



RECOVER

Hackathon

Recover / Cogito NTNU

24.10.2025 - 25.10.2025

Behandling av data

Deltakere som under hackathonet lagrer data mottatt fra Recover—enten lokalt eller i skyen—plikter å slette all slik data umiddelbart etter hackathonets slutt. Bakgrunnen for dette er at materialet kan inneholde informasjon som vurderes som forretningssensitiv. Vi ber derfor alle deltakere være oppmerksomme på dette ansvaret, og sikre at ingen kopi av dataen oppbevares etter arrangementet.

1 Introduksjon

Recover er Skandinavias ledende aktør innen skadebegrensning og gjenoppbygging. I Norge samarbeider vi med alle de store forsikringsselskapene, og utfører omtrent 30 000 oppdrag i året—hovedsakelig forsikringsskader som vann- og brannskader.

Vi er over 1 200 ansatte her til lands, hvor de aller fleste er håndverkere som tømrere og malere. Vi har også en god del *sanerere* som er eksperter på desinfisering, rengjøring og luftfjerning etter uting som sot, sopp, kjemikaliesøl, og blod. “Sanering” er et ord dere kan komme borti i hackathonet, og det betyr rett og slett å *rense, dekontaminere, eller gjenopprette noe til en sunn tilstand*.

Når en skade oppstår, rykker vi ut på kort varsel og setter inn tiltak for å begrense skaden. Deretter dokumenterer vi skadestedet med bilder og rapporter, og kalkulerer skadeomfanget i detalj. Hvis kalkylen¹ får tomme opp fra forsikringsselskapet, starter vi arbeidet med å rive, tørke, og bygge opp igjen slik at skadestedet får tilbake sin opprinnelige stand.

1.1 Forretningsutvikling og analyse

Det er ikke bare håndverkere som jobber i Recover! For ca. tre år siden ble det etablert et team for forretningsutvikling og analyse, bestående av to kybernetikere og to økonomer. Teamet jobber i hovedsak med effektivisering, automatisering og analyse.

I løpet av denne tiden har vi utviklet apper og systemer som i dag er en helt sentral del av hverdagen til alle som jobber i Recover, og vi er blitt bransjeledende på teknologi. Vi jobber hver dag med å automatisere kjedelig papirarbeid, og ikke minst utvikler vi *KAI*, vår Kalkyle AI som hjelper prosjektledere og kalkulatører med å kvalitets sikre kalkylene våre. Nå ønsker deres innspill: hvordan kan *KAI* bli enda smartere?

¹En kalkyle er en lang liste over alle arbeidsoperasjoner som skal utføres på skaden. Til å lage kalkylen, bruker vi en programvare som har flere hundre standardoperasjoner, for eksempel “Riv dørlist”, “Gåtid fra parkering til skadested”, “Bompenger” og så videre. Kalkylen utarbeides enten av prosjektleder, eller spesialister som kalles kalkulatører. Programvaren regner ut forventet tidsbruk på arbeidsoperasjonene, som igjen danner grunnlag for prisen Recover får betalt av forsikringsselskapet for arbeidet, basert på avtalt timespris.

2 Oppgaven

2.1 Manglende arbeidsoperasjoner

Gjennom flere år har vi samlet titusenvís av kalkyler. Dette gir oss et grunnlag for å se på muligheter til å effektivisere arbeidet. Spørsmålet vi stiller oss er: Kan vi bruke erfaringen og datamengden til å hjelpe kalkulatørene i arbeidet—for eksempel ved å foreslå operasjoner basert på tidligere lignende saker? Greier vi å identifisere mønstre i de gamle kalkylene som kan hjelpe oss med å lage gode kalkyler i dag?

Problem 1: Manglende arbeidsoperasjoner (40 %)

I forsikringssaker får vi stort sett ikke betalt per time vi jobber, men per arbeidsoperasjon vi utfører. Det er dermed avgjørende å få med seg hver enkelt operasjon som utføres i hvert rom—for eksempel “*Riv gulvlist, 2.5 meter*” eller “*Riv parkett, 5 m²*”. For mange prosjekter innebærer dette omfattende kalkyler, og det er ikke uvanlig med flere hundre slike arbeidsoperasjoner i én og samme sak.

Kalkulatørene våre må i dag bygge opp kalkylene fra bunnen av, linje for linje. Da gjelder det å ha tunga rett i munnen for at kalkylen skal bli helt riktig.

Målet for oppgaven er å lage en algoritme som predikerer manglende arbeidsoperasjoner i et rom. Dere trenger ikke bry dere om spesifikk *mengde* (for eksempel at det skal være 2.5 meter gulvlist), bare det faktum at arbeidsoperasjonen skal være med i kalkyla. Dersom hver korrekte prediksjon er en *True positive (TP)*, hver feilaktige prediksjon er en *False positive (FP)*, og hver manglende operasjon dere ikke finner er en *False negative (FN)*, beregnes scoren deres slik:

1. Hver TP gir + 1 p
2. Hver FP gir - 0.25 p
3. Hver FN gir - 0.5 p
4. Dersom et rom ikke mangler noen arbeidsoperasjoner, belønnes dere med + 1 p for å predikere en tom liste

Scoren normaliseres slik at 100 % tilsvarer en perfekt prediksjon av hele datasettet, og 0 % tilsvarer prediksjonen av kun tomme lister. Det er dermed svært mulig å få en negativ poengsum, dersom algoritmen presterer dårligere enn en *dummy algoritme* som kun predikerer tomme lister.

Scoren [0 - 100 %] (cappes på 0 dersom dere har negativ score) utgjør 40 % av den totale scoren deres i hackathonet. For mer informasjon om data, sjekk avsnitt 3.

2.2 Tilnærming og presentasjon

Det er helt naturlig at resultatene ikke nødvendigvis blir som forventet. Særlig med begrenset tid og komplekst datasett. Det viktigste er å vise hvordan dere har tenkt, hvordan dere har angrepet problemet og hva dere har lært underveis.

Problem 2: Tilnærming (30 %)

Forbered en presentasjon for lørdag ettermiddag hvor dere setter av ca. 5 minutter til å gå gjennom hvordan dere jobbet med Problem 1: Hvilken tilnærming valgte dere, og hvorfor? Gå gjerne gjennom de viktigste stegene—enten det gjelder idéutvikling, databehandling, modellvalg eller testing. Ta gjerne en titt på avsnitt 3.4 om komplementær data. Kunne tilgang på mer data hjulpet modellen deres?

Selv om dere ikke fikk resultatene dere ønsket, gis inntil 30 % av den totale scoren basert på kreativitet, fremgangsmåte, og presentasjon.

2.3 Bærekraftsinitiativ

Recovers motto er “Beskyttere av hverdagen for en bærekraftig verden”. For oss betyr ikke bærekraft bare CO₂ og klima, men også hvordan vi tar vare på folk; det er miljøet på arbeidsplassen, det er fornøyde kunder og det er effektive prosesser.

Små, praktiske grep kan ha stor effekt når man gjennomfører 30 000 prosjekter i året. I Oslo og Bergen har vi for eksempel tømre som bruker sykkel i stedet for bil. I tillegg har vi begynt å ta vare på gammel parkett, skjære den om, og gjenbruke den på nye prosjekter. Der andre kanskje ville fjernet hele stuegulvet, forsøker vi alltid å bare bytte det som faktisk er skadet. Det krever litt mer finesse, men resultatet er mindre avfall, mindre inngripen, det reduserer gjerne belastningen for kunden.

Problem 3: Bærekraftsinitiativ (30 %)

Recover jobber hver dag for å gjøre arbeidet med skadebegrensning og gjenoppbygging mer bærekraftig—både i form av lavere utslipp, redusert materialbruk og smartere, mer skånsomme løsninger for både kunder og ansatte.

Hvordan kan vi ta dette arbeidet enda et steg videre?

Gruppen skal komme med forslag til ett eller flere konkrete initiativer som kan

fremme bærekraft og/eller effektivitet i Recover. Dette kan være teknologiske løsninger, endringer i prosesser, ny bruk av data, bedre gjenbruk av materialer—eller noe helt annet. Det er også lov å utfordre etablerte rutiner.

Lørdag morgen inviterer vi til frokostseminar hvor vi i Recover forteller litt om hverdagen vår og våre utfordringer. Vi anbefaler at dere tar turen for å forberede dere best mulig til oppgaven. Representanter fra selskapet blir tilgjengelige etter seminaret for spørsmål og sparring.

Forbered en presentasjon på ca. 5 minutter til lørdag kveld.

Kreativitet, gjennomførbarhet og forståelse av bransjen vektlegges i vurderingen.

Presentasjonen utgjør 30 % av den totale scoren deres i hackathonet.

3 Data

En kalkyle kan anses som en samling av flere mindre kalkyler. Kalkylen for et prosjekt deles inn i rom (*gang, stue, kjøkken, ...*) som hver inneholder et sett arbeidsoperasjoner (*Riv taklist, Ny taklist, ...*).

3.1 Arbeidsoperasjoner

En arbeidsoperasjon er i de fleste tilfeller en standardoperasjon, men kalkulatøren kan i tillegg legge til fritekstoperasjoner. Dette er arbeid som ikke finnes i biblioteket. I denne sammenhengen er vi kun interesserte i standardoperasjonene, altså kan vi se bort fra fritekstoperasjoner i denne oppgaven. Fritekstoperasjonen er fjernet fra datasettene dere får utdelt.

3.2 Klynger (Clusters)

Flere av standardoperasjonene har mange fellestrekk. *Ny laminatgulv, Ny parkett, Ny heltregulv* kan alle grupperes under *Ny gulv*. For redusere antallet standardoperasjoner har vi laget klynger for mange av standardoperasjonene, og det er disse som er tilgjengelige i datasettet. Vi har også valgt å kode klyngene som tall. Kalkylen [*Riv flytende gulv, Riv gulvlist, Ny gulv, Ny gulvlist*] blir dermed representert som [44, 49, 46, 70] i datasettet.

3.3 Kalkyle

Her vises strukturen for hvordan en kalkyle er oppbygd:

- Soverom
 - Riv gulvlist
 - Riv gulv
 - Riv isolasjon i gulv
 - Ny isolasjon i gulv
 - Nytt gulv
 - Ny gulvlist
 - Riv veggpanel
 - Riv veggisolasjon
 - Ny isolasjon i vegg
 - Ny veggpanel
 - Tilskjæring av isolasjon
- Kjøkken
 - ...

Eksempelet viser hvordan operasjoner ofte henger sammen. Når man river gulv må man nesten helt sikkert legge nytt gulv.² Samtidig er ikke alle sammenhenger like enkle.

Et kunstig eksempel: Hvis kalkylen inneholder *Riv/Nytt gulv* skal kalkylen i de fleste tilfeller også inneholde *Riv/Ny gulvlist*. Det finnes også mer komplekse sammenhenger enn dette.

Datasettet kan lastes som tabulær data eller som tensorer. I trenings- og valideringsdata vil dere finne kolonnen *is_hidden* (dette representeres som X og Y dersom du bruker tensor-versjonen av datasettet). Denne kolonnen simulerer rader som mangler i kalkylen, det er altså disse radene dere skal forsøke å predikere ved hjelp av den øvrige kalkylen og eventuelt annen tilgjengelig data.

Tips: Dersom dere lager features - husk å filtrere ut radene hvor *is_hidden* = 1 først, slik at dere unngår å lekke informasjon dere "egentlig ikke har" inn i featurene. Testdatasettet vil naturligvis ikke ha kolonnen *is_hidden* tilgjengelig.

project_id	room	room_cluster	work_operation	is_hidden
1	Soverom 2	soverom	49	0
1	Soverom 2	soverom	44	1
1	Soverom 2	soverom	86	0
1	Soverom 2	soverom	279	0
1	Soverom 2	soverom	46	0
1	Soverom 2	soverom	70	0
1	Soverom 2	soverom	345	0
1	Soverom 2	soverom	314	0
1	Soverom 2	soverom	315	0
1	Soverom 2	soverom	330	1
1	Soverom 2	soverom	136	1
...

Det vil bli utlevert trenings- og valideringsdata. Datasettene inneholder alle arbeidsoperasjoner som er kalkulert på et stort utvalg historiske saker. Kalkylene er strengt tatt ikke komplette (det kan mangle arbeidsoperasjoner i grunndataen), men vi antar at dette jevner seg ut ettersom vi har et ganske stort datasett.

For øvrig er datasettet nærmere beskrevet i kodebasen dere får utlevert.

²I forsikringssaker står forsikringstaker (kunden) fritt til å velge *kontantoppgjør* i stedet for å få skaden fikset av oss. Det vil si at han eller hun får en sum utbetalt fra forsikringsselskapet sitt og tar ansvaret for jobben selv. Det er også mulig å velge kontantoppgjør på *deler* av jobben; for eksempel kan kunden velge å bruke Recover til å rive og tørke, og fikse gjenoppbygging selv (eller hyre inn en annen håndverker). I slike saker kan vi dermed ha kalkyle hvor vi river gulv uten å legge nytt.

3.4 Komplementær data

En kalkulator bruker en del data fra forsikringssaken for å bygge opp og revidere kalkylen. Dette er i hovedsak en befaringsrapport, som inneholder en beskrivelse av hva som har skjedd og hvilke konsekvenser det har for forsikringstaker. Denne rapporten er supplert med bilder fra skadestedet.

Underveis i arbeidet med utbedringen av skaden dokumenteres timene som håndverker og prosjektleder jobber på skaden, med en kommentar til hva som er utført de timene. Av og til er beskrivelsene veldig informative, men ikke alltid:

CaseID	Date	EmployeeID	Quantity	Description
1	25.06.2025	10	8.5	innsetting dørkarm, dør og karm lister + gulv lister på bad. panel på tak i kjeller
1	26.06.2025	10	3	veggpanel
1	26.06.2025	11	5.5	hentning av materialer. flytting av skaper og div. tildekking gulv. isolering, spikerslag og gipsing. rydding.
2	27.06.2025	12	7.5	tatt ned kjøkkenet og satt oppatt.
1	27.06.2025	10	7.5	div kontor arbeid
...

Et prosjekt vil også generere innkommende fakturaer. Dette er for eksempel faktura på materialer (*parkett, gips, lister...*). Her fremkommer løpemeter, kvadrat og antall.

Kort oppsummert - det finnes mange faktorer (ikke bare kalkylen i seg selv) som kan hjelpe til med å finne manglende arbeidsoperasjoner. Noen av disse faktorene er vanskelige å anonymisere; det kan for eksempel ligge sensitiv informasjon i rapporten, i timelistene, i bilder, eller i fritekstoperasjonene i kalkylen. Det er nok også formålstjenlig for hackathonet å begrense omfanget av oppgaven til en viss grad.

Følgende metadata er tilgjengelig på project_id-nivå under hackathonet

- damage_address_zip_code
- recover_office_zip_code
- case_creation_year
- case_creation_month
- insurance_company³
- office_distance⁴

³Forsikringsselskapets navn er anonymisert

⁴Avstand i kilometer fra Recovers kontor til skadestedet

Eksempler på data som *ikke* er tilgjengelig (men som kanskje kunne blitt brukt til å forbedre modellen deres?)

- Befaringsrapport - Strukturert tekstdata
- Bilder fra skadestedet
- Kommentarer til utført arbeid - Tekstdata
- Fakturaer med størrelser
- Fritekstoperasjoner
- AnsattID til kalkulator
- Mengde og enhet til arbeidsoperasjonene (f.eks. Riv gulv 25 m², Riv gulvlist 2.0 m)
- Bruke arbeidsoperasjonenes egentlige navn, og ikke clustre dem
- Timelister til håndverkere og prosjektleder
- Kanskje noe annet dere savner?