P6 – Studieretningscase

Matematik og bioteknologi

2.D 2022-2023



# Diabetes og Machine Learning

Denne studieretningscase er et samarbejde mellem matematik A og bioteknologi A, hvor I skal arbejde med kunstig intelligens – nærmer betegnet machine learning. Kunstig intelligens kan beskrives som:

”Man bruger begrebet kunstig intelligens om maskiner, der kan træffe beslutninger selvstændigt og på egen hånd udføre opgaver, som ellers kun en dyre- eller menneskehjerne kan udføre” (Larsen, et al., 2021)

Hvor machine learning er kendetegnet ved:

”Når man udvikler en algoritme i en maskine, der kan analysere data, som den kan lære fra, kaldes det machine learning – på dansk maskinlæring. Maskinen bruger analyserne til at forudsige eller gætte på udfald af noget i den virkelige verden.” (Larsen, et al., 2021)

Studieretningscasen har til formål at træne jer i at undersøge og afgrænse en problemstilling, sådan at I bliver i stand til at lave en problemformulering, som efterfølgende løses ved at skrive en opgave hvori både matematik og bioteknologi indgår.

Opgaven handler om at lave et værktøj vha. machine learning for en praktiserende læge så lægen bliver bedre til at diagnosticere og henvise patienter med diabetes til sygehuset. For mange gange oplever sygehusene at de patienter, der bliver henvist til undersøgelse for om de har diabetes, er henvist forkert. Der findes forskellige former for diabetes – type 1 og type 2 er de mest udbredte men alt i alt mener man at der findes op til 13 forskellige typer af diabetes (Hulgard & Madsen, 2023). Der er forskellige fysiologiske faktorer også kaldet sundhedsdata, som indikerer at man har diabetes. Den typiske er at blodglukoseniveauet, som normalt ligger på 90 mg/100 mL, er forhøjet (Hulgard & Madsen, 2023).

Der skal udvikles en algoritme altså en kode i programmet ”Orange” (Demsar, et al., 2013) efter principperne i machine learning. Denne algoritme bruger en række forskellige sundhedsdata til at sortere patienterne ved hjælp af en GINI-koefficient. Algoritmen skal trænes ud fra data indsamlet fra 768 kvindelige Pima-indianere fra Phoenix i Arizona, USA, med en alder over 21 år (Kaggle, 2023; Github, 2023). Algoritmen i programmet laver et beslutningstræ, som den bruger til at sortere patienterne efter og dermed forudsige om de har diabetes eller ej. Denne algoritme vil altid sortere nogle patienter forkert, men der kan laves ændringer i beslutningstræet så fejlene minimeres.

Din opgave er nu følgende:

* At undersøge og afgrænse en problemstillingen ”For mange gange oplever sygehusene at de patienter, der bliver henvist til undersøgelse for om de har diabetes, er henvist forkert”
* Udarbejde en problemformulering med max 5 spørgsmål som gør det muligt at lave en redegørelse, analyse og diskussion i opgaven.

(En analyse i bioteknologi og matematik er at spørge hvorfor? Grundlaget, altså empirien, for analysen er de data og den algoritme samt resultaterne af programmet lavet ud fra principperne i machine learning.)

* At skrive en opgave som besvarer din problemformulering.

# Formalia

Se dokumentet ”Vejledning til DIO og P6 - Rev. 2021”.

# Tidsplan:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fredag d. 27/1** | **Fredag d. 3/2** | **Onsdag d. 8/2** | **Torsdag d. 9/2** | **Fredag d. 10/2** |
| *3 moduler* | *3 moduler* | *3 moduler bio* | *3 moduler bio* | *3 moduler bio* |
| Præsentation af studieretningscasen  Intro til machine learning Litteratursøgning | Problemformulering  Disposition for opgave | Opgaveskrivning | Opgaveskrivning | Opgaven er færdig kl 11.25 |
| Lektier til d. 3/2:  - Installation af Orange  - Litteratursøgning og -læsning  Fordybelsestid til lektier: 4t | Lektier til d. 8/2:  problemformulering,  disposition, læse litteratur  Fordybelsestid til lektie: 4t |
| *3 moduler mat* | *3 moduler mat* | *3 moduler mat* |
| Opgaveskrivning | Opgaveskrivning | Fagenes metoder og forberedelse til mundtlig fremlæggelse |
| Lektier til d. 17/3:  PP-show til fremlæggelsen |
| Fordybelsestid til lektie: 4t |

# Fremlæggelse

Der er evaluering d. 17.3.2023 i 1. – 3. modul, hvor du skal fremlægge din opgave med efterfølgende samtale og tilbagemelding på opgavens indhold. Evalueringen varer 15 minutter og du har 5 minutter til at fremlægge i.

# Evaluering

Der gives feedback ud fra nedenstående:

|  |  |
| --- | --- |
| **Opmærksomhedspunkter** | **Feedback** |
| Har du i din konklusion besvaret problemformuleringen? |  |
| Bevæger du dig både på et redegørende, analyserende og diskuterende niveau? |  |
| Har du udvalgt og anvendt relevant faglig viden, materiale og metode fra begge fag i din undersøgelse? |  |
| Argumenterer du med belæg for dine påstande, og dokumenterer du undervejs? |  |
| Har du faglig dybde i din opgave – ift. både studieområdet og de indgående fag? |  |

# Referencer

Hulgard, K. & Madsen, C.-M. V., 2023. *Biologibogen: Diabetes.* [Online]   
Available at: https://biologibogen.systime.dk/?id=435  
[Senest hentet eller vist den 25 01 2023].

Hulgard, K. & Madsen, C.-M. V., 2023. *Biologibogen: Insulin og glukagon.* [Online]   
Available at: https://biologibogen.systime.dk/?id=434  
[Senest hentet eller vist den 25 01 2023].

Kaggle, 2023. *kaggle.com.* [Online]   
Available at: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database  
[Senest hentet eller vist den 25 01 2023].

Larsen, T. M., Westen, V. & Skytt, L., 2021. *Kunstig intelligens.* [Online]   
Available at: https://faktalink.dk/titelliste/kunstig-intelligens  
[Senest hentet eller vist den 25 01 2023].