

Statusrapporter

Uke nr. 4

Andrzej:

Gjennomført	Oppdatere kontrakt, initialisere timeliste, forelesning, skrive på visjonsdokument.
Problemer	Lite å skrive på visjonsdokument, data ikke tilgjengelig enda.
Tiltak	Forlate det til senere i håp om at mer dukker opp.
Oppgaver neste uke	

Eivind:

Gjennomført	Oppdatere kontrakt. To obligatoriske forelesninger om vitenskapsteori. Skrevet på visjonsdokumentet. Planlegging
Problemer	Vi har ikke fått data enda. Veldig usikker på hva vi skal gjøre. Henger etter når det gjelder timer
Tiltak	Mase på oppgavestiller. Lete etter artikler som er relevante for oppgaven
Oppgaver neste uke	Lese på relevante artikler og se på ulike metoder for klassifisering

Trygve

Gjennomført	Oppdater kontrakt. To obligatoriske forelesninger om vitenskapsteori. Skrevet på visjonsdokumentet. Planlegging.
Problemer	Vi har ikke fått data enda. Veldig usikker på

Statusrapporter

	hva vi skal gjøre. Henger etter når det gjelder timer
Tiltak	Lete etter relevante artikler og relevant forskning.
Oppgaver neste uke	Prøve å få data. Lete etter relevante artikler og relevant forskning.

Uke nr. 5

Andrzej:

Gjennomført	Leste artikler om domenet, sette opp prosjekt, skrev på powerpoint til obligatorisk workshop. Testet med neuralt nettverk med dense lag, for å kunne se om det er anvendbart i vår oppgave. Forsøke å få det mest grunnleggende til å virke. Sette opp clustering med PCA/visualisering så dataene kan bli mer lesbare når de kommer. Prøve normalisering og dropout.
Problemer	Uenighet om hvordan midlertidige testdata skal se ut, maskinlæring tar lang tid. Fortsatt ingen tilgang til data fra Signicat.
Tiltak	Bruke selvlagde data inntil videre for å ha noe å gjøre.
Oppgaver neste uke	Forberede til workshop, sette opp hos Signicat.

Eivind:

Gjennomført	Lese på relevante artikler om svindel. Forberede powerpoint til workshop. Generere data. Laste ned Pycharm. Lese på klassifiserings strategier. Implementert egen
--------------------	---

Statusrapporter

	k-nearest neighbour metode. Sendt mail til oppgavestiller om arbeidsplass og data
Problemer	Vi har ikke fått data enda. Vi bruker genererte data. Vi har ikke fått kontorplass enda. Egen k-nearest neighbour metode har høy kompleksitet i forhold til SkLearn sin knn metode
Tiltak	Fortsette å mase om data. Ta i bruk sklearn sine metoder i stedet for. Det å implementere egen metode gir grei forståelse, men ikke lav kompleksitet
Oppgaver neste uke	Møte opp hos Signicat for å få kontorplass og adgangskort. Gjør klar lånepc. Eksperimenter med ulike sklearn metoder.

Trygve

Gjennomført	Lage program for å generere data. Teste og sammeligne algoritmer på generert data.
Problemer	Ikke fått noe data enda.
Tiltak	Lage program for å generere data.
Oppgaver neste uke	Møte opp hos Signicat for å få kontorplass og adgangskort. Satt opp tre bærbare maskiner med Ubuntu 18.10 og Python 3.6.7 med biblioteker. Eksperimenter med ulike sklearn metoder.

Uke nr. 6

Andrzej:

Statusrapporter

Gjennomført	Forberedelse og gjennomføre workshop. Fikk maskiner og plass hos Signicat, installerte nødvendig software på maskinen. Prøve forskjellige metoder for å motvirke ubalanserte data (fortsatt med neuralt nettverk), og gjøre koden jeg har jobbet med ryddig. Sette opp tester på en slik måte at de kan lett kjøres sammen og sammenlignes. Prøve å endre på metrikken nettverket bruke for å lære. Detaljert ukerapporter med informasjon fra timelisten.
Problemer	F1_score ble feil. Prøvde å få mcc til å virke som metrikk for neuralt nettverk, men ingenting ville bli godtatt. På nettet har flere hatt samme problem, men få har svart. Får ikke tilgang til det interne nettet hos Signicat. Fikk et eksempel på hvordan data vi skal bruke ser ut (hvilket format).
Tiltak	Anvendte sklearn's f1_score metode, som fikset problemet. Ga opp på mcc som metric inntil videre. Spørre og mase på folk på kontoret for å få tilgang til nettet.
Oppgaver neste uke	Mase mer om data, prøve å sette opp repeterende tester, møte med veileder.

Eivind:

Gjennomført	Forberedelse og gjennomføre workshop. Lastet ned nødvendig software på lånemaskin. Eksperimentert med ulike sklearn metoder (knn, svm, decision tree). Validere resultatene med og uten normalisering og oversampling. Ordnet møte med veileder.
Problemer	Problem med touchpad på lånepc-en. Koble seg på hovednettet.
Tiltak	Prøvd å fikse det selv og spurt om hjelp.

Statusrapporter

	Ingenting hjelper. Tar med meg mus. Prøvd å koble til nettet med vpn, men vi må vente.
Oppgaver neste uke	Veldig lite vi kan gjøre neste uke på grunn av innovasjonscamp og forelesning. Møte med veileder

Trygve

Gjennomført	Forberedelse og gjennomføre workshop. Laste ned nødvendig software på lånemaskin. Eksperimentert med ulike sklearn metoder (random forest). Validere resultatene med og uten normalisering og oversampling(bruke imblearn). Koblet til internnettet. Fått tilgang til prod data. Sammenligne nøyaktighet av metoder.
Problemer	Når oversampling ble brukt så gikk nøyaktigheten ned.
Tiltak	Ikke bruke oversampling
Oppgaver neste uke	Veldig lite vi kan gjøre neste uke på grunn av innovasjonscamp og forelesning. Møte med veileder.

Uke nr. 7

Andrzej:

Gjennomført	Møtte veileder for å snakke om prosjektets framgang og framtid. Fikk tilgang til og utforsket litt Kibana, men det viste seg ikke å være veldig nyttig. Leste også litt på den interne wikien til Signicat. Maste litt mer om data, fikk mer hjelp i denne retningen. Oversatte data til en struktur vi kan bruke i maskinlæring.
--------------------	---

Statusrapporter

Problemer	Ikke lett å få noe ut av Kibana. JSON med data var ikke kompatibel med måten Python ville ha det.
Tiltak	Fikk mer hjelp av de som kjenner til systemet bedre. Det tok lang tid da dette var vanskelig, men vi fikk noe til slutt. Skrev et shellscript for å omforme dataene litt slik at Python godtar det. Det tok en lang stund, men det virker.
Oppgaver neste uke	Se nærmere på data hvis vi får dem, ellers mase mer. Prøve å kjøre noen clusteringalgoritmer på disse nye dataene. Evt. mase om mer data for å få mer nøyaktige resultater.

Eivind:

Gjennomført	Møte med veileder. Skrevet møtereferat. Fikk tilgang til Kibana, og fått hjelp til å hente ut data. Tatt en titt på ulike cluster metoder. Lest på wikien til Signicat
Problemer	Dataene vi fikk har ingen fasit om fraud. Dataene er i json format, og vi sliter med å få pandas.json.reader til å fungere
Tiltak	Siden dataene ikke har fasit, må vi bruke unsupervised learning. Når det gjelder formatet må vi prøve å få det konvertert til csv format eller få json.reader til å fungere
Oppgaver neste uke	Teste ut dataene med cluster-algoritmer

Trygve

Gjennomført	Møte med veileder. Fikk tilgang til Kibana og hentet ut data. Testet og sammenlignet clustering metoder.
--------------------	--

Statusrapporter

Problemer	Noen problemer med NaN på datasett
Tiltak	Prøve å fikse problem med NaN.
Oppgaver neste uke	Fikse problem med NaN

Uke nr. 8

Andrzej:

Gjennomført	<p>Fikk mer data, samt data sortert etter tjeneste/metode. Var med å fikse problem med verdier som blir NaN/uendelig i dataene når de blir lest inn. Kjørte neuralt nett og clustering på disse nye dataene, for å få en oversikt over hvor bra disse virker. Forsøkte å gjøre forsiktige endringer i nettverket for å forbedre resultatet. Implementerte ICA og t-SNE for å visualisere data, slik at det ikke bare er PCA. Gjorde videre forsøk med neuralt nettverk. Laget noe omformatering for tidsstempel slik at tidspunkt kan brukes i maskinlæringen som tallverdier, heller enn strenger. Ryddet koden en del ved å addere flere metoder for å slippe mye dobbel kode. La på litt mer abstraksjon slik at visualiseringsmetoder er mer uavhengige av clusteringalgoritmen. Implementerte ISOMAP og UMAP, samt fikset på små feil og detaljer. Fikset litt på hvordan timestamp blir oppdaget, for å redusere kompleksitet. Testet om endringer ødelegger eksisterende ting. Teste effektene av normalisering og skalering på data før maskinlæring. Fant at RobustScaler ga best resultat for nbid. Prøvde å legge til lag med konvolusjoner, men det gikk dårlig. Leste litt om normalisering og convolutional layers på nettet.</p>
--------------------	---

Statusrapporter

Problemer	Konvolusjonene godtar ikke shape fra laget over. Mangler fortsatt fasit på hva som er fraud og ikke fraud. Vanskelig å se hvor vi skal videre.
Tiltak	Sparer konvolusjoner til senere, spør andre om hjelp. Mase på produksjonsjef om det som mangler, snakke med andre på kontoret om hva de synes.
Oppgaver neste uke	Forhåpentligvis prøve med konvolusjoner igjen, for å se om det gir forbedring. Få en bedre ide av hva som burde gjøres videre. Finne/lese flere artikler relatert til maskinlæring generelt.

Eivind:

Gjennomført	Vi har fått mer transaksjonsdata, og disse organisert basert på metode. Hatt en nyttig samtale med han som skrev oppgaven. Laget en association rule klasse for å se sammenhenger med verdier. Konvertert datafilene fra json til csv, slik at vi kan analysere de i weka. Testet ut clustering på andre transaksjonsmetoder. Laget en grafisk tidsplan
Problemer	Fortsatt unsupervised learning. Vi vil helst ha fraud fasit. Weka klarer ikke å lese alle linjene med data
Tiltak	Sendt melding til personen som har ansvaret for disse dataene. Weka: Luke ut de linjene med å bruke dataframe.head() og dataframe.tail()
Oppgaver neste uke	Forhåpentligvis få data med fasit, men hvis ikke fortsette med unsupervised learning

Statusrapporter

Trygve

Gjennomført	Fikset problem med NaN i datasett. Teste ut datasett i Weka. Visualisering av data og analyse. Automatisere sammenligninger av metoder. Diskutert data som kan brukes og hva vi skal gjøre med den.
Problemer	Ikke noen sannhetstabell på fraud data.
Tiltak	Prøve å få tak i mer data om fraud dataene.
Oppgaver neste uke	Teste flere metoder og mase om data.

Uke nr. 9

Andrzej:

Gjennomført	Gjorde store oppryddinger i det vi har av kode, slik at det blir mer oversiktlig. Fikset det slik at man kan kjøre dimension reduction uavhengig av både clusteringen og den grafiske framstillingen. Dette tillater oss å kjøre for eksempel PCA før clustering, slik at vi kan clustre dataene basert på det vi ser i den grafiske framstillingen. Snakket også en del med noen fra Signicat om å få den fasiten på hva som er og ikke er fraud. Fikset feil som kom av kodeendringer underveis, oppdaterte kode, ryddet i metoder, laget metode for å måle metric på clustering. Oppdaget at outlier detection er en ting, og leste noe om det og hva det er. Implementerte det via sklearn, og testet det ut på dataene vi har. Dette fokuserer på tettheten til punktene, for å bestemme hva som er "normalt". Dette virket ikke uten å ta dimension reduction først, fordi ellers blir alle punktene markert som "normale". Mer outlier detection, prøve noen metoder. Leste artikler rundt temaet, fant ut at kpca fantes også som DR, så implementerte det i tilfelle vi vil prøve med det. Var med på å fikse noe feil som oppstår underveis.
Problemer	Litt tull med ny kode som ikke virket, og mye rotete struktur i prosjektet. Vanskelig å se hvor vi skal videre.
Tiltak	Ryddet litt og feilsøke.
Oppgaver neste uke	Snakke mer om anskaffelse av data, møte veileder. (Lite her ettersom eksamen er snart)

Statusrapporter

Eivind:

Gjennomført	Denne uken har jeg stort sett testet ut mye forskjellig i clustering. Først har jeg testet ut dimensjon reduksjon før clustering. Deretter har jeg lagd funksjonalitet som kombinerer association rule med clustering for å finne kjennetegn hos clusterene. Så lagde jeg en metode i python som tar i bruk en R-fil. Denne R-filen lager valgte som predikerer cluster.
Problemer	Ingen data med fraud fasit
Tiltak	Mast på oppgavestillere, men vanskelig å få tak i disse dataene
Oppgaver neste uke	Møte med veileder. Øve til eksamen

Trygve:

Gjennomført	Prøvd ut flere clustering algoritmer. Sett på outlier detection og testet ut dette. Sett på tidligere forskning. Begynt å strukturert rapport og lære mer Latex.
Problemer	Mangel på data.
Tiltak	Mase på data.
Oppgaver neste uke	Møte med veileder. Øve til eksamen.

Uke nr. 10

Andrzej:

Gjennomført	Prøve outlier detection på flere datasett,
--------------------	--

Statusrapporter

	endre litt på innstillinger. Møte veileder, hjelpe med rapport. Lese litt på artikler. Lage til fargekodingen på isomap, ved å øke contamination i en løkke, og samle opp på hvilken verdi punktene blir "kastet ut". Bruker det videre til å fargelegge punktene fra blå til rød. Får da en fin visualisering av hvor langt "utenfor" punktene er. Skrev også noen avsnitt på hovedrapport, hovedsakelig litt teori og metode om DR, clustering og outlier detection.
Problemer	Lite tid denne uken på grunn av eksamen, og ingen klare gjøremål.
Tiltak	Mer masing som vanlig.
Oppgaver neste uke	Lese mer teori etter eksamen, og planlegge veien videre.

Eivind:

Gjennomført	Vi har ikke gjort så mye denne uken på grunn av eksamen. Vi har hatt et møte med veileder, og jeg har skrevet møtereferatet til møtet. Jeg har i tillegg sett litt på matplotlib for å gjøre ting finere. Resten av tiden har gått ut på å lese om unsupervised learning, hierarkisk clustering og skrevet på hovedrapport
Problemer	Stor usikkerhet på hva vi skal gjøre videre
Tiltak	Skrive på hovedrapporten. Prøve å få veiledning
Oppgaver neste uke	Prøve å finne mer vi kan gjøre

Trygve:

Statusrapporter

Gjennomført	Møte med veileder. Vært på foredrag med Microsoft om maskinlæring. Lest mer om unsupervised learning.
Problemer	Mye lesing til eksamen.
Tiltak	Lese mer til eksamen.
Oppgaver neste uke	Lese til eksamen.

Uke nr. 11

Andrzej:

Gjennomført	Sette opp og prøve ut neuralt nettverk med verdiene fra isolation forest, og fikset en liten feil i koden til neuralt nettverk. Skrev også noe på introduksjon, teori og annet på hovedrapport. Se over hva vi gjorde før eksamenstiden. Leste en del på teori, artikler og dokumentasjon så jeg kan skrive mer på hovedoppgaven. Minimale kodeendringer. Fikk data med IP, som vi så litt på.
Problemer	Eksamen tok opp mesteparten av uken, og noe mangel på mål.
Tiltak	Få tak i et nytt datasett.
Oppgaver neste uke	Tilpasse/Omformatere nytt datasett, slik at det kan brukes med metodene vi allerede har satt opp.

Eivind:

Gjennomført	Det ble gjort lite denne uken siden de tre første dagene gikk til eksamen. Mesteparten av tiden har gått til rapportskrivning og research siden det fortsatt er usikkerhet på
--------------------	---

Statusrapporter

	hva vi skal gjøre videre. Vi fikk vite at vi ikke kunne få data med fasit på fraud. Utenom dette har jeg lagd valgtre basert på outlier detection. Helt til slutt fikk vi mer data
Problemer	Det er ingen mulighet for å kunne bruke supervised learning. Vi kan ikke bekrefte at våre metoder fungerer uten de dataene vi er ute etter. Vi fant også ut at lokasjon hadde mye å si under outlier detection. Vi er veldig usikker på hva vi skal gjøre videre.
Tiltak	Skrive mer på hovedrapport. Få veiledning. Ta en titt på forflytning av lokasjon med samme person.
Oppgaver neste uke	Se på de nye dataene vi har fått.

Trygve:

Gjennomført	Satt opp Flask Server. Lest på sklearn pipelines. Skrevet på rapport.
Problemer	Mye jobb med eksamen tidlig i uken.
Tiltak	Gjøre eksamen.
Oppgaver neste uke	Jobbe på GUI.

Uke nr. 12

Andrzej:

Gjennomført	Utforske nye data med IP mer, kjøre med forskjellige dimensjonsreduksjoner, se på hva forårsaker divergens, kommunisere med ansatte om data og oppgave. Bruke mer forskjellige outlier detection metoder for å se hvordan det ser ut.
--------------------	---

Statusrapporter

	<p>Fikset noen småfeil i koden. Satte opp VPN til NTNUs intranett. Skrev på teori og metode, samt litt analyse/diskusjon på rapporten. I tillegg skrev jeg senere mer på konklusjon og referanser etter hvert. På slutten av uken argumenterte jeg også for noen av valgene jeg tok underveis, og beskrev problemer med prosjektet. Forsøkte å illustrere hvordan data ser ut for bare en kunde. Prøvde ut et nyere og større datasett. Møte med veileder. Forsøkte å se om PCA fortsatt gir geoip som viktigst. Det går enda saktere nå med over 115 tusen datapunkter. Leste litt nærmere om UMAP, og forsøkte å se om forskjellig metric gir bedre resultat. Argumentere også for valg av metoder på rapporten, skrev om og rettet feil. Gjorde mindre kodeendringer så det blir lettere å bruke neuralt nettverk, samt lagre noen kjøringer slik at bildene kan brukes på rapporten.</p>
Problemer	<p>Memory begynner å bli for lite på maskinen når vi får store datamengder og Python. Det tar lang tid å kjøre maskinlæring med så høy kompleksitet. Mangler også en konkret retning med prosjektet, og bestemte mål.</p>
Tiltak	<p>Forsøke å fjerne noen kolonner som gir lite ny informasjon, men skaper høy dimensjonalitet. I tillegg, snakket vi en god del med de som bestemmer hva vi holder på med i et forsøk på å komme på en mer meningsfull retning.</p>
Oppgaver neste uke	<p>Få mer konkrete oppgaver ved å mase intenst på oppgavestiller. Begrunne noen av valgene bedre med bilder.</p>

Eivind:

Gjennomført	<p>Denne uken har jeg analysert outlier detection med association rules for å finne ut hva som kjennetegner en outlier. Skrevet en del på hovedrapport, hovedsakelig teoridelen. Jeg har også brukt en del tid på kildeleting gjennom google scholar og universitetsbiblioteket til NTNU. Vi har også hatt et møte med veileder samt samtaler med oppgavestiller for å finne ut veien videre. Jeg skrev møtoreferatet fra møtet. Brukte en del</p>
--------------------	--

Statusrapporter

	av den siste dagen til å se på tkinter
Problemer	Det viser seg at den viktigste årsaken til outliers er lokasjon. Vi er veldig usikre på hva vi skal gjøre videre, og det blir mye rapportskriving bare for å noe å gjøre.
Tiltak	Det er mulig vi får data angående abuse eller tjenester som er i ferd med å feile. Vi kan deretter prøve å forutse dette.
Oppgaver neste uke	Hvis det ikke dukker opp noe fra oppgavestiller, så går tiden til å lage en GUI-løsning med tkinter og å skrive på rapport

Trygve:

Gjennomført	Skrive på rapport. Oppdatere referanseliste plugin i Latex. Lese masse artikler. Møte med Nils. Diskutere mulig nytt datasett. Lage GUI for å finne lignende transaksjoner(tkinter).
Problemer	tkinter er slem.
Tiltak	Gi Eivind mye jobb med tkinter
Oppgaver neste uke	Se på timeseries predictions.

Uke nr. 13

Andrzej:

Gjennomført	Lage visualisering for begrunnelse, hjelpe med GUI, skrev på rapport om begrunnelse. Det tar meget lang tid å kjøre, så å lage visualiseringene er det som tar mest tid. Har lite annet å gjøre en del av uken. Gjøre ferdig alle illustrasjonene til begrunnelsen av dimensjonsreduksjon og outlier detection, samt skriver mer utfyllende på avsnittene der. Begynte å lese på tidsserier med flere variable og hvordan maskinlæring virker
--------------------	---

Statusrapporter

	med dem. Lærte da grunnleggende metoder og egenskaper ved tidsserier. Laget en versjon av tkinter-GUIet med outlier detection. Fant ut at å vise matplotlib i tkinter er tregt hvis det er for mange punkter på en gang. Skapte et kommandolinje UI fordi tkinter er bare stress, og det er for tregt. Generaliserte noen metoder slik at de kan brukes i flere tilfeller. Videreutviklet UI med clustering, og gjorde hele saken mer oversiktlig og robust. Fikk tilgang til et sted hvor vi kan hente data. Hjalp litt med tkinter, og fikset noen småting som ikke virket så bra med UI. Testet at mye av det gamle virker med det nye.
Problemer	Å kjøre med mange datapunkter tar mye tid, får ikke gjøre veldig mye annet samtidig. Det virker også som at å vise matplotlib i tkinter er noe tregere.
Tiltak	Lage tekst-UI som et alternativ til tkinter, som virker mye bedre.
Oppgaver neste uke	Se gjennom og utforske alternativt datasett, finne ut hva kan gjøres der, og hvordan jeg skal hente inn data fra prometheus. Legge på noen ting til UI hvis det er aktuelt.

Eivind:

Gjennomført	Mesteparten av tiden denne uken har gått til å lage gui for våre metoder. Jeg har brukt tkinter til dette og det har skapt mye problemer, men det gikk fint etterhvert. I starten brukte jeg mye tid på sette meg inn i hvordan tkinter fungerer. Jeg har lagd en main applikasjon som tar i bruk databrowseren som Trygve har laget. Jeg har generalisert den slik at den kan gjenbrukes. Deretter har jeg lagd en enkel app til plotting av datapunktene med cluster. Siste delen av tiden gikk til layout.
Problemer	Det gikk ikke an å kombinere layoutsystemer i tkinter som grid, pack o.s.v. Det var vanskelig å slå sammen frames som fungerte individuelt til å fungere sammen i en main-applikasjon.

Statusrapporter

	Det tok litt tid å få matplotlib til å fungere i tkinter. Det virker også som at plot i tkinter går tregt
Tiltak	Jeg valgte ett layoutsysteem, grid. Prøvde meg mye fram til å forstå frames i tkinter. Når det gjelder plottingen kan jeg tilby å kjøre plot utenom tkinter.
Oppgaver neste uke	Mer tkinter, skrive på rapport. Evt se på ny data

Trygve:

Gjennomført	Lese på tidligere forskning om event prediction. Teste biblioteker for timeseries. Satt opp LSTM modell for å jobbe med et vær datasett der det skulle ta tidligere data og forutse framtidig data.
Problemer	Har ikke tilgang til ytelsesdata fra produksjon.
Tiltak	Prøve å få tilgang til produksjonsdata.
Oppgaver neste uke	Prøve å få tilgang til produksjonsdata. Ordne mer på LSTM modellen.

Uke nr. 14

Andrzej:

Gjennomført	Se nærmere på prometheus og hva som ligger der. Fikk vite hvor error er, og se noen grafer, og så hvordan jeg kan få grafer for spesifikke variabler. Fikset litt rundt shellscript som skal hente ut data derfra, og lage ett for å fjerne uinteressante deler som inneholder HTML informasjon. Hjalp litt med tkinter GUI. Jobbe videre med shellscript og prometheus for å hente ut nyttige data. Forbedret en del på det jeg satte opp før, fikk mye hjelp fra noen som jobber her. Kan nå hente ut både data for når
--------------------	---

Statusrapporter

	<p>alarmene gikk og dataene som blir brukt til å beregne dem. Valgte ut noen variabler, og satte dem sammen til et datasett. Videre laget jeg ett bedre shellscript som laster ned datasettet på en effektiv måte, som de andre kan lett bruke for å få seg data å kjøre maskinlæring på, samt satte opp innlesning og sammenslåing av data i python. Brukte også en del tid på å se på hvordan variabler oppfører seg over tid. Ble litt bedre kjent med hvordan spørringene i prometheus virker også, som lot meg utforske hva som ligger der mer. Søkte mer i Grafana etter uttrykk som gir interessante variabler for datasettet, og la til noen flere. Fant mer ut hvordan jeg kan benytte Grafana, og hvordan jeg må oversette spørringene der til å kunne bruke dem via prometheus. Tenkte ut løsninger for å omforme shellscript slik at det går å hente data for bare ett alias, samt begynte å redusere antall alias brukt i det sammensatte datasettet, slik at mer relevante deler av dataene blir igjen. undersøkte flere funksjoner i prometheus for å forstå hva som foregår i kallene bedre.</p>
Problemer	<p>Det er veldig mye i Prometheus, så det er mye jobb å finne fram nyttige ting. Ukjent med syntaks og hvordan Prometheus virker.</p>
Tiltak	<p>Benyttet Grafana for å se hva som det allerede finnes grafer for, for å få et bedre inntrykk av hva som er viktig.</p>
Oppgaver neste uke	<p>Gjøre ferdig det som er igjen av splitting med alias, eventuelt finne fram mer data som kan være nyttig. Mulig det blir mer rapport også.</p>

Eivind:

Gjennomført	<p>Forbedret utseendet til tkinter applikasjonen ved å legge til fargefunksjoner. Endret layouten til databrowseren og fikset feil i innhenting av data. Lagt til en databrowser med cluster label og en association applikasjon i cluster_app. Størrelsene på komponentene i cluster_app kan endres ved bruk av paned window.</p>
Problemer	<p>Litt problemer med tkinter treeview. Det er vanskelig å endre på layout</p>

Statusrapporter

Tiltak	Mye googling
Oppgaver neste uke	Legge tkinter på hyllen og se på de nye dataene vi har fått. Evt så kan jeg legge til LSTM i tkinter applikasjonen, men da må alt i LSTM bli omgjort til metoder

Trygve:

Gjennomført	Satte opp mer robust LSTM pre-processing og post-processing. Fått tilgang til ytelsesinformasjon fra produksjon. Kjørt LSTM med produksjonsdata fra siste måned.
Problemer	Det kan ta veldig lang tid å kjøre LSTM modellen(40 min for en kjøring).
Tiltak	Lars har bestilt en kraftigere PC.
Oppgaver neste uke	Forbedre LSTM modellen. Sette opp vaslene fra Prometheus til outputen på LSTM.

Uke nr. 15

Andrzej:

Gjennomført	Jobbet videre med shellscrip og henting av data. Kan nå ha flere sett med data for seg selv samtidig, og kan hente ut det som finnes for bare ett alias. Tok med enda flere kolonner etter mer skattejakt i Grafana. Leste også litt på tidligere arbeid innen lignende tema. Gjorde et forsøk på å gjøre koden til LSTM mer ryddig og delt i metoder. Fant også ut at prometheus ikke lar meg hente ut mer enn 11000 datapunkter per request, så begynte å pønske ut løsninger for å splitte opp datointervallene i flere datoer i shellscrip. Skrev også litt på rapporten om diskusjon. Etter mye tid med shellscrip, klarte jeg etter hvert å oppnå oppdeling av datoene. Nå deles de opp slik at det ikke vil hentes mer enn 10000 punkter på en gang, og etterpå blir resultatene fra de mindre tidsintervallene samlet og gjort om til bare en tabell. Dermed har jeg omgått begrensningen, slikt at vi kan hente ut større datasett. Testet også at LSTM kjører med et datasett som er større enn vi kunne bruke tidligere. Begynner å mistenke at implementasjonen er bygd på feil grunnlag, så må
--------------------	---

Statusrapporter

	<p>undersøke det nærmere. Forberedte også noen avsnitt til rapporten, slik at etter møte med veileder kan jeg lime det inn i det formatet oppgaveformatet vi bestemmer oss for. Forsøkte litt kjøring med LSTM ved å kjøre med ekstra mange data, og bare forutse en tidslinje på en gang. Leste også litt grunnleggende teori om det, siden jeg ikke vet hva det er. Prøvde å legge på convulusjons-LSTM, men hadde problemer med shape så det ble ikke noe av det. Var også på møte med veileder. Undersøkte nærmere hvordan vår nåværende implementasjon av LSTM virker, slik at jeg kunne forenkle litt og rette et par feil. Jeg satte det også opp slik at det er mulig å se hva loss blir på test verdiene, og at det skal være mulig å lagre modellen. Testet det også litt for å se om modellen er på rett spor.</p>
Problemer	<p>LSTM er noe uryddig og vanskelig å tolke. Får ikke hentet ut nok data på grunn av grenser i prometheus. Vanskelig å gjøre det i shell script. LSTM tar lang tid.</p>
Tiltak	<p>Forsøke å forstå og rydde LSTM. Splitte opp kall til prometheus slik at hvert enkelt kall ikke går over grensen.</p>
Oppgaver neste uke	<p>Forsøke å sette opp LSTM til å forutsi verdier lengre fram ved å bruke rekursive metoder. Skrive på hovedrapport.</p>

Eivind:

Gjennomført	<p>Denne uken har satt meg inn i å forutsi grafene til driftsdataene. Jeg har lagd <code>timeseries_predictions.py</code> med en metode for data prosessering, plot og en metode som kombinerer disse to med en regresjonsmetode. Foreløpig er det lagt til 5 ulike metoder: Random forest regresjon, linear regresjon, lasso, elastic net og support vector machines. Jeg har også brukt en del tid på å finpusse og gjøre metodene mer generelle og robuste. La også til funksjonalitet der man kan bruke flere tidligere datapunkter til å forutsi grafene. Til slutt begynte jeg smått</p>
--------------------	--

Statusrapporter

	på en tkinter applikasjon for å vise de ulike prediksjonene
Problemer	Prøvde å kopiere kode fra LSTM for å bruke dette i timeseries_predictions.py, men det var litt uoversiktlig og sklearn krevde 2d array som x input og ikke 3d
Tiltak	Lagde hele timeseries_predictions selv
Oppgaver neste uke	Fortsette med timeseries_predictions.py, timeseries_app og legge til ekstra funksjonalitet

Trygve:

Gjennomført	Optimalisert LSTM model. Skrive på rapport. Lære mer om Latex.
Problemer	LSTM programmet bruker for mye minne og krasjer maskinen.
Tiltak	Mer optimal innlasting av data og tvinge garbage collection før kjøring av LSTM model.
Oppgaver neste uke	

Uke nr. 16

Andrzej:

Gjennomført	Omformet LSTM til å gjøre rekursive prediksjoner framover i tid, framfor å bare estimere neste verdi. Dette gikk ikke like bra å forutse videre, så forsøker små variasjoner for å forbedre det. La til en loss beregning for dette også, slik at det er lettere å se om modellen gjør det bra. Skrev også litt på rapporten mellom kjøringene. Omformet LSTM videre for å dele opp i batches, samt kjøre flere men kortere tester for å få bedre metrikk på loss. Så også på resultatet av en kjøring som fortsatte over natten, og så at rekursjonene ble mye bedre etter mange nok epoker. Satte også opp LSTM slik at det er mulig å spesifisere hvor mye skal brukes til å forutse hvor mange punkter framover
--------------------	---

Statusrapporter

	<p>per batch. Utvidet nettverket litt også, slik at modellen kan bli mer komplisert for lange kjøring. Siden vi ikke skal være på kontoret de neste dagene på grunn av påsken, forsøker jeg å kjøre en ekstra lang kjøring, som vil kjøre i flere dager hvis ikke programmet krasjer eller maskinen går tom for minne av en eller annen grunn. Vi har fortsatt svake maskiner, så å kjøre maskinlæringen med mange epoker tar fortsatt veldig lang tid. Skrev på rapport om LSTM, time series data, samt utvidet introduksjon, forord og diskusjon om fraud noe. Forberedte et avsnitt om LSTM teori, samt søkte etter noen bra referanser/teori til det. Skrev om også noen deler med dårlig språk. Tok også med litt på det generelle avsnittet om teori på fraud delen. Fikk på plass en del om teori og metode til time series, samt en stor blokk med diskusjon. Skrev også litt om prosessen vår i metode til fraud delen.</p>
Problemer	<p>Ikke nok minne til å videre utvide datasettet. Vanskelig å komme seg til kontoret i påsken.</p>
Tiltak	<p>Kjøre med flere epoker i stedet for å oppnå bedre nøyaktighet. Jobbe på skolen med rapport.</p>
Oppgaver neste uke	<p>Se resultatet av lang kjøring, prøve å forbedre videre. Sannsynligvis skrive videre på rapport, spesielt LSTM teori.</p>

Eivind:

Gjennomført	<p>Startet uken med å lage møtetreferat. Deretter har jeg laget en rekursiv predikerings metode som bruker sine egne y output som input. I tillegg har jeg lagt til funksjonalitet i timeline_predictors som predikerer flere dager fremover. De siste to dagene brukte vi tiden på rapporten. Vi startet med å se over og forbedre forordet og introduksjonen. Deretter skrev jeg om connectivity based og distribution based clustering</p>
Problemer	<p>Memory leak error hvis jeg tar i bruk for mange dager bakover i timeline_predictors.</p>
Tiltak	<p>Ikke bruke for mange dager, eller prøve ut</p>

Statusrapporter

	maskinen som er bestilt for oss
Oppgaver neste uke	Fortsette med timeline_predictors eller skrive mer på rapport

Trygve:

Gjennomført	Vært på hytte uten nett og pc.
Problemer	Ikke pc eller nett.
Tiltak	Komme tilbake til sivilisasjonen.
Oppgaver neste uke	

Uke nr. 17

Andrzej:

Gjennomført	<p>Ryddet LSTM opp i bedre metoder, for at det skal kunne anvendes andre steder enklere. Så gjennom resultatet fra lang kjøring, det gikk bedre, men ikke helt så bra som ønskelig. Testet nye endringer, for å være sikker på at alt virker som det skal. Vi fikk ny og sterkere maskin. Begynte å lage en versjon av LSTM som trener på å direkte forutsi en serie med punkter, heller enn rekursivt, for å se om det blir bedre. Ordnet LSTM slik at det også går å forutsi en rekke med punkter på en gang, heller enn rekursivt. Ryddet deretter koden til begge disse, ved å samle det som er likt for begge. Brukte en del tid på å prøve å forbedre resultatet til LSTM, og til dette implementerte jeg også en graf av loss-historikken for både trening og vurdering. Dette tillater oss å se hvordan modellen utvikler seg over tid, og kan gi verdifull informasjon om mulig forbedringspotensiale. Forsøkte videre å utforske hvordan endringer i LSTM nettverket påvirker resultatet, og fant at å lage et større ett gjør ikke det nødvendigvis bedre enn et lite ett. Utforsket flere aktiveringsfunksjoner, samt ting som toveis LSTM. Bruke den nye maskinen noe for å se hvordan modellen blir etter lang tid, for å finne ut om den stopper å bli bedre, blir overfittet eller faktisk øker nøyaktigheten. Anvendte feature selection på LSTM, slik at det er mer relevante kolonner for modellen, som førte til noe lavere loss. Forsøkte videre å forbedre modellen, men det eneste som ga bedre resultat var å øke data. Separerte innlasting og skalering av data.</p>
--------------------	--

Statusrapporter

Problemer	Klarer ikke å gjøre modellen mer nøyaktig, men resultatet er fortsatt ikke helt optimalt enda. Modellen treffer av og til i rett retning (opp eller ned), men er på ingen måte pålitelig enda.
Tiltak	Forsøke både mer avanserte og forenklede modeller. Forsøke å bruke mer data, men det er ikke så mye mer å ta av.
Oppgaver neste uke	Prøve noen andre nettverkslag for å se om det hjelper, men hovedsakelig rapport. Se om høyere resolusjon (altså flere datapunkter) gir bedre resultat.

Eivind:

Gjennomført	På mandag skrev jeg litt på rapport i teoridelen om clustering. Resten av uken gikk til å forbedre timeline_predictors samt og fremstille dette i appen timeseries_app. Jeg har lagt til feature selection i timeline_predictors slik at prediksjonen går mye raskere. Måten dette gjøres på er at jeg tar i bruk valgtre for å fjerne unødvendige features. Det er også mulig å fremstille grafene i tkinter med knapper som kan skru av og på grafene. Jeg har også brukt tid på å fikse den rekursive metoden i timeline_predictors
Problemer	Hatt noe problemer med tkinter. Det var vanskelig å endre på den rekursive metoden etter endring i dataprosessering metoden
Tiltak	Jobbet meg rundt problemet i tkinter. Mye debugging angående den rekursive metoden
Oppgaver neste uke	Helst skrive mye på rapporten og dokumentere kode.

Statusrapporter

Trygve:

Gjennomført	Fått ny pc og satt opp den. Optimalisert LSTM for å kjøre på GPU(100 ganger raskere). Ordnet så alle verdier blir skalert tilbake til original skala etter trening. Testet ut Hyperas for Bayesian Optimization
Problemer	<ol style="list-style-type: none">1. PCen fryser helt på tilfeldige tidspunkt uten belastning.2. Ikke noe Python 3.6 på Ubuntu 19.04.3. Problemer med nVidia driver og CUDA installasjon.
Tiltak	<ol style="list-style-type: none">1. Oppdatert BIOS og deaktivert strømspare innstilling.2. Lastet ned kildekoden og compilet dette.3. Byttet til Ubuntu 18.04.
Oppgaver neste uke	Skrive rapport.

Uke nr. 18

Andrzej:

Gjennomført	<p>Skrev om metode på time series, og en del om metode på fraud delen. Samtidig kjørte jeg noen forsøk på å forbedre, vurdere, og vise fram modellen. Skrev mer konklusjon, diskusjon og andre små endringer mens jeg gjorde noen enkle forsøk med andre optimizere for LSTM (rmsprop for eksempel). Forsøkte å kjøre med data med resolusjon på 1 sekund, men matplotlib er for svakt. Forsøkte å skrive litt om RNN og LSTM, samt exploding/vanishing gradient. Leste litt om gradient. Omformulerte en del av det jeg skrev før. Utvidet diskusjonen på fraud delen, og la på en del på konklusjon og data delene på time series. Ryddet litt i koden, slik at det blir lettere å se hva som er relevant. Skrev også noe mer på konklusjon og diskusjon på time series. Leste om vitenskapsteori forelesninger for å fornye kunnskapen om det. Letet etter noen gode referanser, og utvidet introduksjon litt. Forsøkte å endre måten treningen skjer på, men forkastet det da jeg må bruke tid på rapport og det virket ikke. Mye lesing av relaterte artikler. Reformaterte mye tekst slik at det blir mer elegant, samt fant noen gode kilder. Skrev også litt om over- og underfitting, og forsøkte å skrive mer</p>
--------------------	---

Statusrapporter

	om skalering.
Problemer	Matplotlib er svakt.
Tiltak	Prøve noe annet.
Oppgaver neste uke	Mer rapport.

Eivind:

Gjennomført	Uken har stort sett gått til å gjøre ferdig teoridelen i rapporten. Jeg startet uken med å skrive om regresjon, lasso, ridge, elastic net. I tillegg implementerte jeg inn ridge og lars i timeline_predictors. Enkle forbedringer ble gjort i timeseries_app ved at man kan kjører en og en modell med mellomlagring og mulighet for tilbakeblikk. Deretter skrev jeg ferdig delen om density based clustering. Neste del jeg skrev ferdig var om least angle regression (Lars). Etter dette fikk vi veiledning over telefon om hvordan vi skulle sette opp rapporten. På fredag gjorde jeg ferdig teoridelen om valgtrær og random forest.
Problemer	Vi var usikre på hvordan vi skulle sette opp rapporten våres siden den er todelt. I tillegg var vi litt usikre på hvordan vi skulle sette opp kildene
Tiltak	Vi tok kontakt med faglærer i emnet om bacheloroppgaven. Hun kom med råd om å ha de to delene i samme rapport, men med god adskilling.
Oppgaver neste uke	Teoridelen er stort sett ferdig. Målet blir å gjøre ferdig metodedelen og kanskje resultatdelen.

Statusrapporter

Trygve:

Gjennomført	Skrevet rapport, ordnet referanser, sette opp Bayesian Optimization biblioteket Talos og bruke det til å optimalisere modellen.
Problemer	Mange kilder som har blitt brukt er ikke bra nok(har ikke referanser).
Tiltak	Finne bedre kilder
Oppgaver neste uke	Mer rapport

Uke nr. 19

Andrzej:

Gjennomført	Fjernet litt duplikater i rapporten, og beskrev hvilken variabel vi benytter som eksempel i rapporten. Endret også litt på datasettet for å fikse node_load variabelen. Fylte på litt mer på metode på time series, og begynte å sette opp analyse for time series. Laget og plottet en naiv metode for å forutse framtidige verdier, slik at vi har en baseline å sammenligne alle metodene med. Definerte en metrikk for å sammenligne modellene, der hvis verdien er større enn 0, er modellen bedre enn den naive, mens hvis den er lavere, er den dårligere. Laget også noen illustrasjoner for analyse på fraud, og skrev litt rundt disse på både analyse og diskusjon. Laget også et avsnitt om hvorfor en naiv metode er viktig. Laget en mer avansert naiv metode, slik at det er mer å teste mot. Skrev om forordet litt. Skrev mer på begge introduksjonene, ved å beskrive innholdet i rapportene. Gikk gjennom og noterte hvor det hadde passet med flere referanser, og foreslo endringer. Møte med veileder. Hjalp litt med å skrive på skalering, skrev mer om data og resultat på fraud, og diskusjon og naive metoder på time series. Utvidet begge konklusjonene. Endret språket på fraud problemstillingen. Gikk gjennom hele oppgaven igjen og oppgraderte språk, og flyttet på ting og omformulerte. Skrev litt om GRU, og leste på RNN.
Problemer	Vanskelig å forstå seg på RNN, Talos tar lang tid å kjøre.
Tiltak	Lese mer om RNN.
Oppgaver neste uke	Gjøre ferdig rapporten, skrive mye om

Statusrapporter

	resultatet av Talos. Dokumentere og rydde kode.
--	---

Eivind:

Gjennomført	Jeg har testa hvordan modellene i timeline_predictors reagerer på mer og mer data til modellen med expanding window. Jeg har også tatt i bruk samme dataoppdeling som LSTM hvor det deles i trenings- , validerings- og testdata. Valideringsdataen blir urørt av de enkle metodene. Deretter har jeg lagt til GRU i timeline_predictors for å sammenligne metodene. Når det gjelder rapporten har jeg skrevet på metodedelen om clustering og association rule. Jeg har også sett tilbake på fraud-delen for å finne ut hva mer vi kan skrive om, og prøve å legge til mer innhold på resultat (bilder etc.). Skrevet på delen om support vector machine i rapporten. Møte med veileder samt at jeg har skrevet møtetreferat. Satt opp tidspunkt for presentasjon
Problemer	Maskinen jeg bruker sliter med å håndtere store datamengder med fraud_data. Får memory error. Den stasjonære maskina er opptatt med timeseries
Tiltak	Jeg får vente til neste uke og deretter gjør ferdig analysedelen på fraud-delen
Oppgaver neste uke	Gjør ferdig og finpusse rapporten. Dokumentere kode

Trygve:

Statusrapporter

Gjennomført	Oppdatert illustrasjoner, ordnet ny datasplitter(expanding window og sliding window) for datasettet, skrevet rapport, ordnet kilder
Problemer	Deler av rapport mangler relevante kilder.
Tiltak	Finne kilder
Oppgaver neste uke	Skrive ferdig rapport.

Uke nr. 20

Andrzej:

Gjennomført	Skrev litt om robust scaler på teori, hjalp andre med Engelsk på rapporten. Mer rapport, og mye rydding av kode for å klargjøre for innlevering. Begynte på sammendrag, skrev litt om normalisering, og på time series diskusjon. Laget også en liste så vi ser hva som er igjen, slik at det blir lettere å se. Gikk gjennom noen av retningslinjene for bachelor for å se om vi har med alt. Sammenligningen kjørte ferdig, så beskrev resultatene fra den på rapporten. Fikk gjort mesteparten av konklusjon og analyse på time series, og utvidet sammendraget litt. Flyttet litt på og redigerte mange avsnitt. Skrev om gradient descent. Utvidet time series diskusjonen en del. Skrev readme, ryddet mer kode. Skrev om kjøretid, og var med på gjennomgang av rapporten. Refleksjonsnotat.
Problemer	Sammenligningen kjørte feil, så noen bilder mangler fortsatt. Noen ting er fortsatt usikre angående innlevering.
Tiltak	Kontakte veileder. Kjøre sammenligning på nytt, og møte opp i helgen.
Oppgaver neste uke	Ferdiggjøre innleveringen og framføring.

Eivind:

Gjennomført	Denne uken har gått til å dokumentere kode og skrive ferdig rapporten. Det første jeg gjorde var å skrive teori og metode for support
--------------------	---

Statusrapporter

	vector machine. Jeg følte det manglet litt på diskusjon på fraud delen, så jeg fylte mer der. Deretter skrev jeg på metodedelen om feature selection og sklearn metodene. Utenom rapport og dokumentering, brukte jeg en dag på å fremstille en god sammenligning mellom sklearn metodene og RNN metodene. Til slutt gikk tiden til å finpusse rapporten
Problemer	Vi var usikre på va vi skulle levere og om vi kunne levere kode gjennom github
Tiltak	Vi ringte veileder og vi kom fram til at vi kunne levere kode gjennom github. Vi fant også ut hva som skulle leveres på inspera
Oppgaver neste uke	

Trygve:

Gjennomført	Tegnet illustrasjoner for GRU, LSTM, RNN, NN. Finpuss av dokument, dokumentasjon av kode. Formatering og omskriving av rapport. Skrevet mer på teori.
Problemer	Ikke alt klart til innlevering
Tiltak	Gjøre alt klart til innlevering.
Oppgaver neste uke	Levere oppgave på mandag.