfor maskinlæring, som bruker innsamlede data til å lære opp en algoritme til å eksempelvis kategorisere [3].

Videre vil vi i denne rapporten undersøke hvordan man ved å benytte ulike algoritmer, kan klassifisere tre forskjellige pingvinarter, basert på et datasett som viser art, \*\*\*lengder\*\*.

Decision Tree er en av metodene vi vil teste. Denne \*\*\*……\*\*\*.

Den neste metoden vi vil se litt på er perceptron.

\*\*\*\*I denne rapporten er målet å se om vi kan \*\*\*…..\*\*\*. Vi vil se om \*\* konvergerer\*\*. Finner man riktig klassifisering.

\*\*\*\* I denne rapporten vil vi se nærmere på hvordan de forskjellige egenskapene til perceptron og beslutningstrær kan variere fra lineært separerbare datasett, og de som ikke er lineært separerbar. Videre vil vi etter hvert sammenligne resultater for de forskjellige klassifiseringsmetodene.

# 2. Teori

\*\* Mer grundig forklaring av decision tree\*\*

\*\* Mer grundig forklaring av perceptron\*\*

\*\*NB: Beskriv: «vekter», «bias» og «gini impurity».\*\*

\*\* INKLUDER ligninger der variabler forklares. \*\*

\*\* Forklare hvordan metoder først trenes, deretter anvendes. \*\*

# 3. Metode

\*\* Beskriv alle valg som er gjort. Begrunn disse. \*\*

\*\* Dokumenter med referanse til koden hvor det er relevant \*\*

\*\* Skal være slik at andre kan gjenta utifra det som er beskrevet \*\*

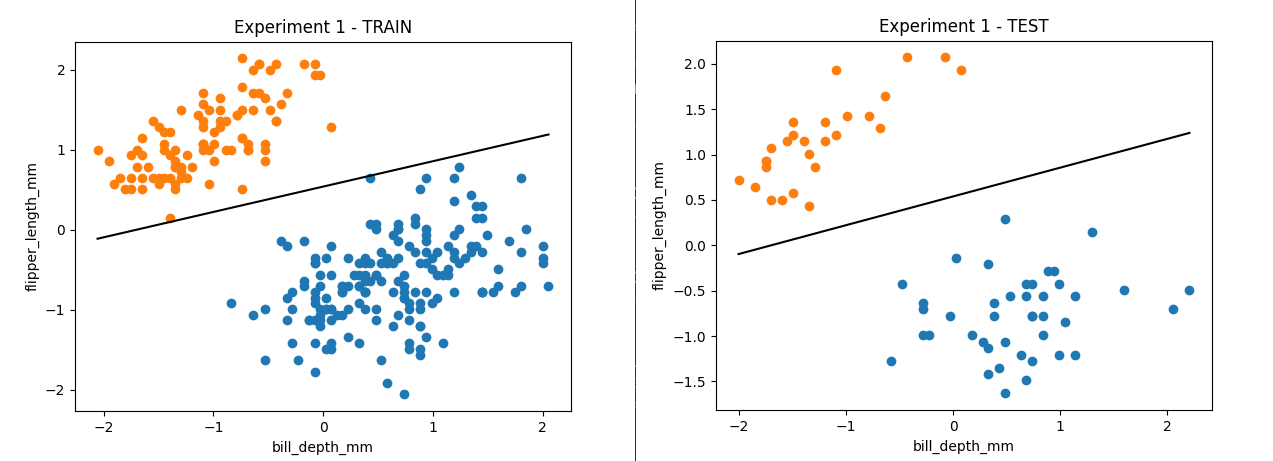
\*\* Hvilke verktøy/metoder som er benyttet \*\*

\*\* Implementasjonsdetaljer \*\*

Splitter datasettet i 80/20 [4].

# 4. Resultat

**Eksperiment 1:**

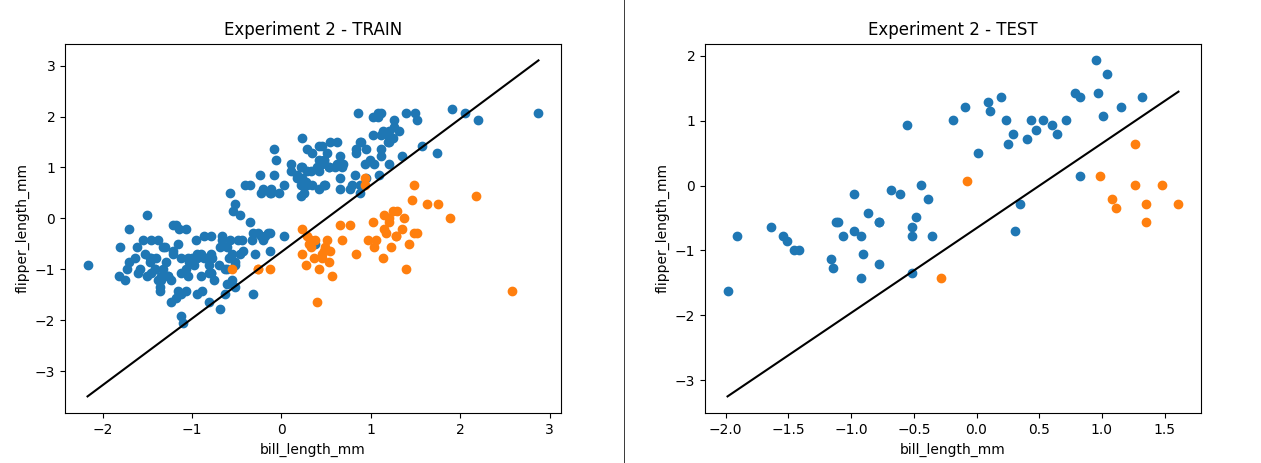
 Bilde x.xx – Viser hvordan treningsdata og testdata blir linjert skilt i eksperiment 1.

Konvergerer hver gang når man trener perceptron. Nøyaktighet (accuracy) mellom 0.96 og 1.0.

Bias: 0  
Epochs: 1000

Treningssett: 80%  
Testsett: 20%

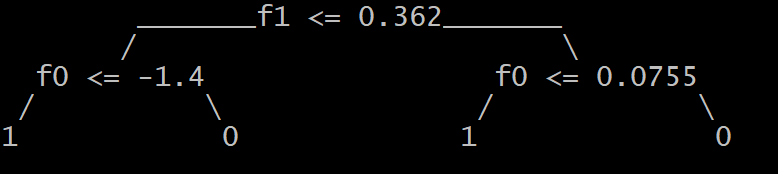
**Eksperiment 2:**

  
Bilde x.xx – Viser hvordan treningsdata og testdata blir linjert skilt i eksperiment 2.

Konverger ikke. Nøyaktigheten er 0.97 i dette eksempelet, men varierer mellom 0.85 og 0.96.  
Bias: 1  
Epochs: 1000

Treningssett: 80%  
Testsett: 20%

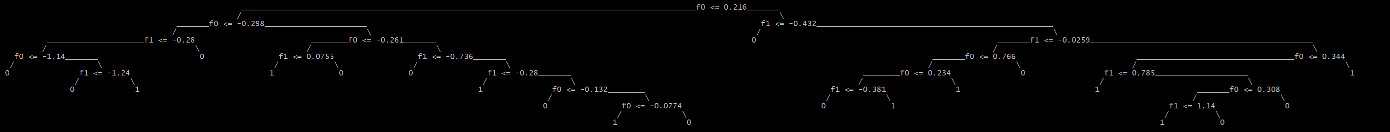
**Eksperiment 3:**

  
Bilde x.xx – Viser Decision Tree for eksperiment 3.

Treningssett: 80%  
Testsett: 20%

Vi kan se at utifra Decision tree så er spørsmålet fra index 1 det beste å stille som første spørsmål i root. Deretter bruker man spørsmålet fra index 0 for begge de to gren nodene.

**Eksperiment 4:**

  
Bilde x.xx – Viser Decision Tree for eksperiment 4.

Bildet over er noe utydelig, men det viser allikevel at det blir noe mer komplisert enn forrige eksperiment.

Treningssett: 80%  
Testsett: 20%

Nøyaktighet: Varierer mellom 0.90 og 0.96.

Sammenlignet med perceptron testen for samme art og feature, kan vi se at nøyaktigheten er noe bedre på decision tree. Dette er nok grunnet at det ikke er like lett å skille denne arten lineært basert på de to egenskapene som er valgt for denne testen.

**Eksperiment 5:**

Treningssett: 80%  
Testsett: 20%  
Simuleringer 1000

Ved tusen simuleringer fikk vi en nøyaktighet på 0.96

\*\* Resultater som er relevant til problemstillingen. Presentert på en objektiv og oversiktelig måte \*\*

\*\* Eksperimentene er designet for å undersøke relevante spørsmål, f.eks: variasjon i ytelse/tidsbruk når modellparametere endres \*\*

\*\* Dersom det er tilfeldigheter, så må man gjenta flere ganger, og deretter beskrive ved hjelp av statistikk. \*\*

\*\* Gjerne benytte tabeller, grafer, illustarasjoner og/eller bilder. \*\*

\*\* Presentere resultater opp mot tidligere beskrevet teori. \*\*

\*\* Kommenter gjerne kort i teksten for å fremheve de mest interessante / overraskende resultatene \*\*

**Perceptron - 2**

* Lag et nytt perceptron som skal skille arten **Chinstrap** fra de to andre. Bruk kun kolonnene *bill\_length\_mm* og *bill\_depth\_mm* fra X-matrisa (ikke samme som over!). Tren perceptron'et med ( X train , y train ). Merk at det ikke er sikkert at modellen konvergerer - forklar i så fall hvorfor. Visualiser "decision boundary" på samme måte som over. Mål nøyaktigheten til modellen med ( X test , y test ).

**Decision tree - 1**

* Lag et decision tree for å skille pingvinarten **Gentoo** fra de to andre artene basert på *bill\_depth\_mm* og *flipper\_length\_mm*, på samme måte som med perceptron. Mål nøyaktigheten og visualisér treet (rot, greinnoder, løvnoder). Kommentér: Gir verdiene til grein- og løvnodene mening?

**Decision tree - 2**

* Lag et decision tree for å skille pingvinarten **Chinstrap** fra de to andre artene basert på *bill\_length\_mm* og *bill\_depth\_mm*, på samme måte som med perceptron. Mål nøyaktigheten og visualisér treet. Sammenlign med resultatene for perceptron.

**Decision tree - 3**

* Gjenta følgende eksperiment flere ganger, med en ny tilfeldig "stokking" og oppdeling av ( X train , y train ) og ( X test , y test ) for hver gang: Lag et decision tree basert på alle 4 features i datasettet, som skal skille mellom alle 3 arter i datasettet. Bruk ( X test , y test ) for å måle nøyaktigheten. Oppgi statistikk for resultatene.
* Kommentér i rapporten: Hvorfor er klassifisering av mer enn to kategorier mulig med et decision tree, og ikke med et perceptron?

# 5. Diskusjon og konklusjon

\*\* Fremhever de mest overaskende/interessante delene av resultatet. \*\*

\*\* Kommer med mulige forklaringer til hvorfor resultatene ble som de ble. \*\*

\*\* Sammenligner relaterte resultater. \*\*

\*\* God selvinnsikt som står til levert kode og opnådde resultater \*\*

\*\* Drøfting av mulige feilkilder. \*\*

\*\* Beskrivelse av mulig fremtidig arbeid \*\*

**Sammenligning og oppsummering**

* Sammenlign resultatene for de ulike modellene du har trent og testet. Er det forskjeller mellom resultatene? Hva er sannsynlige grunner til det, i så fall?

I eksperiment nummer 4 kunne vi se at perceptron oftere hadde en høyere nøyaktighet enn decision tree. Dette kan være grunnet hvilken data som blir plukket ut til test og trening. Dersom testdataen er linært separerbar, så vil man ved å benytte perceptron kunne klassifisere punktene med god nøyaktighet. Dersom datapunktene ikke er liniært separerbare, så vil decision tree kunne være mer nøyaktig enn perceptron.

# 6. Kilder

1 - <https://www.kode24.no/artikkel/en-fjerdedel-av-koden-hos-google-blir-na-ai-generert/82156162>

2 - <https://optimar.no/solutions/product/species-recognition>

3 - <https://cloud.google.com/discover/what-is-supervised-learning>

[4] – *Ahmed*: The Motivation for Train-Test split. *Medium*. Hentet 23. November 2024 fra  
<https://medium.com/@nahmed3536/the-motivation-for-train-test-split-2b1837f596c3>

X – UiT Norges Arktiske Universitet (2024, 23. November) Gini Impurity – Introduksjon. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1026570>

X – UiT Norges Arktiske Universitet (2024, 23. November) Bruk av Gini Impurity for å vurdere splitt av datasett. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1026571>

X - UiT Norges Arktiske Universitet (2024, 23. November) Bruk av Gini impurity for å bygge beslutningstrær. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1026572>

X - UiT Norges Arktiske Universitet (2024, 23. November) Formatere datasett som (X,y)-par for trening og testing. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1026549>

X - UiT Norges Arktiske Universitet (2024, 23. November) Nevrale nettverk og perceptrons – Del 1 Introduksjon. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1026565>

X - UiT Norges Arktiske Universitet (2024, 23. November) Torsdagsmøte 20241107. Canvas.  
<https://uit.instructure.com/courses/34928/modules/items/1088128>

X - UiT Norges Arktiske Universitet (2024, 23. November)