

Prosjektplan¹

Navn på sandkasseprosjekt:

- *EpiAnon: Anonymisering av epigenetiske data til bruk i maskinlæring*

Navn på virksomhet(er):

- Age Labs AS

Deltakere fra virksomhet:

Navn	Rolle	E-post	Telefon
Karl Trygve Kalleberg (KTK)	CEO		
Espen Riskedal (ERI)	CTO		
Arne Søråas (AVS)	CSO		

Deltakere fra Datatilsynet:

Navn	Rolle	E-post	Telefon
Eirik Gulbrandsen (EGU)	Kontaktperson, Senioringeniør		
Grete Alhaug (GRA)	Senior Rådgiver, jurist		
Sindre Dyrhovden (SID)	Rådgiver, jurist		

1. Kort beskrivelse av prosjektet og formålet med planen

Age Labs lager tidlig diagnostikk for aldersrelatert sykdom basert på maskinlæring og epigenetikk. Kjernen i selskapet er en plattform som produserer "algoritmedrevne blodprøver". Ved å lese av den epigenetiske profilen til en pasient ut fra en blodprøve, kan algoritmene brukes til å identifisere mønstre som sier noe om fremtidig sykdomsutvikling. Dette muliggjør persontilpasset prediktiv diagnostikk.

I forbindelse med utvikling av slike algoritmer i FoU-kontekst er det vanlig å benytte aidentifiserte, sensitive helseopplysninger. Det er to grunner til å benytte anonymiserte data: (1) dele til tredjeparten med tanke på validering og reproduserbarhet (vitenskapelige publikasjoner), og (2) til bruk for forbedring av algoritmen (etterlæring). I dette sandkasseprosjektet ønsker vi spesielt å fokusere på det siste av disse punktene. Samtidig vil bruken av aidentifiserte/ pseudonymiserte data være nødvendig i deler av prosessen for å understøtte klinisk bruk av løsningen.

Formålet med denne planen er at Age Labs og Datatilsynet har en felles plan som beskriver mål, leveranser, aktiviteter, ansvarsfordeling og tidsfrister i sandkasseprosjektet. Planen kan endres ved behov.

¹ Eksempel på [prosjektplan fra Finanstilsynet sin sandkasse](#)

2. Mål og forventede leveranser i prosjektet

Age Labs er i FoU-stadiet for sin første diagnostiske test for revmatologisk sykdom. Vi ønsker at sandkasseprosjektet fokusere på muligheten for å anonymisere og/eller aidentifisere helsedata som er nødvendig for akkurat denne testen.

- Hvordan sikrer vi anonymisering av epigenetiske data og kliniske kovariater?
- Hvordan går vi frem, med tanke på anonymisering og dataminimalisering, når vi skal "etterlære" algoritmen vår etter at den er satt i produksjon, og hvilke hjemmel/behandlingsgrunnlag kommer til anvendelse?

Mål 1: Behandlingsgrunnlag

- Leveranse 1.1: Vurdering av behandlingsgrunnlag som trengs
- Leveranse 1.2: Vurdering av påvirkning av ny helselov (evt)

Mål 2: Teste hypotesen vår om at epigenetiske helsedata kan anonymiseres i vår kontekst

- Leveranse 2.1: Praktisk lakmustrer for om epigenetiske data er anonyme eller ei
- Leveranse 2.2: Demonstrator (programvare) som viser dette på et ekte datasett
- Leveranse 2.3: Delrapport om erfaringene fra lakmustrer og demonstrator

Mål 3: Avklare alternativer til anonymisering for etterlæring, dersom anonymisering ikke lar seg gjennomføre

- Leveranse 3.1: Detaljert oversikt over informasjonsflyt for etterlæring (diagram)
- Leveranse 3.2: Risikovurdering for denne arkitekturen (DPIA)

3. Aktiviteter, metode og arbeidsfordeling

Prosjektplanen er laget for å teste den initielle hypotesen til Age Labs om at det vil være mulig å anonymisere epigenetiske datasett for selskapets konkrete diagnostiske anvendelse. Slik anonymisering vil ha anvendelse i selskapets FoU-fase og i en senere produksjonsfasen. Da hypotesen kan vise seg å ikke holde, er siste del av prosjektet satt av til å vurdere alternative fremgangsmåter for produksjonsfasen, der man da vil benytte seg av identifiserte eller aidentifiserte data.

Metode. Prosjekt vil primært benytte seg av arbeidsmøter (workshops) med eksperter fra DT og AL. Det kan være aktuelt å invitere inn fageksperter i vårt eget nettverk for enkelte av møtene. Det er aktuelt å gjøre litteratursøk for å se på eksisterende arbeid på feltet. Det planlegges også utvikling av en liten demonstrator (programvare) for å illustrere dataminimalisering og anonymisering for et helt konkret og praktisk tilfelle.

Workshop 1: Anonymisering av epigenetiske helsedata (tilnærming / teknikker / metoder), del 1

- Diskusjon av eksempeldatasett for å vurdere graden av anonymitet
- Hva som skal til for at individdata er å anse som anonyme?
- Sammenligning/inspirasjon: hvilke kriterier legger SSB til grunn for at et datasett er anonymt?
- Forberedelse til litteratursøk for anonymisering av epigenetiske data

Workshop 2: Risikovurdering og vurdering av behov for samtykke ved etterlæring

- Vurdering av behandlingsgrunnlag / roller (inkl skisse)
- Ny lovhjemmel (i kontekst spesifikk løsning); spesielt vurdering rundt om det er rett å tenke på etterlæring i vår situasjon som beslutningsstøtte, eller om det er kvalitetsforbedring

Workshop 3: Anonymisering av epigenetiske helsedata (tilnærming / teknikker / metoder), del 2

- Gjennomgå litteratursøket
- Diskutere mulige anonymiseringsstrategier basert på egen kunnskap og litteratursøk
- Forberedelse til utvikling av demonstrator

Workshop 4: Gjennomgang av demonstrator for anonymisering

- Har vi oppnådd anonymisering ut fra kriteriene vi satt?
- Hvilke usikkerhetsmomenter gjenstår evt?
- I hvilken grad er dette generaliserbart til andre omics?

Sluttrapport: Oppsummering av erfaringene fra prosjektet i en strukturert sluttrapport.

4. Tidsplan for sandkasseaktivitetene

Tidspunkt	Aktivitet	Deltakere
30. april – 13:00-14:30	Workshop 1	KTK, ERI, AVS, EGU, GRA, SID
	Litteratursøk	KTK, AVS
10. mai – 11:00-13:00	Workshop 2	KTK, AVS, EGU, GRA, SID
	Utvikling av demonstrator	KTK, AVS, (ERI)
11. mai – 12:00-12:45	Podcast / intervju	KTK, EGU
31. juni (TBD)	Workshop 3	KTK, AVS, EGU, GRA, SID
Sommerferie		
15. august (TBD)	Workshop 4	KTK, ERI, EGU, GRA, SID
15. september (TBD)	Sluttrapport	KTK, ERI, AVS, EGU, GRA, SID

5. Kommunikasjon

Prosjektet vil dele resultater og erfaringer både eksternt og internt. Som et minimum vil prosjektplanen og sluttrapporten gjøres tilgjengelig utad, i tillegg til at milepæler og resultater underveis i prosjektet kan publiseres på prosjektets nettside eller i sandkassens podcast.

Det utarbeides en egen kommunikasjonsplan for prosjektet med detaljer rundt mål, målgrupper og planlagte tiltak for eksternt og intern kommunikasjon. Ansvar for framdrift og oppfølging av kommunikasjonsplanen ligger hos Datatilsynets kontaktperson i sandkasseprosjektet.