

# MLOps på Google Cloud Platform

## Sluttrapport

Ingvild Andersen, Jon Akselberg Langholm, Erik Flæsen Dalen

19. mai 2021



**Statens vegvesen**

## Revisjonshistorie

Dato	Versjon	Beskrivelse	Forfatter
18.05.2021	1.0	Første versjon av sluttrapport	Erik F.D., Ingvild A. og Jon L.
20.05.2021	1.1	Språkvask, mindre revideringer etter tilbakemelding fra veileder	Erik F.D., Ingvild A. og Jon L.

## Innhold

<b>1 Forord</b>	<b>5</b>
<b>2 Innledning</b>	<b>6</b>
2.1 Dokumentets hensikt . . . . .	6
2.2 Definisjoner og forkortelser . . . . .	6
2.3 Oversikt over dokumentet . . . . .	7
<b>3 Oppgavebeskrivelse</b>	<b>7</b>
<b>4 Hvordan ble oppgaven løst</b>	<b>8</b>
4.1 Litteratur og Internett . . . . .	8
4.2 Metoder og standarder . . . . .	8
4.3 Maskinvare . . . . .	9
4.4 Programvare . . . . .	9
4.5 Arbeidsfordeling . . . . .	10
4.6 Dokumentasjon . . . . .	11
<b>5 Gjennomføring av prosjektet</b>	<b>12</b>
5.1 Prosess . . . . .	12
5.1.1 Oppstart . . . . .	12
5.1.2 Utførelse . . . . .	14
5.1.3 Ferdigstilling . . . . .	17
5.1.4 Planlagt og faktisk arbeidsprosess . . . . .	17
5.2 Måloppnåelse . . . . .	18
5.2.1 Effektmål . . . . .	19
5.2.2 Resultatmål . . . . .	19
5.2.3 Prosessmål . . . . .	20
5.3 Timeregnskap . . . . .	21
5.4 Refleksjon rundt prosessen . . . . .	21
5.4.1 Suksessfaktorer og risikovurdering . . . . .	23
5.5 Endringer . . . . .	24
<b>6 Videre arbeid</b>	<b>24</b>
<b>Referanser</b>	<b>25</b>

## Figurer

2	Enkel oversikt over oppgaven. . . . .	8
3	Struktur i zip-filen oppgaven leveres som . . . . .	12
4	Stipulert arbeidsprosess . . . . .	18
5	Faktisk arbeidsprosess . . . . .	18

## 1 Forord

Denne sluttrapporten skrives i sammenheng med bacheloroppgaven «MLOps på Google Cloud Platform», forfattet av Erik F. Dalen, Jon Langholm og Ingvild Andersen. Den er den fjerde av fire rapporter produsert i forbindelse med faget IDRI3001 *Bacheloroppgave i Informatikk, drift av datasystemer*". Hensikten er å presentere leser for hvordan prosjektperioden har utfoldet seg, samt erfaringer prosjektgruppen sitter igjen med.

Oppgavens hensikt var at prosjektgruppen skulle lære hvordan man gjennomfører et prosjekt på vegne av en reell arbeidsgiver, samt de tilpasningene som behøves for å møte bedriftens mål og krav. Prosjektgruppen skulle kombinere kunnskap fra studiet med ny kunnskap for å løse bedriftens problemstilling. Fra prosjektgruppens side var det også et ønske å øke medlemmenes kompetanse i maskinlæring og maskinlæringsverktøy i Google Cloud Platform. Oppgaven har vært faglig krevende, og gruppen har måttet tilegne seg mye ny kunnskap for å kunne møte bedriftens krav. Dette innebar å sette seg inn i hvordan maskinlæringsmodeller fungerer, samt hvilke produksjonsbehov disse har for å kunne fungere optimalt i skyplattformen. I tillegg til økt kompetanse i bruk av Google Cloud Platform og maskinlæringsverktøy, endte gruppen også opp med ny erfaring i bruk av TensorFlow Extended, Kubeflow og Kubernetes.

Både oppgaven og resultatet hadde ikke blitt til uten hjelp fra Statens vegvesen, konsulenter hos Bekk og veileder hos NTNU. Vi ønsker derfor å rette en takk til alle de involverte.

### **Torgeir Thoresen (BEKK)**

Vår veileder på vegne av Statens vegvesen. Takk for tett oppfølging, veiledning, tilbakemelding på rapporter og oppklaring rundt bedriftens krav og ønsker. Du har hatt en stor positiv påvirkning i prosessen.

### **Safurudin Mahic (BEKK)**

Sensor på vegne av Statens vegvesen. Takk for all veiledning, og alle møter underveis i prosessen. Vi setter også stor pris på at du tar deg tid til å sensurere oppgaven på vegne av oppdragsgiver.

### **Jostein Lund (NTNU)**

Veileder og vår universitetslektor hos NTNU. Takk for god oppfølging og veiledning, både gjennom dette bachelorprosjektet og de siste tre årene med skolegang.

### **Stein Meisingseth (NTNU)**

Vår universitetslektor hos NTNU. Oppgaven hadde ikke blitt til uten din hjelp. Vi vil også takke for god veiledning og støtte gjennom de tre siste årene med skolegang.

### **Lars Meisingseth (SVV)**

Produkteier på vegne av Statens vegvesen. Takk for at du ga oss sjansen til å skrive hos dere. Vi har også satt pris på veiledningen vi fikk i startfasen, og at du har fulgt opp underveis.

### **Øvrig**

Det har vært flere kloke hoder som har vært innom prosessen og som har kommet med gode råd og veiledning. Takk til alle som har vist interesse i Statens vegvesen, og maskinlæringsgruppa hos Bekk som har bidratt når vi har stått fast.

## **2 Innledning**

### **2.1 Dokumentets hensikt**

Dokumentets hensikt er å konkludere og reflektere rundt bachelorprosessen i sin helhet. Det skal gå inn på hvordan oppgaven ble løst, og hvilke verktøy som ble tatt i bruk. Den informerer leser om hvordan prosjektgruppen har fordelt arbeidet, og hvordan prosjektet ble gjennomført. Dette innebærer informasjon om tidsbruk og refleksjon rundt hvordan prosessen ble lagt opp. Videre sammenstilles målene som ble satt i forstudien mot reell måloppnåelse, samt endringer som har blitt gjort i ettertid.

### **2.2 Definisjoner og forkortelser**

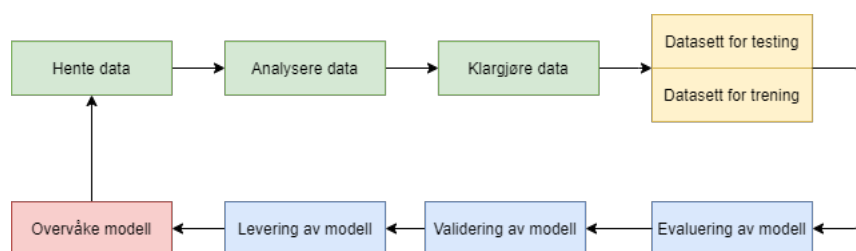
Begreper og forkortelser er samlet inn i en felles prosjektordbok.<sup>11</sup>

## 2.3 Oversikt over dokumentet

Første del av dokumentet (2) beskriver dokumentets hensikt (2.1), hvor man finner definisjoner og forkortelser (2.2) og gir en overordnet oversikt i dokumentstrukturen (2.3). Neste kapittel gir en beskrivelse av oppgaven (3), før leseren presenteres med hvordan denne har blitt løst (4). Helt konkret innebærer dette bruk av litteratur og internettsøk (4.1), metoder og standarder (4.2), maskinvare og programvare (4.3, 4.4), arbeidsfordeling (4.5) og dokumentasjon (4.6). Her beskrives altså verktøy og metoder som ble brukt for å komme frem til dagens løsning. Videre beskrives selve prosjektgjennomføringen (5), altså gjennomføringen av prosjektet. Her gjennomgås prosjektprosessen (5.1), man sammenstiller ønskede mål og oppnådde mål i måloppnåelse (5.2), ser på forventet timebruk mot reelt timebruk (5.3), før refleksjon rundt prosess (5.4) og til slutt endringer som avviker fra opprinnelig forstudie (5.5). Siste kapittel omfatter videre arbeid (6).

## 3 Oppgavebeskrivelse

Statens vegvesen ønsket et 'Proof of Concept' i form av en produksjonspipeline for maskinlæringsmodeller i GCP. Denne skulle kunne ta i mot en ferdig definert maskinlæringsmodell, og behandle denne slik at man på andre siden kunne hente ut prediksjoner gjennom API-kall. Målene definert i samarbeid med oppdragsgiver var at pipelinen skulle kunne ta i mot data, analysere og forberede data, samt dele data opp i trenings- og testsett som skulle brukes til å trene og teste modellens prediksjoner. Øvrige krav gikk ut på at man skulle kunne evaluere modellens prediksjonkvalitet, validere at den kan tas i bruk, levere modellen til et endepunkt og overvåke den når den er i bruk. Disse stadiene gjenspeiler løst prinsippene i MLOps. Prinsippene i MLOps er gjennomgått i designrapporten.<sup>9</sup> Oppgaven ble etter ønske fra arbeidsgiver løst med bruk av skyplattformen Google Cloud Platform.



Figur 2: Enkel oversikt over oppgaven.

## 4 Hvordan ble oppgaven løst

### 4.1 Litteratur og Internett

Informasjon som ble brukt til å løse oppgaven kommer først og fremst fra offisiell dokumentasjon på Internett. Helt konkret fra dokumentasjonssidene til GCP,<sup>3</sup> Tensorflow,<sup>4</sup> Kubeflow<sup>5</sup> og Kubernetes.<sup>6</sup> Det har også blitt utført nettsøk for øvrig informasjon i startfasen av prosjektet, spesielt da gruppen måtte lære seg opp i hva maskinlæring er og hvordan dette fungerer. Her ble blant annet nettsiden <https://medium.com> tatt flittig i bruk, da denne er kilde for mange artikler om emnet. I et konkret tilfelle valgte også gruppen å ta i bruk nettsiden Stack Overflow (<https://stackoverflow.com>) for å innhente hjelp, men dette førte ikke med seg noen løsning. Fysiske kilder ble ikke tatt i bruk.

### 4.2 Metoder og standarder

Prosjektdeltakerne har benyttet innslag fra utviklingsmetodene Scrum og Kanban i prosjektgjennomføringen, da især stand-ups hentet fra Scrum, samt Kanbantavler for oversikt over forefallende arbeid, hentet fra Kanban.

En stand-up er et kjapt møte, hvor prosjektdeltakere melder om hva de har arbeidet med siden sist og om det er noen utfordringer de ser med videre arbeid. Stand-upene ble gjennomført for å bidra til problemløsning, løse opp eventuelle blokkeringer som gjorde at folk stod fast, og for å oppdatere hverandre på fremdrift. Her har også veileder fra SVV, Torgeir Thoresen vært med og bidratt stort til problemløsning. Stand-ups har blitt gjennomført to ganger i uken gjennom prosjektperioden, henholdsvis tirsdager og torsdager,



med unntak av få dager, hvor stand-up har utgått grunnet ulike årsaker, som at prosjektgruppen har fokusert på andre emner i perioder.

Videre har prosjektgruppen, som nevnt, benyttet Kanbantavle for oversikt over forefallende arbeid. Til dette har verktøyet Trello (<https://trello.com/>) blitt brukt. Ved bruk av denne metoden har prosjektdeltakerne hatt god oversikt over hva som må gjøres i de ulike rapportene og utviklingsstegene, samt hvem som til enhver tid gjør hva.

Standarder har vært lite relevant for prosjektgruppen, men det har likevel blitt tatt utgangspunkt i etablerte maskinlæringsoperasjoner (MLOps) for å elevere prosessen fra en maskinlæringsmodell er utviklet til den ligger i produksjon. Dette innebærer standardprosedyrer en maskinlæringsmodell må gjennom før den er klar for å brukes. I dette prosjektets tilfelle har det vært aktuelt å trekke inn uthenting av data, analysering av data, forbereding av data, trening av modellen, evaluering av modellen, validering av modellen, levering av modellen og overvåkning av modellen. Grundigere beskrivelse av prosjektets bruk av MLOps kan leses i driftsrapporten.<sup>10</sup>

Gruppen har også forholdt seg til TensorFlow, som er de facto standard for maskinlæringsmodeller; ihvertfall for maskinlæring som baserer seg på nevrale nettverk.

I tillegg til dette har prosjektgruppen fulgt standarden IEEE som referansestil.

### 4.3 Maskinvare

Det ble ikke brukt noen maskinvare direkte, da løsningen ble utviklet og kjørt på Google Cloud Platform (GCP). Det var et ønske fra oppdragsgiver at dette skulle brukes fremfor frittstående maskiner. Metoden ved å utvikle en pipeline på en skyplattform fremfor lokale maskiner forenklet arbeidet på flere måter; bachelorgruppen slapp å forholde seg til fysiske maskiner direkte, ei heller installere tjenester på et underliggende operativsystem.

### 4.4 Programvare

Følgende programvare ble brukt under arbeidet med oppgaven:

Programvare	Brukes til
Google Cloud Platform	Skytjeneste som all underliggende programvare kjører på.
TensorFlow Extended (TFX)	Rammeverk for å definere pipeline.
Kubernetes	Plattform for å kjøre og skalere applikasjoner. Brukes for Kubeflow Pipelines som kjører pipelineen.
Kubeflow Pipelines (KFP)	Kjører selve pipelineen. Tar imot en kompilert pipeline fra TFX, og sørger for at komponentene kjøres i riktig rekkefølge.
Jupyter Notebooks	Brukes som utviklingsmiljø for pipelineen. Her kan TFX-kode utvikles, kompileres og sendes til KFP. Bachelorgruppen har brukt dette under utvikling, men dette er ikke nødvendig for videre utvikling.
AI Platform	Plattform for tjenester rund kunstig intelligens (KI) på GCP. AI Platform: Jobs brukes for trening av maskinlæringsmodeller, og pipelineen sender ferdige modeller til AI Platform: Models.

## 4.5 Arbeidsfordeling

Bachelorgruppen har hatt en viss arbeidsfordeling. Jon har hatt hovedansvar for utvikling, Erik har hatt hovedansvar for infrastruktur og vært formell prosjektleder, og Ingvild har hatt hovedansvar for dokumentasjon.

Alle tre bidro likt i starten med research. Etter hvert som arbeidet med oppgaven ble mer praktisk, falt arbeidsfordelingen naturlig ut i fra hvilke områder de forskjellige gruppemedlemmene har mest erfaring med.

Forutenom research, har Ingvild satt opp strukturen i alle rapporter, og skrevet nesten alt innhold som ikke er rent teknisk. Dette gjelder for designrapporten<sup>9</sup> og driftsrapporten.<sup>10</sup> Forstudierapporten<sup>8</sup> og sluttrapporten har hatt jevnere

fordeling av skriving.

Innledningsvis i utviklingen av pipeline, jobbet Jon og Erik tett. Undersøkelse av hvilke tjenester som skulle brukes, og diverse feilsøking ble gjort sammen. Etter hvert som brikkene falt på plass, overtok Jon utviklingen av selve pipeline, mens Erik fokuserte på infrastruktur. Avslutningsvis i implementasjon av løsning, jobbet Jon og Erik tilnærmet uavhengig med hvert sitt. Gruppemedlemmene oppdaterte likevel hverandre underveis gjennom stand-up-møter og jevnlig kommunikasjon ellers. Dette ble gjort slik at noen andre kunne steppe inn om den ansvarlige skulle bli syk eller lignende.

## 4.6 Dokumentasjon

Prosjektprosessen har ført til følgende dokumentasjon:

- Forstudierapport
- Designrapport
- Driftsrapport
- Readme
- Sluttrapport
- Prosjekthåndbok
- Presentasjon

Rapportene leveres skjøttet sammen i ett dokument. I tillegg blir alle vedlegg levert som en zip-fil. De ordinære vedleggene ligger under mappen *vedlegg*. Mappen *rapporter* inneholder de fire rapportene som enslige pdf-filer. Strukturen i zip-filen er slik:

```
vedlegg.zip
├── vedlegg/
│   ├── presentasjon.pptx
│   ├── readme.pdf
│   ├── prosjekthåndbok.pdf
│   ├── ordbok.pdf
│   ├── kildekode.zip
│   └── rapporter/
│       ├── forstudierapport.pdf
│       ├── designrapport.pdf
│       ├── driftsrapport.pdf
│       └── sluttrapport.pdf
```

Figur 3: Struktur i zip-filen oppgaven leveres som

## 5 Gjennomføring av prosjektet

### 5.1 Prosess

Prosjektet ble delt inn i tre faser: oppstart, utførelse og ferdigstilling. Vanligvis vil utførelsesfasen være den desidert største fasen av et prosjekt, men i dette tilfellet var det mye nytt å sette seg inn i, så det ble brukt mye tid i oppstartsfasen. Under kommer tre lister med fasenes deloppgaver.

#### 5.1.1 Oppstart

Arbeidet i denne fasen gikk til:

1. Planlegging med bedrift
2. Innføring i skymiljø
3. Utarbeide forstudierapport
4. Undersøke mulige teknologier og fremgangsmåter

Planleggingen med SVV startet allerede høstsemesteret 2020. Gruppen ble foreslått to forskjellige oppgaver, men valgte den vanskeligste, selv om SVV informerte om at det var en stor oppgave.

Etter valg av oppgave, kommuniserte ikke partene før vårsemesteret når

arbeidet skulle starte. Det ble klargjort grundigere hva SVV var ute etter. Sammen med SVV kom partene frem til en rekke behov. Oppgaven ble konkret å utarbeide og/eller undersøke hvordan man kunne implementere funksjonalitet som støttet opp under disse behovene. Det ble foretatt et møte for å utarbeide brukerhistorier for pipeline.

I januar fikk bachelorgruppen en innføring i GCP av veileder Torgeir Thoresen. En oversikt over forskjellige funksjoner i GCP, samt forklaring av miljøer, tjenestebrukere og andre ting som kunne skape forvirring om man ikke var informert om deres funksjon. Samtidig ble det gjort avtaler på at det skulle avholdes møter annenhver uke for å oppdatere prosjekteier Lars Meisingseth, sensor fra SVV Safurudin Mahic og veileder fra NTNU, Jostein Lund, med status underveis. Dette ble kun gjennomført et par ganger, da det viste seg lite hensiktsmessig med disse møtene i dette prosjektet. Stand-ups ble dekkende nok for prosjektgruppen og oppdragsgiver.

Arbeidet med å skrive forstudierapport startet helt i begynnelsen av semesteret, og foregikk parallelt med de to tidligere punktene. Selve arbeidet med forstudierapporten gikk stort sett knirkefritt, da bachelorgruppen hadde god erfaring med å utarbeide slike rapporter fra tidligere i studiet. Et par kapitler, spesielt de som omhandlet mål og brukerhistorier, krevde ekstra kommunikasjon med SVV før de kunne skrives.

Den siste delen av oppstartsfasen, som omhandler undersøkelse av teknologier og fremgangsmåter var spesielt tidkrevende i dette prosjektet. Denne fasen ble brukt som utgangspunkt for resten av arbeidet i oppgaven, og måtte bli gjort på en god måte for at gruppen i det hele tatt skulle kunne klare å levere på noen mål i det hele tatt.

Det ble lest mange medium-artikler og fulgt mange Jupyter-notebooker. De førstnevnte var stort sett veldig vage, og de sistnevnte fungerte ikke. Gruppen forstod at dette var et veldig nytt felt som var lite dokumentert. Gruppen hadde blitt varslet om dette fra SVV allerede første møte.

Etter noen ukers graving, oppdaget gruppen TensorFlow Extended (TFX). Tilsynelatende kunne dette rammeverket brukes for å løse de oppgavene gruppen hadde fått. En veldig lovende Youtube-video<sup>2</sup> som ga et overblikk

over rammeverket overbeviste gruppen om at det var dette som skulle brukes. De innebygde komponentene i TFX kunne brukes uten videre konfigurasjon. Etter ønske var det også mulighet til å tilpasse de innebygde komponentene eller skrive helt nye komponenter. Dette i tillegg til å kunne bruke hvilken som helst orkestrator var de to viktigste punktene som gjorde at gruppen valgte dette rammeverket. Gruppen var også frustrerte av å tilsynelatende ikke komme noen vei med prosjektet etter å ha undersøkt i flere uker, og var veldig klare for å komme seg videre i prosjektet.

Det var tilgjengelig Jupyter-notebooker for å sette opp grunnleggende TFX-pipeliner, så gruppen valgte å ta utgangspunkt i dette for å bruke rammeverket. Dessverre fungerte ikke disse, og gruppen forstod at det krevde bedre dybdekunnskap for å skulle få til prosjektet. Jon brukte om lag to uker på å gjennomgå hver enkel TFX-komponent for å forstå hver enkelt komponents rolle i pipelinen, og hvordan disse skulle konfigureres. Mens Jon satte seg inn i TFX, testet Erik de individuelle funksjonene i GCP som ville bli brukt senere av pipelinen for å integrere pipelinen med GCP.

En prototype ble utarbeidet etter den dypere undersøkelsen av TFX. En populær orkestrator som virket lovende var Kubeflow Pipelines (KFP). En orkestrator krevdes for å kjøre pipelinen, men gruppen hadde litt vanskeligheter med å få brukt KFP helt i starten. Etter en vurdering av alternativer, fant gruppen ut at KFP sannsynligvis var det rette valget, og at det ville være verdt arbeidet med å finne ut hvordan man satt opp dette. Erik tok ansvar for dette, og til slutt hadde gruppen en teknisk sett fungerende pipeline.

En gjennomgående utfordring under hele prosjektet var at MLOps er et veldig nytt felt, som fører til at utarbeidelse av pipeliner og feilsøking av disse er dårlig dokumentert. Til dels svak eller manglende dokumentasjon på TFX, samt fraværet av diskusjoner på Stack Overflow, gjorde at gruppen måtte klare seg foruten krykker som en ellers er vant til å ha i andre utviklingsprosjekter.

### 5.1.2 Utførelse

Arbeidet i denne fasen gikk til å:

1. Utarbeide designrapport
2. Implementere tiltenkt løsning
3. Utarbeide driftsrapport

Å si at utførelsesfasen kom etter oppstartsfasen er ikke helt korrekt. Arbeidet med å skrive designrapport startet samtidig som gruppen undersøkte teknologier. Dette er naturlig, da designrapporten skal rapportere det en har undersøkt og bestemt seg for.

Selve arbeidet med designrapporten gikk fint uten betydelige utfordringer. Etter gruppen hadde landet på en løsning og valg av teknologi, handlet det bare om å få nedfelt dette i rapporten.

Neste steg var å implementere selve løsningen som var skissert i designrapporten. Dette var utfordrende, men med bakgrunn i forarbeidet som ble gjort i forrige fase, hadde gruppen gode forutsetninger for å få gjennomført oppgaven. Ved hjelp av en eksempelpipeline som kan finnes som mal i TFX, kunne gruppen utvikle en hel, fungerende pipeline. De elementer som ble brukt fra malen var konfigurasjon som var spesifikk for en bestemt modell og data. Resten av arbeidet, som gikk på å sette sammen pipelinen med denne konfigurasjonen var altså helt generell. Konfigurasjonsfilene fra malen ble brukt for å teste pipelinen. Tilsvarende konfigurasjonsfiler skulle SVV utvikle på egenhånd i ettertid, og var utenfor bacheloroppgavens rammer.

Det oppstod stadig feilmeldinger for hver nye komponent som ble lagt til pipelinen. Stort sett var det vanskelig å tyde feilmeldingene, og som tidligere nevnt var det mangel på dokumentasjon og foruminnlegg om disse problemene. Problemene ble stort sett løst gjennom prøving og feiling. Stort sett var det den spesifikke konfigurasjonen for de enkelte komponentene som skapte trøbbel. En ting gruppen ikke fikk rettet på, var mellomlagring av utdata. Dette viste seg å skyldes en bug i KFP, så mellomlagring måtte skrus av for at pipelinen skulle kjøre.

Til slutt hadde gruppen en fullt fungerende pipeline som kjørte i KFP. Neste steg var å integrere de individuelle stegene i pipelinen med funksjoner i GCP. Dette var en forholdsvis enkel sak, og gruppen var takknemlig for at de hadde

valgt TFX, da dette er en av rammeverkets sterke sider.

Etter gruppen hadde utviklet en pipeline som kjørte helt og holdent på GCP med GCP-funksjoner i hvert ledd, gikk resten av arbeidet ut på å legge til rette for videre utvikling. Koden som var utviklet gjennom en lang prosess med prøving og feiling, var som et resultat blitt meget rotete. Rydding i dette, flytting og samling av konfigurasjon og kommentering av kode preget resten av selve utviklingen av pipeline. Jon tok ansvar for å ferdigstille pipeline, mens Erik lagde skript for å automatisk sette opp infrastrukturen som krevdes av pipeline på GCP. I denne perioden vekslet Ingvild på å bistå Jon og Erik og på å skrive driftsrapport.

Bachelorgruppen hadde tidligere i prosjektet etterspurt en maskinlæringsmodell fra Statens vegvesen for å teste pipeline med denne. Maskinlæringsteamet til Statens vegvesen var ikke klar med en modell som kunne brukes til testing før i slutten av april. På dette tidspunktet var bachelorgruppen opptatt med å runde av utviklingen av pipeline og dokumentere denne, så dette var dessverre litt for sent. Maskinlæringsteamet til Statens vegvesen virket positive til pipeline som var blitt utviklet. De satt av to dager til å sette seg inn i hvordan pipeline kunne tilpasses modellen sin. Mye tid gikk til å få kjørt pipeline lokalt, fremfor på GCP, for å forenkle utviklingen. Videre oppstod det problemer med å skulle gjøre all preprosessering av data bare ved bruk av TensorFlow-tensorer.

Videre skulle driftsrapporten skrives. Ingvild satte i gang arbeidet med dette samtidig som Jon og Erik gjorde det avsluttende praktiske arbeidet. Selve arbeidet med driftsrapporten gikk nok en gang greit, om noe tidkrevende. Her var det mye teknisk dokumentasjon som skulle skrives. Samtidig som Jon og Ingvild jobbet med driftsrapporten, skrev Erik en readme<sup>12</sup> som forklarer hvordan infrastrukturen og pipeline skulle settes opp på GCP. Erik skrev også på driftsrapporten om infrastruktur.

Den største fartshumpen som bachelorgruppen støtte på kom helt på slutten, da de måtte ta en og en halv uke pause fra jobbingen med oppgaven for å lese til eksamen i IDRI2006 Organisasjon og ledelse. Etter dette hadde gruppen en drøy uke på å skrive ferdig driftsrapporten og sluttrapporten.



### 5.1.3 Ferdigstilling

Arbeidet i denne fasen gikk til å:

1. Utarbeide sluttrapport
2. Levere oppgave
3. Presentere oppgave

Arbeidet med sluttrapporten gikk greitt for seg. Her handlet det bare om i løpet av en uke få oppsummert hele prosjektet i en rapport. Etter en jevn fordeling av innhold som skulle skrives av hver enkelt i gruppen, har arbeidet med sluttrapporten vært et av de enklere stegene i prosjektet. Gruppen hadde så god tid at de tok seg råd til fri på 17. mai.

Oppgaven leveres 20. mai. Selve leveransen av oppgaven vil inneholde denne rapporten og alle andre rapporter tidligere skrevet i prosjektet. Den vil også inneholde ordbok, prosjekthåndbok, readme, kildekode og lysbildene for presentasjonen. Til slutt vil leveransen inneholde en zip av 11. mai-utgaven av GitHub-pakkebrønnen.<sup>7</sup> Alle disse tingene leveres samlet i en enkelt zip-fil.

Gruppen har avtalt tidspunkt for presentasjon av oppgaven. Dette skjer 21. mai klokken 15:00. Invitert er veileder fra NTNU, Jostein Lund, veileder fra Statens vegvesen, Torgeir Thoresen, sensor fra Statens vegvesen, Safurudin Mahic, produkteier hos Statens vegvesen, Lars Meisingseth og emneansvarlig hos NTNU, Stein Meisingseth.

Statens vegvesen har også etterspurt en ekstra presentasjon som heller vil ha fokus på prosjektets produkt og vil bli en teknisk gjennomgang for ansatte og en innføring i hvordan dette kan brukes i praksis. Denne faller den 25. mai klokken 13:00.

### 5.1.4 Planlagt og faktisk arbeidsprosess

Den planlagte arbeidsprosessen er hentet fra antatt tidsbruk per rapport, på bakgrunn av tidligere bacheloroppgaver fra instituttet. Det har vist seg vanlig å bruke 2-5 uker på forstudierapport, avhengig av oppgavens omfang. Prosjektgruppen satte derfor opp fire uker til dette, da forhåndskunnskapene rundt teknologi og løsning var relativt liten. Videre har designrapporten et

gjennomsnittlig tidsomfang på 2-3 uker. Driftsrapport er den rapporten som vanligvis opptar mest tid, 8-12 uker. Prosjektgruppen budsjetterte derfor med 10 uker til denne, for at tiden skulle brukes opp. Til slutt skal sluttrapport, vurdering av samarbeid og presentasjon være relativt lite tidkrevende. Se figur 4 for prosjektgruppens stipulerte arbeidsprosess.

Uke	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Forstudie																				
Systemkrav-/designrapport																				
Driftsdokument/driftsrapport																				
Sluttrapport																				
Vurdering av gruppesamarbeid																				
Presentasjon																				

Figur 4: Stipulert arbeidsprosess

Under følger prosjektgruppens faktiske arbeidsprosess (figur 5).

Uke	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Forstudie																				
Systemkrav-/designrapport																				
Driftsdokument/driftsrapport																				
Sluttrapport																				
Vurdering av gruppesamarbeid																				
Presentasjon																				

Figur 5: Faktisk arbeidsprosess

Av dette kan en se at det ble brukt en uke mindre enn planlagt på forstudie, tre uker mer på designrapport, to uker mindre på driftsrapport og en uke ekstra på sluttrapport, samt noe forflytninger på oppstart og slutt med rapportene. Rapportene ble utarbeidet parallelt med utforsking og testing av teknologi underveis, hvor uke 11 gikk kun til testing av valgt teknologi. For grundigere Gantt-diagram, se underkapittel ?? Gantt-diagram.

## 5.2 Måloppnåelse

Målene som ble satt for prosjektet, kan finnes igjen i kapittel 3 i forstudierapporten.<sup>8</sup>

### 5.2.1 Effektmål

Effektmålene som ble satt, er som følger:

- *Ta i bruk mer moderne teknologi i veisektoren.*

Skulle Vegvesenet ta i bruk pipelinen som bachelorgruppen har utviklet, vil dette målet være oppnådd.

- *Innhente forslag til hvordan de kan kjøre det foreslåtte systemet i produksjon.*

Dette er essensen i bacheloroppgaven. Pipelinen som bachelorgruppen har levert, er både systemet og i seg selv et eksempel på hvordan man kan kjøre dette i produksjon.

- *Et konkret produkt som kan testes ut.*

Pipelinen er et konkret produkt som kan testes ut.

- *Kunne vurdere om dette er retningen de ønsker å bevege seg i, eller om de vil utforske alternativer.*

Oppgaven er et forslag til hvordan man kan kjøre maskinlæringspipeliner i produksjon, og danner et beslutningsgrunnlag for nettopp dette målet.

### 5.2.2 Resultatmål

Resultatmålene som ble satt, er som følger:

#### Hovedmål:

- *Et ferdigstilt 'Proof of Concept' på en produksjonspipeline for en maskinlæringsmodell i GCP.*

Dette har bachelorgruppen utviklet.

- *Ha resultatet klart innen 20.05.2021*

Resultatet er klart, og levert innen 20.05.2021.

#### Delmål:

- Kunne kjøre ut en maskinlæringsmodell i GCP.
- Kunne kjøre ut nye versjoner av en maskinlæringsmodell i GCP.

- Kunne eksponere en maskinlæringsmodell for input og bruke den fra andre applikasjoner.
- Finne ut hvordan man kan overvåke en maskinlæringsmodell og dens evne til å predikere riktig.
- Finne ut hvordan man kan skalere en pipeline med hensyn på kapasitet og ytelse.

Alle disse delmålene er nådd. En grundig forklaring for hvordan og hvorfor hvert mål har blitt nådd, finnes i kapittel 6 i driftsrapporten.<sup>10</sup>

Det ble tatt forbehold om at gruppen ikke nødvendigvis ville klare å nå alle delmålene, og kunne snevre inn oppgaven om det ville være behov for dette. Derimot var ikke dette nødvendig, og bachelorgruppen har klart å levere på alle mål.

### 5.2.3 Prosessmål

Prosessmålene som ble satt, er som følger:

- *Økt kunnskap innen GCP og dens underliggende verktøy*

Gruppen har absolutt blitt bedre kjent med GCP. Ingen på gruppen hadde noe erfaring med GCP ved prosjektstart.

- *Kompetanse på hvordan man planlegger og innfører en produksjonspipeline for maskinlæringsmodeller i GCP*

Gruppen har lært hvordan man designer og innfører en produksjonspipeline for maskinlæringsmodeller kun ved bruk av tjenester på GCP.

- *Styrkede samarbeidsevner i teamarbeid med færre gruppemedlemmer*

Gruppen hadde allerede god erfaring med å samarbeide med hverandre, og fikk videreutvikle dette. Samarbeidet fungerte godt, selv om medlemmene var vant med å være litt flere på en gruppe.

- *Økt kunnskap om individuelle styrker og svakheter i gruppesammenheng*

Gruppemedlemmene har blitt bedre kjent med hverandres styrker og svakheter når det kommer til samarbeid og kompetanse.

- *Erfaring rundt kommunikasjon og samarbeid med en ekstern, reell bedrift*

Dette var en ny erfaring for gruppemedlemmene, og alle medlemmer har definitivt fått mye mer erfaring med dette enn de hadde i begynnelsen.

- *Høy faglig måloppnåelse*

Dette gjenstår å se, men gruppen er optimistisk.

### 5.3 Timeregnskap

Se Gantt-diagram, ukesrapport og timefordeling i vedlagt prosjekthåndbok.<sup>13</sup>

### 5.4 Refleksjon rundt prosessen

Det positive med oppgaven var at den ga bachelorgruppen muligheten til å tilegne seg mye ny kunnskap innen maskinlæring, MLOps og Google Cloud Platform. Listen over krav fra Statens vegvesen var fleksibel, noe som la opp til at gruppen kunne konsentrere seg om å skape en best mulig løsning fremfor å fokusere på å få med alt". Det var både spennende og utfordrende å sette seg inn i et såpass lite utprøvd felt, noe som ga både mestringsfølelse og frustrasjon om hverandre. Oppgavens omfang ble konkretisert gjennom ukentlige møter med Statens vegvesen, slik at løsningen aldri lå i fare for å havne utenfor hva oppgavestiller ønsket. Denne tette oppfølgingen var svært positiv, da det ga en trygghet for både bachelorgruppen og Statens vegvesen. Målene som ble satt i forstudien ble nådd med god margin, selv om man til tider fryktet at noe funksjonalitet måtte sløyfes. Gruppen sitter igjen med mye nyttig kunnskap om Google Cloud Platform og drifting av maskinlæringsmodeller, samt hvordan det er å arbeide med en slik oppgave tett opp mot en bedrift. Tidsrammene ble litt flytende da oppgaven behøvde mye tid på research, men fungerte godt sett i sin helhet. Dette ble løst ved at man delegerte ut arbeidsansvar, slik at man kunne jobbe i parallell. Denne metoden fungerte tilfredsstillende for hele prosjektgruppen. Samarbeid mellom prosjektgruppen, bedrift og veileder har i hovedsak skjedd over

Microsoft Teams og Slack. Førstnevnte fungerte godt innen det mer formelle samarbeidet, der bedrift og veileder hadde tilgang på dokumenter, samt en oversikt over felles møter i kalender. Slack var en mer uformell plattform mellom bachelorgruppen og bedrift, der man kunne stille spørsmål om oppgaven når man sto fast. Denne fungerte også godt, og ga gruppen muligheten til å be raskt om innspill utenfor møtetidene. Det kan også trekkes frem at oppgaven er nyttig for bedriften, noe som gir en mening rundt utførelsen.

Det er mye man i ettertid kan si at skulle ha blitt gjort annerledes. Hadde man skrevet en slik refleksjon i starten av prosessen kunne det blitt argumentert for at oppgaven var for krevende, da gruppen sto mye fast. Dette løste seg fint, noe som heller ga en erfaring om at hvis man jobber på så går det bra. I gruppens interne prosess kan man kritisere at det ble brukt mye tid på research i startfasen som viste seg å ha lite funksjon for sluttproduktet, og at man skulle ha begrenset omfanget av teknologi tidligere. Spesielt ble man sittende mye med gjennomganger fra offisiell dokumentasjon som ikke fungerte. Som nevnt under prosesskapittelet måtte gruppen til slutt bruke mye tid på å gå igjennom hver enkelt komponent, noe man kunne ha satt i gang med tidligere. Bachelorgruppen må også ta selvkritikk for ikke å bruke tid på å sette seg inn i kostnadsfunksjonene i Google Cloud Platform, da det i første del av prosessen ble høye kostnader rundt utforskningen inne i GCP. Når det kommer til dokumentasjon valgte gruppen å bruke Overleaf/LaTeX fremfor den mer tradisjonelle Microsoft Word. Da størsteparten av medlemmene var mer vant med Word i begynnelsen, tok dette noe ekstra tid i startfasen av prosjektet, selv om det fungerte godt når alle lærte teknologien.

Et negativt aspekt ved perioden som må trekkes frem er Covid-situasjonen, som førte til at gruppen ikke kunne arbeide fysisk hos Statens vegvesen som planlagt. Dette kunne påvirket oppgaven i en dårlig retning, da effektiv kommunikasjon er mye vanskeligere over nett. Veileder hos Statens vegvesen tilpasset seg godt til situasjonen, og hans tilstedeværelse to ganger i uka gjennom hele perioden førte til at konsekvensen av Covid ikke var så stor som den kunne ha blitt. Intern veileder hos NTNU var også tilgjengelig på videosamtale og e-post når dette var nødvendig. Covid skapte også problemer internt i gruppa. Da dette var en praktisk oppgave fungerte gruppa

best under fysisk samarbeid, og var derfor avhengig av å finne grupperom på NTNU. Dette fungerte dårlig i de periodene NTNU stengte campus, i tillegg til at det var vanskelig å booke grupperom i tide. I en ideell situasjon hadde gruppa hatt tilgang til et fast arbeidssted gjennom hele perioden.

#### 5.4.1 Suksessfaktorer og risikovurdering

I begynnelsen av prosjektperioden ble det satt opp et sett av suksessfaktorer. I en refleksjon rundt prosessen kan det være nyttig å se tilbake på disse. For det første ble det sett på som nødvendig med aktiv involvering og tilrettelegging fra Statens vegvesen. Som nevnt tidligere i refleksjonen stilte ekstern veileder opp over all forventning, og alt gruppen behøvde hjelp med ble tilrettelagt. Disse hyppige møtene oppfylte også suksessfaktoren som handlet om regelmessig informasjonsutveksling. De neste suksessfaktorene baserte seg på tilgjengelighet i Google Cloud Platform og at løsningen måtte basere seg på GCP-komponenter. Gruppen har ikke opplevd noen problemer med funksjonaliteten til GCP, og klarte å utvikle en løsning som kjører på skyplattformen. All funksjonalitet ble godt dokumentert, og etter mye prøving og feiling klarte gruppen å tilegne seg tilstrekkelig kunnskap til å utvikle ønsket løsning.

En annen analyse som ble gjort i starten av prosessen er risikoanalysen. De uønskede hendelsene som ble beskrevet her er også relevante å trekke frem i en diskusjon om prosessen. Som nevnt tidligere i refleksjonen var gruppen i begynnelsen bekymret for at oppgaven hadde for stort omfang, og ga denne hendelsen moderat sannsynlighet. Dette viste seg å være feil, og gruppen klarte å møte de kravene som ble satt sammen med oppgavestiller. Andre risikofaktorer som ble trukket frem var sykdom og konflikt i gruppa. Gruppa har samarbeidet godt, og har ikke vært rammet av verken Covid eller øvrig sykdom. Videre risiko omhandler problemer med selve teknologien, spesielt at produktet ikke fungerer til demo. Når arbeidet løsnet etter en lang periode med research, fungerte teknologien fint. Det var mye feilsøking, men det er noe man kan forvente i et slikt prosjekt. Gruppen har allerede holdt en demo for Statens vegvesen og intern sensor som har vært vellykket, og forventer derfor det samme på offisiell presentasjon av oppgaven.

## 5.5 Endringer

Endringer omfatter endringer i løsning og teknologi.

Arbeidet med oppgaven har vært såpass vellykket at det ikke har vært behov for endring. Å bruke ekstra lang tid på å orientere seg rundt teknologier lønnet seg, da valget av teknologi viste seg å fungere godt. Det har ikke blitt foretatt noen utskiftninger i teknologistakken.

Det ble i begynnelsen av prosjektet tatt forbehold om at det kunne være behov for å innsnevre oppgaven om det skulle være behov for dette. Gruppen har klart å nå alle målene som ble satt opprinnelig. En liten avklaring ble gjort, nemlig at pipelinen kun vil fungere med Tensorflow-modeller.

## 6 Videre arbeid

Det neste som gjenstår med pipelinen, er å sette denne i produksjon. Her må maskinlæringsteamet hos Statens vegvesen legge inn en innsats for at pipelinen skal passe til den spesifikke maskinlæringsmodellen de utvikler. Dette faller utenfor omfanget til prosjektet, men dokumentasjonen som gruppen har skrevet i driftsrapporten<sup>10</sup> vil være til hjelp med dette arbeidet.

En grundigere gjennomgang om videre arbeid og potensielle forbedringer av produktet, finnes i kapittel 5,4 i driftsrapporten.<sup>10</sup>



## Referanser

- [1] Regjeringen: *Statens vegvesen*, 2021.  
Hentet fra:  
<https://www.regjeringen.no/no/dep/sd/org/underliggende-etater/statens-vegvesen/id443412/>.  
Lastet ned: 15. januar 2021
- [2] Youtube: *TFX: Production ML pipelines with TensorFlow (TF World '19)*, 2019.  
Hentet fra:  
<https://www.youtube.com/watch?v=TA5kbFgeUlk>.  
Lastet ned: 12. mai 2021
- [3] Google: *Documentation | Google Cloud*, 2021.  
Hentet fra:  
<https://cloud.google.com/docs/>.  
Lastet ned: 20. april 2021
- [4] Tensorflow: *TFX API Reference*, 2021.  
Hentet fra:  
[https://www.tensorflow.org/tfx/api\\_overview/](https://www.tensorflow.org/tfx/api_overview/).  
Lastet ned: 20. april 2021
- [5] Kubeflow: *Kubeflow Pipelines*, 2021.  
Hentet fra:  
<https://www.kubeflow.org/docs/components/pipelines/>.  
Lastet ned: 20. april 2021
- [6] Kubernetes: *Kubernetes Documentation*, 2021.  
Hentet fra:  
<https://kubernetes.io/docs/home/>.  
Lastet ned: 20. april 2021
- [7] GitHub: *Bacheloroppgave for Statens vegvesen - Release Innlevering av bachelorarbeid*, 2021.  
Hentet fra:  
<https://github.com/efdalen/bachelor2021/releases/tag/1>.  
Lastet ned: 11. mai 2021

- [8] J. A. Langholm, I. Andersen og E. F. Dalen, "Forstudierapport", *NTNU*, Februar 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.
- [9] J. A. Langholm, I. Andersen og E. F. Dalen, "Designrapport", *NTNU*, Mars 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.
- [10] J. A. Langholm, I. Andersen og E. F. Dalen, "Driftsrapport", *NTNU*, Mai 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.
- [11] J. A. Langholm, I. Andersen og E. F. Dalen, "Ordbok", *NTNU*, Mai 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.
- [12] J. A. Langholm, I. Andersen og E. F. Dalen, "README", *NTNU*, Mai 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.
- [13] J. A. Langholm, I. Andersen og E. F. Dalen, "Prosjekthåndbok", *NTNU*, Mai 2021. [Vedlegg].  
Hentet fra:  
Vedlegg i innlevering av oppgave.