

Logbook

Learning goals

At kunne træne AI modeller på data

At kunne forstå datasæt

Være bekendt med at kode machine learning modeller

At kunne bruge forskellige machine learning modeller

At kunne evaluere modellernes performance

At bruge statistiske tests til evaluering

At kunne bruge modellerne til at bestemme hvilke patienter, der er i høj eller lav risiko

At kunne bruge f.eks. gaussian modeller til at forudsige manglende punkter
(manglende huller i modellen)

Template

- Questions
 - Questions here, and who to ask
- Reading, who and what
 - What was read?
- Implementation, who and what
 - What was implemented and worked on
- Results, who and what
 - Results of these implementations.
 - Do we need to reevaluate the results next time
- Decisions, who and what, what do you do alone, what do you do together
 - What decisions were made?

Project meetings

07/02/2024

- Questions
 - When do we get access to the project dataset? Ask supervisor
- Reading, who and what
 - No material was read today.
- Implementation, who and what
 - A colaboration-contract has been made containing the aproach and what is expected of eachother.
 - We have formulated some questions we want to ask the supervisor for the next meeting.
- Results, who and what
 - We have finished the colaboration - contract which is a part of the assignment the 28/2.
- Decisions, who and what, what do you do alone, what do you do together
 - A meeting has been set up with the supervisors next Tuesday 13/2 16:00. This will be a virtual meeting.

21/02/24

Vi (Ditte, Lucia, Viktor) mødtes til vejleder mødet, hvor vi efterfølgende, opdaterede vores gruppemedlem der ikke hørte starten af vejledermødet grundet forsinkelse på en halv time (Muneer). Dertil arbejde vi på at få udarbejdet en projekt beskrivelse, samt læse de links vi har fået givet omkring EuroSCORE.

Vi havde en 5-6 forskningsspørgsmål, vi lige ville lade ligge til efter forelæsningsen, og lige få klart syn på hvad vi ville i fagprojektet. Efter forelæsningsen havde vi møde med vores "søster-gruppe" hvor vi brainstormede på forskningsspørgsmål, dette møde deltager Ditte, Lucia og Viktor.

Brainstorm forskningsspørgsmål

No description has been provided for this image

Efter vi havde brainstormet, kiggede vi på vores forskningsspørgsmål og næsten færdig gjort vores udkast til projekt beskrivelsen. Projektbeskrivelsen, skrives primært at Ditte, Lucia og Viktor, da Muneer ikke havde sin computer med, og ikke har responderet på meddelser på Discord ang. ændringer i projekt beskrivelsen

27/02/24

kl 10.54 har Ditte Lucia og Viktor godkendt nyt vejleder møde, efter samtale på discord, hvor Muneer ikke har deltaget i samtale

28/02/24

Ditte, Lucia og Viktor mødes efter forelæsningen og skriver projekt beskrivelsen færdig, med de rettelser der er kommet fra vejleder.

Dertil udarbejdes der en mail til Muneer, ang. gruppe samarbejdet.

Der undersøges litteratur til vores projekt.

Ingen besked fra Muneer om fraværelse.

Projektbeskrivelse, Gantt Chart, Samarbejdsaftale, Project Canvas og Læringsmål afleveret.

Forløbige artikler

- <https://findit.dtu.dk/en/catalog/652c8a0f2fa76a1b70c0f344>
- <https://findit.dtu.dk/en/catalog/64bf66db76bd98312b7d8cd7>
- <https://findit.dtu.dk/en/catalog/6420dfb03a65a13c29450683>
- <https://findit.dtu.dk/en/catalog/5c47be8bd9001d019d32f1db>

Mail sendt til Muneer:

Hej. Vi oplever at der er nogle problemer og mangler i forhold til vores samarbejde.

Vi synes, at det er svært at have et samarbejde om projektet, når du kommer for sent til møderne, ikke dukker op til det aftalte, og ydermere ikke kommer og har en computer med til at kunne bidrage med at skrive på projekt beskrivelsen.

Dertil ønsker vi at der kommer kommunikation, som også har været manglende på det seneste. Jeg referer til vores sammenarbejdskontrakt, hvor vi har skrevet under på at man giver besked, når man ikke kommer. Dette hører vi intet om, idet svar og reaktioner på vores discord server er manglende.

I flere tilfælde har vi meddelt opdateringer i forhold til, hvor langt vi er med det nuværende arbejde, hvilke tanker vi har gjort os for fremtidige opgaver samt, hvad lægerne har meddelt over teams. Siden vi ikke hører noget fra din side, er vi usikre på, i hvor stor grad du er up-to-date med vores nuværende projekt status. Yderligere, ved vi heller ikke om du har sendt en underskrevet kontrakt til lægerne, hvilket et kravet for at kunne bruge computeren, og om du er klar over, at en nyt møde finder sted fredag med en hovedansvarlig fra hospitalets IT-afdeling.

Dette gør, at vi har nogle krav til dig, som du skal have bevist inden vores "midsvejsaflevering" om 3 uger. Er disse ting ikke blevet bedre, må vi desværre

meddele at du ikke kan være en del af vores gruppe derefter.

Hold øje med Discord og svar. Være der til, det vi aftaler (også selvom man har travlt, for det har vi alle) . Det forventes 9 timer om ugen og dette har vi ikke krævet over. Hold øje med teams og discord og svar på tingene (så det er tydeligt for os, at du har set beskederne og er opdateret). Medbringe computer hver gang.

Mvh Ditte, Lucia og Viktor

28/02/24 kl 15:40

Muneer bliver smidt ud af gruppe ved samtale med Morten Mørup, da han ikke har svaret i 2,5 uge. Hverken på Discord eller mail.

Muneer bliver fjernet fra Discord og git. Samme aften kontakter Muneer grupper, med besked hvad er der sket med vores discord gruppe.

Vi (Ditte, Lucia og Viktor) får at vide af Morten Mørup, at da han nu har svare skal vi stadig give ham en chance og derved skrive til ham. \

Mail koorspondance med Muneer:

06/03/24 kl. 19:50\

Hej,

Jeg bliver nødt til at indrømme at jeg først ser denne mail nu. Jeg har simpelthen haft en dårlig periode i den sidste måned grundet personlige årsager, der har forårsaget massere af fravær fysisk og skriftligt på DTU, og jeg er ikke selv tilfreds med mit bidrag og ustabilitet. Jeg er virkelig ked af jeg hvis jeg ikke kan være med fremover for jeg har mange gode ideer og tænker tit på projektet (har bare ikke haft mulighed på at arbejde på det tekniske grundet vi manglede data.) Jeg er tilbage på fødderne nu, og jeg håber i kan give mig en chance til, hvor jeg så lover jeg overholder vores aftaler mht. samarbejdskontrakten, for det er virkelig et projekt jeg vil arbejde på og have det bedste ud af.

Venlig hilsen,

Muneer\

07/03/24

kl 08:36

Kære Ditte, Lucia og Victor,

Kan se Muneer igår har svaret på jeres besked, inden jeg i dag ville skrive ud til ham og jer om gruppesamarbejdet.\

Jeg synes vi skal holde fast i planen, vi aftalte oprindeligt. Dvs. give Muneer opgaver til midtvejsafleveringen, og hvis der ikke leveres på disse som aftalt og samarbejdet med hans indsats stadig ikke er tilfredsstillende, må skridtet tages videre til at han

ikke kan være en del af jeres gruppe.

Vil I skrive til Muneer og sikre han har klare opgaver at levere på til jeres samarbejde, som aftalt.

Bedste hilsener

Morten

kl 08:40

Hej Morten.

Vi svarer Muneer på hans mail og holder fast i planen.

Mvh

Ditte Gilsfeldt

Studie nr: s210666

kl 08:52

Hej Muneer.

Vi giver dig en chance til i gruppen, men må indrømme at selvom man personlig har det svært, så er man også nødt til at være ansvarlig og få kommunikeret til sin gruppe. At der intet er hørt fra dig i 2 en halv uge, det er ikke okay. Vi kan jo lidt se du følger med på discord siden du samme dag opdager at du egentlig er ude af gruppen, samt flere gange er online, til at vi meget har undret os over den manglende kommunikation.

Vi vælger dog stadig med vores ret, at hvis tingene ikke bliver bedre inden midtvejsevalueringen, så kan du ikke være en del af gruppen.

Det gælder at du overholder følgende:

Får underskrevet kontrakten med Rigshospitalet, før dette kan du ikke bruge computeren.

Kommunikere til os og vejleder

Kommer til det aftalte til tiden og også med computer.

Dertil har vi andre heller ikke kunne arbejde på dataen da vi først får computeren i morgen, men det betyder ikke at der ikke har været ting der skulle laves.

Mvh

Ditte Gilsfeldt

Studie nr: s210666

kl 14:20

Hej,

Kan du hjælpe mig med den kontrakt der skal sendes til Rigshospitalet? Hvor er den og hvordan skal den sendes? Tak på forhånd.

Venlig hilsen,
Muneer

kl 17:11

Send mig lige din oplysninger til git og discord. Så tilføjer jeg dig igen. Og så ligger den der inde.

Mvh
Ditte Gilsfeldt
Studie nr: s210666

kl 17:55

Git: muneer-kayali
Discord: muneerkayali

Mvh,
Muneer

kl 20.07

Prøv lige at tjek om du har adgang nu. Ellers gør jeg det lige i morgen på computeren

Mvh
Ditte Gilsfeldt
Studie nr: s210666

08/03/24

kl 09.08

Hej Muneer.
Du burde have adgang når du godkender invitationerne.

Mvh
Ditte Gilsfeldt
Studie nr: s210666

kl 14.10

Hej, har stadig ikke adgang.

Mvh,
Muneer Kayali

kl 15:27

Har du fået invitationen?

Mvh Ditte Gilsfeldt Studie nr: s210666

Dertil kommer der ikke flere mails, og Ditte har sendt utallige invitationer til Discord gruppen samt venne anmodning der ikke bliver svaret. Ditte kontakter Muneer på Messenger:

10/03/24

15:28

Hej Muneer. Jeg har forsøgt i 2 dage at få dig ind i discord gruppen. Det er svært at få dig med i tingene når du kun svarer 1 gang på to timer. Forventer ikke du svarer hele tiden, men hvis du vil ind i discord gruppen, er du nødt til at kommunikere lidt mere med mig, så det kan lykkes.

Hilsen Ditte

Hvor Muneer svarer at han beklager men at han stadig ikke er med i gruppen på discord og at jeg kan sende ham en invitation, dertil svarer Ditte at der er sendt mange link, men at det ikke er muligt før han godkender venneanmodning og hertil kommer det op og kører.

13/03/2024

Ditte, Muneer og Viktor mødtes kl 10:00. Lucia er ikke til stede da hun er på ferie.

Idag blev der skrevet på rapporten, da midtvejsafleveringen snart skal afleveres. Der blev kigget på flere artikler for at få en bedre forståelse af hvordan mortality kan forudsiges ved brug af AI. Der blev også undersøgt hvilke muligheder for syntetisk data, vi har til rådighed.

potentielt library til syntetisk data: <https://github.com/sdv-dev/SDV>

06/06/2024

Lucia, Ditte og Viktor er tilstede for at starte med projektet. Intet hørt fra Muneer, og ingen resons på discord.

I dag, har vi startede med at kigge videre med at sortere dataen, samt kigge på om vi kan få syntetisk data op og kører.

Dertil har vi kun haft mulighed for at arbejde på en computer (rigshospitalets computer), da vi ikke kan flytte dateen fra computeren. Detter har gjort at vi sammen har kigget på det, men dertil er der også blevet kigget på det feedback der er blevet givet, som er ved at blive rettet til i vores rapport.

Det er blevet aftalt at mødes med "søster" gruppen, for at spare med hianden, på mandag kl 12.

10/06/24

Alle i gruppen er mødtes.

Vi har i dag haft møde med søstergruppen (se "referat" længere nede). Vi er gået i gang med at finde ud af hvad for dataen vi skal bruge og ikke bruge samt hvordan vi skal bruge det. Dertil har vi kigge lidt mere på rapport skrivningen, og fået rette vores state of the art, samt vores forskningsspørgsmål, da vi har fundet ud af at vi ikke kan sammenligne med EuroSCORE, som vi havde tænkt os.

11/06/24

Viktor, Lucia og Ditte er mødt op på biblioteket, men intet hørt fra Muneer.

State of the art er blevet skrevet færdig, og der er blevet skrevet til med stikord/ en disposition til metode afsnittet, hvor vi har været mere sikre på, hvilke modeller og evalueringstest, der skal bruges (med inspiration fra artiklerne vi har læst i state of the art).

Datasættet er blevet ryddet op yderligere, og det er næsten blevet delt op i forhold til, hvornår hvert person (ID nummer) har haft deres operationer.

Der er også blevet lavet udkast til kode, som vi skal bruge til at evaluerer vores modeller (altså når vi får defineret "prediction" vektorer og har "actual values" vektorerne).

12/06/24

Vi var alle tilstede for at arbejde videre på projektet. Noget af teorien blev færdig og vi kom langt med teori afsnittet. Vi fik kigget på nogle forskellige modellen, vi skal have ændret i og kørt på datasættet.

Vi er stadig i gang med at få cleanet dataen, og vi fik klaret en del af det i dag.

13/06/24

Ditte, Lucia og Viktor er tilstede, ingen besked fra Muneer. Vi er fortsat på at få cleanet dataen, og dertil laver vi videre på teori afsnittet. Vi har forsøgt og med at starte på noget metode afsnit, men det vi nu kan skrive om indtil videre.

Dertil er der sendt en mail afsted til Morten Mørup, ang vejleder møde på mandag.

14/06/24

Vi har alle siddet hjemmefra i dag, og arbejde videre på projektet, for at undersøge teori og metode.

15/06/24

Vi mødtes alle på DTU, og fik lavet mere kode hvor vi ryddede op i dataen. Dertil har vi kigget på FN's verdensmål som skal tages i betragtning til diskussionen.

16/06/24

Lucia, Viktor og Ditte mødtes på DTU som aftalt kl 11, og arbejde videre med det vi lavede i går, samt forberedte til vejleder møde mandag og tirsdag med vejleder.

Muneer sender besked til Viktor kl 13.50 om at der er EID så han tager til sin familie, men måske ville kigge på det senere.

Status

CSV filer gennemgået:

- Alkohol og rygning
- Anæstesiændelse
- Population
- Højde
- Vægt
- Problemliste
- Spirometri

Mangler

- KAG data
- Lab svar
- EKG
- Medicin (meget er givet under og efter operationen)
- Dialyse (pre-operativt (hvis de har nyresvigt inden operation ved lægerne, at det kommer til at gå virkelig skidt))

Forkastes:

- Anæstesidata (target variable)
- Alle operation lavet en uge efter den første operation
- ADT
- Avancerede hæmodynamoske mål
- Cirk

- Respirator data
- ITA (target variable)
- Blødning
- Ekko
- Viewpoint
- Pleuraddren
- Samlede udskillelse
- Total indgift
- Intellisapce

Ikke relevant for nu

- Dialyse (post-operativ)
- Urin (ikke bruges for pre-operativ, urin viser tegn på hvordan deres nyre er)

17/06/24

Vi har i dag holdt vejleder møde (1 time) med Lars, for at se om vi var på rette spor ang. deres ønske, samt få styr på nogle data filer som vi manglede at have styr på.

Dertil har vi arbejdet videre med at få implimenteret dataen og sorteret i de forskellige filer.

Dertil er der blevet kigget mere på rapport skrivningen.

Lavet en plan for resten af ugen:

- Mandag:
 - EKG og KAG bliver sorteret i dag og tilføjet til datasættet
- Tirsdag:
 - Medicin tilføjet til datasættet
 - Træning / test af modeller for pre operativt
- Onsdag:
 - Oprydning i under og efter operationen data
- Torsdag:
 - evt fortsætte, ellers går vi i gang med DL med dialyse og lungesyg
- Fredag:
 - fortsætter med deep learning
- Lørdag:

- Søndag:
 - Skrive på repporten
- Mandag:
 - Skrive repporten færdig

18/06/24

Vi havde i dag vejleder møde med Morten, der gave os nogle indput til hvordan vi skulle håndtere daten, samt fik justeret vores forskningsspørgsmål, samt modeller. Vi har derfor efter følgende arbejdet med den feedback vi fik af vejleder, samt arbejdet med rapportskrivning og implementering af modeller. Dertil har vi arbejde med at håndterer post operativt data, da vi er færdig med pre operativ data.

19/06/24

Ditte var på arbejde hele dagen, hvor hun kiggede på GBM modellen.

Muneer, Lucia og Viktor arbejdede med datasættet og selecte de forskellige features.

20/06/24

Dagen er gået med at kigge på de forskellige krav til opgaven, samt at få skrevet det sidste metode i forhold til feature selecting, og data. Dertil er der fortsat blevet arbejdet med de forskellige modeller, for at få dem til at køre.

21/06/24

I dag har vi kigget på nogen af resultaterne vi har fået gennem natten for modellerne for pre operativ data. Dertil havde vi møde med Morten hvor vi fik gennemgået nogen af de ting vi var i tvivl om.

Dertil har vi rettet til ud fra hvad vi skulle ændre fra vejleder mødet.

22/06/24

Dagen er gået med at skrive på rapporten, samt arbejde på modellerne, for at få resultater vi skal og kan bruge for alle modeller, og både pre, post og predi operativ data.

23/06/24

I dag er vi alle mødt for at skrive videre på reporten og kører det sidste på modellerne.

Vi har fået skrevet vores resultater ind, og mangler ikke meget på resusltager, så har vi aftale at vi hver i sær laver videre til i morgen, så vi kun har det sidste at skrive i morgen.

Muneer: Starter på diskussion i forhold til diskussion Lucia: Kigger resultater, og starter på model compareson, læse preface Viktor: kigger på model compareson og retter metode Ditte: kigger på AI værktøjer, læser igennem og se om vi opfylder krav, kigger på diskussion om etik og lægemiddelstyrrelsen

Alle skrive hvordan de har brugt AI værktøjer

Konklussion:

- Svar på research question

24/06/24

I dag har vi arbejdet på det sidste af rapport skrivningen, hvor vi alle har arbejde med at skrive samt lave alt det færdig der skal afleveres.

Supervisor meeting

13/02/2024 (virtual)

Questions:

- Da vi ikke er medicin studerende, hvordan ved vi hvad der er vigtigt?
 - Hvilke metrics kan vi evaluere?
- Hvordan finder vi balancen mellem hvor meget medicin' teori vi skal have med? - Hvornår får vi adgang til dataen
- Vi husker at du nævnte, at datasættet er meget stort. Uden medicinsk baggrund, hvordan skal vi evaluerer hvilke features der er "værd" eller relevante at tage med for træning af modeller (hvis vi ikke kan træne på alle features)?
 - Eller er det anbefalet at lave en form for feature selection eller statistiske analyse for at finde relevante features (analyse af correlation of påvirkninger)?

Notes from the meeting:

- Morten Mørup står for al ML delen og Lars står for Data
- Lars = Anastasi læge (Hjerte afdeling)
- Samlet data over 5 årigeperiode
- Data:
 - 8000 patient forløb
 - Alder, køn, fejl inden hjerte opretation, data under operation
 - data under operation er bla det de har samlt ind
 - Hvad data ser vi kun på en gang i mellem
 - Hvad data kigger vi mest på
 - Hvordan går det for patienter
 - er det her en risiko patient
 - er det end er kan risikere at dø, eller at det går dårligt
 - evt skal de ikke operaes, hvis de har risiko for at dø
 - Det her kan vi da lave noget ML til.
 - eller er det en lav risiko patient
 - Gemmer hvert minut dataen
 - Størrelse af datasættet
 - representativ
 - Nemt at håndtere
 - Overskueligt data
 - Tænk over anomyseringen
 - Jurisk er anomyseringen ikke godt nok
- Outcome:
 - Hvor længe overlever de
 - hvor meget medicin skal de have
 - Hvor meget support har de brug for
- Sampler sjældent
 - medicin
 - krugien
- Sample hyppigt
 - Fx EKG
- Der er allerede lavet ML på dataen
- Ny og spændende og værdifuldt for dem

- Kan vi undervej lavet en eller anden modellering undervejs af data der akn fortælle os nogen om tidligt i opreationen om det kommer til at gå dårligt
- Data pejer på at patienten ender i dialyse
- Euroscore siger noget om overlevelse men ikke om hvor lang tid forløbet tager
- Der laves en kontrakt til "ansættes" på rigshospitalet
- Få en laptop, der kan tilgå dataen.
 - Laptoppen har ingen GPU og kan kun kører det den kan
 - Kan godt tage laptoppen med sig
 - 1 laptop til hver gruppe
- Det er brgrænset hvilke informations kilder der hentes.
- Taler overlevelse
 - fx 90 dage
 - På et år er det deres samlde
- Tid på intensiv og tid på hospitalet ikke på operations bordet
- Hvis ikke man tager død så
- Forskningsspørgsmål:
 - i hvilken grad kan man få explatertbart, men samtidig forstå ...
- Teams gruppe
 - invitere vejleder med
- Mail til Lars
 - chaoqun.zheng@regionh.dk
- Mail til Chao
 - lars.groenlykke@regionh.dk
- baseline model (euro score)
 - samlede .. overlevede
 - logistisk regression
 - lave Euro score selv på data
 - se hvordan den beregner
 - se hvordan den laver logistisk regression
- Forstå modellen, eller om det bare at preditere godt?
 - Klinisk relevant del
 - State på behandling eller gøre om, så
 - kendes betydningen af enkelte komponenter

- vil gerne kunne handle
 - **Explanability er vigtig**
 - Højt sat,
 - er der en biologisk plausibilitet
 - hvad kan være problematisk
 - kan forbygge
 - modeller hvor det er nemmere at forklare
- Skal kende de enkelte datakilder og redegøre og hvor de kommer fra
 - Vigtigt for at kunne lave modellerne korrekt
 - Standard mål, der er velbeskrevet
 - Det medicin vi skal kunne vil fx være, blodtryksmedicin til at sænke blodtrykket
- Lars og Chao kan hjælpe med:
 - Hvornår er blodtrykket ligesom nonsens
 - Er det her overhoved biologisk muligt
 - Data afgrænsning (god kvalitet af dataen)

21/02/2024

- data.gov
- GMM af dataset for brug på vores egen computer
- aflevering i denne uge/næste uge
- align spørgsmål med hospitalet
- Arbejdsspørgsmål
 - 30/90 dages overlevelsesrate(vigtig)
 - Bedre performance end tidligere model
 - Hvor meget mere giver data under operationen?
 - indlæggelsestid
 - længde på operation
 - cor mellem de forskellige target variables
 - Hvad betyder det hvis patienter dør under operationen
 - Bliver de bedre til operationen over tid?
 - 3 forskningsspørgsmål
- Forskningsspørgsmål kan godt ændres hvis det er nødvendigt
- Hvis vi ikke får bedre resultater end logistic regression er det ok, bare det er

gjort rigtigt

- Transformer Deep Learning
- Opdeling i grupper af dataset
- Samme performance metrics for alle models

1/03/2024

- Møde med Lars og Theis omkring computer og data
- Alle i gruppen er til stede
- **Noter omkring data:**
 - Data fås i Excel format
 - Måles i følgende "attributes":
 - CPR, Alder på operationstidspunkt, Køn, Dødsdato, Død inden for 1 år af operationen, operationsdato, Operations SKS kode og operationsnavn
 - Vi kigger på 8000 patienter, men der er målinger for op til 46000 patienter
 - Hvis der er et eksempel men en værdi er 0 (som pulsen) men blodtryk og andre værdier ikke er 0, så er der sket en fejlmeldning
- Omkring operationens gennemgang:
 - Stareter med induktion og går derefter til bypass indtil hjertet stopper, derfra kan man tage Aorta tungen af. Herefter stoppes bypass og data indsamling stoppes.
- Fokuspunkter på forskningsspørgsmål:
 - Gå efter target "død"

22/03/2024 (virtual)

- Alle var tilstedeværende til mødet
- Spørgsmål omkring computeren blev stillet, men løsning vedrørende adgang til at pip installere pakker mangler stadig
- Lars gennemgik diverse Excel data filer, herunder hvad de indeholdte og forklarede, at hver patient havde et "falsk" ID navn for notationens skyld (denne, for hver patient, er det samme gennem alle filerne)

- For eksempel viser et Excel dokument Alkohol og rygning forbundet med en vis usikkerhed (under samme filnavn)
- Enheden for målinger i Excel filen "Advancerede hæmofynamiske mål" er målt i mililiter kviksøl (står udmiddelbart ikke i selve filen)
- Problemlisten er en vigtig film som lister patienters sygdomme, som for eksempel at nogen har haft blodprop eller hjertestop før eller lider af sukkersyge

Flere noter med hensyn til variable i diverse Excel filer:

Alkohol:

- Anæstasi er fra respirator
- Blodtryk måles på 2 forskellige måder
- Kateter i pulsåre i håndleddet og bånd i armen
- Giver samme resultat, kater mere præcis
- Puls og frem er fra monitor alt over er fra respirator

I cirk:

- Ting kan optræde flere gange alt efter hvor de henter data fra. Map tryk kan udregnes ud fra blodtrykket
- Missing values fordi de ikke er gået i gang endnu
- Nogen får målt tryk i længere ikke alle og derfor vil der være mange missing values Anæstasihændelse
- Induktion er når patienten bliver bedøvet på stuen. Stop er når patient bliver kørt ud

Cv bypass:

- At klemme aorta af er en vigtig måling
- Har betydning for hvordan de går dem. Blødning
- Kan lave en summering af hvor meget de har blødt
- Dialyse er ikke vigtig lige nu
- Outcome i hvordan går det dem
- Er de i dialyse eller ej

Ekko:

- Skannet deres hjerte med ultralyd
- Nogen er ikke relevante og det finder chao, Theis og Lars ud af

Intellispace:

- Samme som ekko bare med meget mere data kan have noget der er målt to gange

Ita:

- Hvor længe er de på intensiv, ligegyldig om det er på riget eller Hvidovre

Kag:

- Undersøgelse af krans pulsåre
- Mange patienter har fået lavet bypass.
- Krans kan have været forsnævret ellers helt lukket af
- Hvilke kar og hvor meget er de lukket af

Lab svar:

- Alle de blodprøver de får målt
- Her skal alt ikke bruges
- Men stiger nyretallet og stiger blodprocenten.
- Får de infektioner

Dræn:

- Lunge og væskhule
- Væske de taber
- Ikke så mange af dem
- Problemliste.
- Hvor syge var de i forvejen?
- Hvad fejlede de?
- Har stor indflydelse på hvordan det går dem

Respirator:

- Indstillinger fra respirator
- 2 forskellige maskiner så dataen kan godt være lidt forskellig
- Hvis de stopper med at trække vejret, hvornår skal maskinen så gøre noget. Det er ikke vigtigt

Spir:

- Hvor lunge syge er de når de går til operation

Total:

- Hvad får de ind
- Væske og medicin
- Filtre på tiden, så kan man se under operationen

17/06/24 (Vituel møde med Lars)

DI og DJ problemliste er alle de diagnoser de har ud over det, der kan man ikke sige at noget ikke er relevant, Det burde mordellerne finde, (hvis en har en øjnsygdom, så kommer den ikke til at bone ud.)

- Husk: DE, DN (Nyreinsufficciens??)
 - Kan godt sortere i DN
- Sorteres efter hyppighed, 10-20, hvis der er nogen han ikke har sagt, kan vi tage det med Populationen af patienter skal tages for population.csv, deres operationskoder,
- Mange vil have sukkersyge og forhøjet blodtryk, det vil den bone ud på
 - Blodsygdomme, leversygdomme, og nyresygdomme
 - Infektion i blodet kommer ikke frem i DI, Det vil være under DA, men det vil være noget de har bagefter, det er ikke noget de kommer til operationen med
 - Hvad har folk medst?

Operations koderne er det relevant at tage det hele mange vil gå igen, og nogen vil være mindre hyppige og der har vi ikke nok om data til at det bliver brugbart.

Univariat regression og en multivariat regression

God idé med chi-squared

Ekko data: droppe det, fordi der er mange ting der måler det samme men på forskellige måder og burde være meget korreleret, men med forskellig usikkerhed.

Hvordan skal vi håndtere de forskellige vi mangler?

- EKG: der kan sagtens være at qtc intervallet er relevant, men at rtc ikke er relevant
 - Tage hver ting for sig
- KAG: Den viser hvilke kar der er lukket sammen, hvor sidder aflukningen af det kar
 - 40% halv åben
 - LAD 90% er værre end en ACA på 90%
 - lave en interaktion mellem de to
 - Det er værre at have en ACA på 90% end en LAD på 40%
 - Lav en variable for hver af de kar
- LAB svar: ID1, EGFR/1 er patientes nyrefunktion
 - hænger sammen med kretanin
 - Hæmoglobin (hvor mange blodplader har de)
 - Lav hæmoglobin er at det nok ikke er optimalt til at starte med, så det kan have en indflydelse
 - Plastman glykose kan man også kigge på (pre sukkersyge?)
 - Eosytter, forhøjet kan være tegn på inflammation
 - Tromposytter for lave kan være at patienten har en form for kronisk påvirkning, patienten ikke er i balance
 - Troponin,
 - Kretinkrinase NB
 - Kretinkrinase P
 - Hjertet ikke får blod nok
 -
- Fint bare at tage den første operation, da det er den det drejer sig om
- ADT kan vi ikke bruge til noget, mere administrativt
- Lungesygdom, dialyse, indlæggelses tid
- Ikke pleaudren
- Urin er et udtryk for nyrefunktion, nul eller lidt urin er et tegn på det ikke blive produksere, skal ikke bruges som pre,
- Medicin:

- Til træning er det kun det medicin de får inden
- Ikke noget der nødvendigvis er irrelevant
- medicin for forhøjet blodtryk
- vandrivende medicin for nyre svigt
- Smertestillende, folk der tager fast smertemedicin
- Timestamp:
 - (administrations tid)
 - Dato ?
 - ny medicin fil med timestamp
- Meget af medicinen er givet under, eller
- er kategoriseret i ATC ?
- I forhold til dosis: Dosis kigger vi ikke på, JA NEJ rent binært
 - post
 - meget af det bliver ikke givet som hjemme medicin
 - typemedicin så skal man tilføje antal slags
 -
- viewpoint er under EKKO
- Hvis de allerede er dialyse patienter, så går det rigtig dårligt for dem - Kig på den også
- Prioritere hvis vi kommer under pres?
 - dødlighed under pre operative
 - Forudsige forskellige grupper ud for korrelation
- Lars forlår at kigge informative for helt sætteoptte
- bare tage en 30 eller 90
- om vi får prediteret alle de forkellige ting, nyre er fint nok
- det sjove er at kunne få data til at arbejde for sig.
- Kan være muligt at fortsætte på sit bachelor projekt (spørg Morten)
 - De er muligt for Lars

18/06/24 - Møde med Morten

- Perspektivering
 - Undersøge fairness
 - Kritisk blik
 - Hvad skal der til når mam tager den ud i hospitalet

- Bias etik og fairness
- Hvis en læge blindt stoler på AI
- Figur tekst, oversæt til engelsk, eller fodnote
- For future research, forsknings spørgsmål 2
 - Ikke have det som et research question
 - ikke noget forskningsspørgsmål, vi ikke får svar på
 - Forklare hvorfor det er væk
 - kan godt være vi godt kan nå det
- performance
- Forsknings spørgsmål: ændre fra korrelation til performance
 - Mere interessant
- Referet til artikler \citep \citetet (Giver forfatternavn)

Genlæs kontrakten

- Resultater
 - Kan man se enkelte datapunkter eller kan man ikke
- Data:
 - Ikke interpolation
 - droppe features der har rigtig mange NaN
 - Hvis det er NaN, er det så overhoved noget man kan bruge i virkeligheden
 - Det går ikke at droppe næsten halvdelen
 - Binær variable, gode eller dårlige lunger
 - Drop dem selvom de er vigtige, hvis man har målt den, er fordi man nok har tænkt der var noget, som kan give bias i vores population
- Medicin:
 - Data matrice one-out-of-k, og data komprimere (PCA på datamatrice)
 - Group
- Ingen RNN, vi kan ikke nå det, brug det i perspektivering
- MLP på tabulær data
- Tabulær data,
- Alt vi har gjort at overvejelser kan vi skrive i diskution
- For den bedste model, performance

- RUC AUC og PA AUC og accuracy, som performance mål
- Hyperparameter tuning
- Logiske regression skal ...
- RF , to kryds validering, kigger på deres hyperparameter
 - indre folder tuner vi modeller 5 fold minimal (vis error)
 - ydre folder ..
- Valg vi har truffet:
 - Drop chi-squared
 - Start med nyresvigt
 - Drop de data der har alt for meget missing
 - tidsrækker, der skal omkonverteret til tabulær data, mean, max, 4 variabler pr tidsrække ...
 -
- Skrive i metode at vi tager udgangspunkt i den her, som også outperforms euroscore.
- hvad fandt vi ud af, hvad var godt.
- De modeller de har fat i, er det vi tager fat i.
- Kort fortælle om vores modeller
 - SVM, kender vi ikke
 - Gradient boosting, kender vi ikke
 - De andre modeller kan vi godt kort fortælle om, ikke tænke ud om
 - Skal kunne redegøre for vores modeller
 - PA AUC kender vi ikke, tærskel
 - RUC AUC kender vi
- Gennemsnit og standard deviation
- hvordan er metoder i forhold til hinanden
 - McNemar, til at sammenligne
 - Bygger på accuracy ..
 - hvad er accuracy og baseline i accuracy,

21/06/24 - Møde med Morten

Morten hjalp med os med at få styr på det sidste på vores modeller og med at ændre vores neurale netværk, så vi ikke har for mange hidden layers og alt for mange

noder. Dertil hjalp han os med at få styr på det formelle til rapporten og eksamen. Men også med de ting vi have tvivl om i forhold til vores rapport skrivning, hvortil vi ændrede opstillingen og skrev metode og teori sammen og droppede teori afsnittet. Dertil indførte vi et afsnit om implementering.

Spørgsmål til møder:

18/06/24

Møde med Morten

- Må vi bruge plots det er lavet på det rigtige dataset, eller skal plots og performance også være fra et syntetisk dataset?
 - eller rigtige tal
- Kan Morten hjælpe med at kigge på vores kode og metode og se om vi er på rette vej, eller om han kan guide i hvilken vej vi skal
- Er vores forskningsspørgsmål overhoved muligt?
- Hvordan håndterer vi de flere diagnoser fra hver patient? (SKS)
- Kan Morten hjælpe med hvad vi skal gøre med missing values?
 - skal de fjernes? (væsentlig reduceret datasæt)
 - eller skal vi gætte (interpolation) på de missings values ud fra vores andre variabler? (ranges og laver et kvalificeret gæt)
- Format ang. FN verdensmål
 - hvordan referer man?
- Hvordan formelt referer man til andre artikler
- Hvordan formelt referer man til citater fra andre artikler
- Kan det være muligt at fortsætte til sit bachelor projekt, Lars siger de altid gerne vil samarbejde
- Hvor meget skal vi gå i dybde med metode i forhold til hvad vi har gjort i koden?
- Hvordan gør vi med Medicin tabellen?

17/06/24

Møde med Lars

- hvordan skal vi håndtere diagnoserne, i forhold til vigtigheden af deres sygdomme?
- Hvilket data er mest relevant
- Hvad er mest relevant i ekko, lab svar, ?
- Skal vi måle kigge på FVC og FEV1 og FEV1/FVC (ratioen), eller er det nok bare med FEV1/FVC (ratioen)?
- Hvad er veipoint?
- Hvad med Dialyse?
- Hvad er Pleuraddren?
- Hvad er KAG
- Finde ud af hvor meget af det her er relevant? og om vi må forkaste tingene vi tænker at forkaste
- Der var lungesygdom, dialyse og hvad er den sidstesygdom som vi skal forudsige?

Møde med feedback gruppe

Onsdag den 3 april, blev der afholdt feedback møde angående det første "færdige" del af fagprojektet.

Vi aftalte med den anden gruppe at mødes kl 14 i 324 på DTU.

Inden mødet, mødtes vi alle 4 for at udarbejde et skriv til feedback gruppen, baseret på de 3 c'er og burger modellen.

Vi alle 4 mødtes med feedbackgruppen. Mødet gik godt, de tog godt imod vores feedback og var alle aktiv lyttende. Der var dertil også samtale om nogen af kommentarerne, hvilket fungerede godt.

Stemningen var god, og alle virkede glade.

Feedback gruppen gav god feedback til vores opgave, og vi var alle dertil også aktiv lyttende.

Der var dertil god samtale om ting i hinanden opgave, der var inspirerende, som hver

gruppe ville tilføje til deres egen opgave.

Mødet varede 45 minutter, og det blev aftalt at sende feedbacken på mail til hinanden.

Efter mødet, blev Viktor, Lucia og Muneer og lige skrev det sidste få linjer på feedbacken så den kunne blive sendt afsted til den anden gruppe. Grundet til at de sidste linjer ikke blev gjort færdig, var at feedback gruppen, kom 10 min før aftalt tid.

Det blev aftalt at Ditte skrev logbog, hvor de andre skrev det sidste.

Opdatering af projektplan

Tirsdag den 1. maj gik vi i gang med at arbejde på den opdateret plan af projektet. Vi mødtes virtuelt og alle var til stede. Her lavede vi en Gantt chart, hvori vi lagde en plan for hvilke datoer vi ville arbejde med hvad og sat deadlines på de forskellige sections. Gantt charten hjalp os med at få et overblik over hvor lang tid vi skal bruge på de forskellige sectioner, og om det er feasible. Derudover har vi opdateret vores project canvas så det passer med alt vi har gennemarbejdet og alt vi har planlagt at lave. Sidst opdaterede vi vores projektbeskrivelse hvor der blev taget højde for den feedback vi fik fra vores peers.

Møder med "søster" gruppe

10/06/24

Kl 12.00

Alle i vores gruppe mødtes med søster gruppen for at spare ang. fagprojektet. Vi talte om hvordan data bliver sorteret samt det at lave syntetisk data eller ej. Dertil var vi inde på hvilke forskellige modeller vi hver især benytter, samt hvad vores fokus er / forskningsspørgsmål. Vi er blevet klogere på at vi ikke skal bruge mere fokus på at finde ud af at generere syntetisk data, da det kræver enormt meget. Dertil kan vi lave python fil, som kun indeholder kode, som bliver det der bliver uploadet til git, så dataen ikke kommer til git.

Dertil er vi blevet klogere på at forstå daten, da vores søster gruppe har brugt meget tid på at forstå dataen.

Ydermere, har vi talt om hvordan vi forholder os til EuroSCORE, og vi har måtte erkende at vi ikke kan holde vores modeller op til EuroSCORE, da vi ikke kan få EuroSCOREN på patienterne i dataen, og dertil har vi ikke de samme features i

datasættet, som EuroSCOREN bygger på.