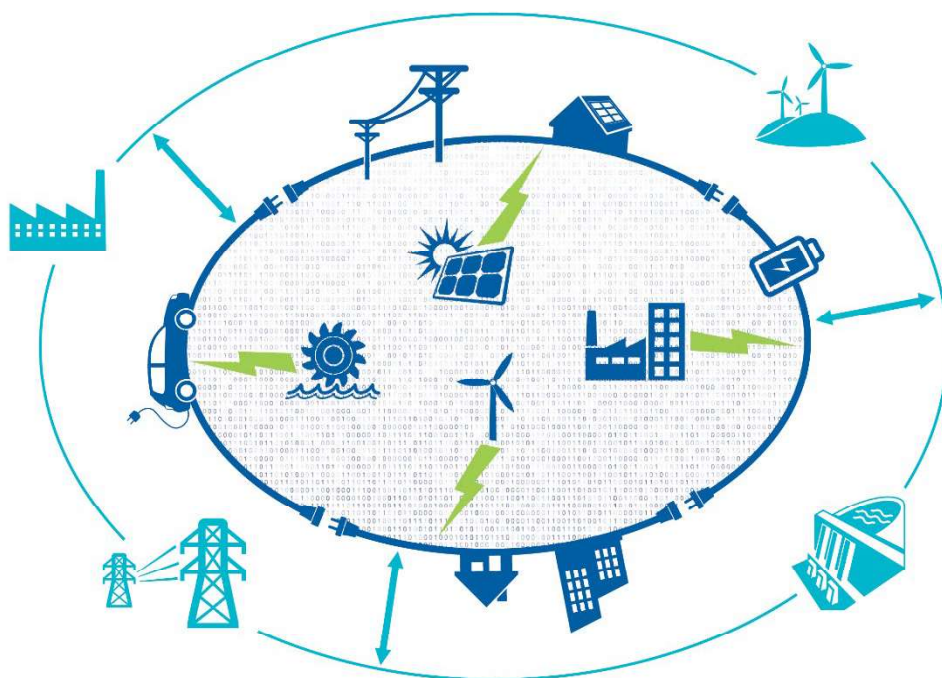


Pilot sluttrapport

Pilot Prediksjon av belastning i strømmettet

Forfatter: Per Oddvar Osland



CINeLDI

Centre for intelligent electricity distribution
- to empower the future Smart Grid



Centres for
Environment-friendly
Energy Research

CINELDI - Centre for intelligent electricity distribution


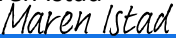
SINTEF and NTNU are the main research partners, with grid operators, technology providers, public authorities and international R&D institutes and universities as partners.

The research centre is financed by the Research Council of Norway and the Norwegian partners through the Centre for Environment-friendly Energy Research (FME) scheme. The FME scheme consists of research centres of limited duration that conduct concentrated, focused and long-term research on a high international level to solve specific challenges related to energy and the environment.



Centres for
Environment-friendly
Energy Research

Prosjektnotat

TITTEL			
Resultat og erfaringsnotat for Pilot <i>Prediksjon av belastning i strømnettet</i>			
WORK PACKAGE	VERSJON	DATO	ANTALL SIDER
WP Pilot	1.0	2024-08-06	13
FORFATTER(E)		WP-LEDER	GRADERING
Per-Oddvar Osland  <small>Per-Oddvar Osland (Oct 21, 2024 16:25 GMT+2)</small>		Maren Istad  <small>Maren Istad (Oct 21, 2024 18:32 GMT+2)</small>	Åpen

SAMMENDRAG

Det har blitt utviklet maskinlæringmodeller for predikering av belastning på transformatorer i nettstasjoner. Modellene er basert på felles teknologi, men er trent opp separat for å predikere belastning for ca 130 nettstasjoner. Nyttien består i at man er i stand til å unngå overbelastning, og dermed unngå/ redusere risiko for driftsavbrudd (og dermed avbruddskostnader), samt forlenge levetid på komponenter i nettet. Løsningen er i drift hos Glitre Nett.

Innholdsfortegnelse

1	Bakgrunnsinformasjon om pilotprosjektet.....	3
2	Om Piloten og fysisk pilotområde.....	5
3	Resultater og innovasjoner fra Piloten.....	6
3.1	Resultater fra delaktivitet 1: Kartlegge behov og problem	6
3.2	Resultater fra delaktivitet 2: Vurder ulike modeller	9
3.3	Resultater fra delaktivitet 3: Utvikle, trene og teste modeller	9
3.4	Resultater fra delaktivitet 4: Evaluering og analyse	9
3.5	Resultater fra delaktivitet 5: Sette løsning i operativ drift	9
3.6	Innovasjoner fra Piloten.....	11
4	Tekniske/faglige erfaringer fra Piloten.....	11
4.1	Oppsummering	12
5	Kost-/nyttevurderinger basert på resultatene for Piloten.....	12
5.1	Kostnader	12
5.2	Nyttevurderinger	12
6	Referanser.....	13

1 Bakgrunnsinformasjon om pilotprosjektet

Tabell 1: Bakgrunnsinformasjon

	Fra malen "planlegging av pilotprosjekt"	Viktige endringer i løpet av pilotperioden
Målsetting	<p>Utvikle metoder/modeller for prediksjon av belastning i strømmettet, med fokus på belastning på nettstasjoner.</p> <p>Vurdere "godhet" / treffsikkerhet for modeller som er utviklet.</p>	Ingen endring.
Problemstilling	<p>Belastning på vitale komponenter i strømmettet varierer som følge av periodisk forbruksmønster (døgnsyklus, ukesyklus, sesongsyklus), trender (økonomisk utvikling, elektrifisering), spesielle hendelser (ferier, helligdager, arrangementer), og variasjon som følger av den stokastiske naturen som ligger i individuell oppførsel. I sum gir dette et bidrag som det er krevende å predikere på en god måte.</p> <p>Belastning kan måles på punkt i nettet der det finnes måleutstyr. For AE Nett sin del er dette transformatorstasjoner, nettstasjoner og hos enkeltkunder. På transformatorstasjoner er forbruket summert opp av flere tusen slutt kunder, og den stokastiske andelen er dermed liten. Prediksjon er dermed lettere enn lenger "nede" i nett-topologien.</p> <p>På sluttkundenivå er den stokastiske andelen stor, og prediksjon dermed mindre treffsikkert. Samtidig er det mindre relevant å predikere forbruk for en enkelt kunde.</p> <p>Nettstasjoner er en nettressurs som vi er interessert i å unngå overlast på, både for å unngå overoppvarming / havari, og for å unngå avkorting i levetid. Dette er også et område det har vært fokusert lite på, delvis fordi måledata fra nettstasjoner ikke alltid er tilgjengelig. Vi ønsker å få mer innsikt i dette området, og vil derfor fokusere på prediksjon av belastning på nettstasjoner. Metoden som er utviklet er imidlertid anvendbar</p>	Ingen endring. Piloten er gjennomført med fokus på nettstasjoner, som planlagt i utgangspunktet.

	for andre områder også, dette må eventuelt vurderes fra tilfelle til tilfelle.	
Aktiviteter	<ol style="list-style-type: none"> 1. Kartlegge behov og problem I hvor stor grad forekommer overbelastning? Hva er typiske belastningsmønstre for en nettstasjon? 2. Vurder ulike modeller Statistiske modeller ML-baserte modeller 3. Utvikle, trene og teste modeller 4. Evaluering og analyse Hvordan er ytelsen på valgt modell i forhold til andre Hvordan er kompleksitet og kostnader sett i forhold til alternative modeller 5. Sette løsning i operativ drift 	Ingen endring. Aktiviteter er i stor grad gjennomført som planlagt.
Kostnadsestimat	<p>Timer: 400 timer, herav</p> <ul style="list-style-type: none"> • Data scientist 250 timer (utvikling av modell, analyse) • Data Engineer 100 timer (produksjonssetting av modell) • Prosjektstøtte 50 timer <p>IKT-kostnad: Kr 50 000</p> <ul style="list-style-type: none"> • Kjøretid på Azure: Trening av modeller, lagring av data, prediksjon 	Reelt timebruk har blitt noe større enn forventet, delvis pga Covid og delvis fordi det var behov for mer tid til trening og idriftsetting. Det har blitt gjennomført en masteroppgave som del av piloten. Tidsforbruk hos studenten er ikke tatt med i regnskapet.
Innovasjonspotensial		
Forventet resultat	<ul style="list-style-type: none"> - Metode for prediksjon av belastning - Løsning i operativ drift hos Glitre Nett 	
Tidsplan	1.1.2020 – 31.12.2020	1.1.2020 – 31.12.2022 Piloten ble forlenget, delvis pga Covid og delvis fordi det var behov for mer tid til trening og idriftsetting.

2 Om Piloten og fysisk pilotområde

Tabell 2: Piloten og pilotområdet

Pilotområdet	Data ble hentet fra nettstasjoner hos Glitre Nett sør, tidligere Agder Energi Nett.
Måledata og andre data som samles inn og lagres fra Piloten	<ul style="list-style-type: none"> - Tidsserie for belastning (kWh/h) over flere år for utvalgte nettstasjoner - Type montasje av nettstasjon (montasje i maste, kiosk eller innvendig i bygg) - Merkeytelse (kVA) for transformator i nettstasjon (dvs maks anbefalt belastning)
Personvern og/eller kraftsensitiv informasjon	<ul style="list-style-type: none"> - Personvern: Ingen personopplysninger i denne piloten - Kraftsensitiv informasjon: Detaljert teknisk informasjon om nettstasjoner er begrenset til belastning, merkeytelse og typemontasje. Koblingsskjema inngår ikke.
Måle- og kommunikasjonsinfrastruktur	Det ble brukt data fra AMS-målere som er montert i alle nettstasjoner hos Glitre Nett Sør. Dette var en infrastruktur som alt var etablert da piloten startet, og det var dermed ikke behov for å installere mer utstyr.
Use-case-beskrivelser og testplaner	Use-case og testplaner er vist i masteroppgaven til Adrian Langemyr [1] og delvis i presentasjoner fra konferanser og webinar [2][3][4].
Regulering og forskrifter	Ingen reguleringer og forskrifter er spesielt relevante for denne piloten.
Barrierer og løsninger	Samenstillinger og dataanalyser har vært krevende, men fremstår ikke som barrierer. Den største barrieren er i praksis overgangen fra pilotering til å ta løsninger i bruk aktivt i operativ drift. Dette er diskutert i presentasjonen «Maskinlæring – overgang fra pilot til drift» [3].
Hvem skal eventuelt ta resultater fra Piloten i bruk?	Resultatene brukes primært i nettselskap i avdeling for vedlikehold og avdeling for operativ drift. Etersom prediksjonen strekker seg rundt 2 døgn frem i tid, vil resultatene være mest relevante for operativ drift i områder der fleksibilitet kan brukes til å redusere effektpådraget.
Hvem er erfaringene relevant for?	Nettselskap, forskningsmiljø og leverandører av programvare for overvåkning og styring av nett.
Hva påvirkes av resultater fra Piloter?	Piloten er et viktig bidrag til mer dynamisk nettdrift. Det vil være mulig å etablere rutiner for overvåkning og varsling av kommende overbelastning, noe som gir grunnlag for å gjøre aktive tiltak for å unngå overbelastning av nettet. Tiltak kan for eksempel være kjøp av fleksibilitet, eller bytte til større transformator i nettstasjonen.
Informasjonsdeling mellom aktørene før/underveis/etterpå	Resultat fra piloten har blitt delt gjennom åpen masteroppgave [1] og presentasjon i flere fora: CINELD-webinar november 2020 [2],

	CINELDI-dagene november 2020 [3] og konferansen AI&ML for the Smart Grid september 2020[4]
Er det laget planer for videreføring? Skalering/fullskala implementering?	

3 Resultater og innovasjoner fra Piloten

Dette kapitlet gir en kort oversikt over resultat og innovasjoner. Merk at belastning og overlast måles i hver enkelt nettstasjonskrets. En nettstasjon vil som regel ha en nettstasjonskrets, men kan i enkelte tilfeller ha flere. For detaljer, se innhold i referanser i kap 6.

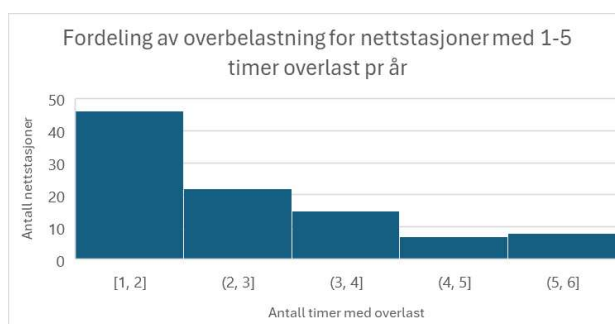
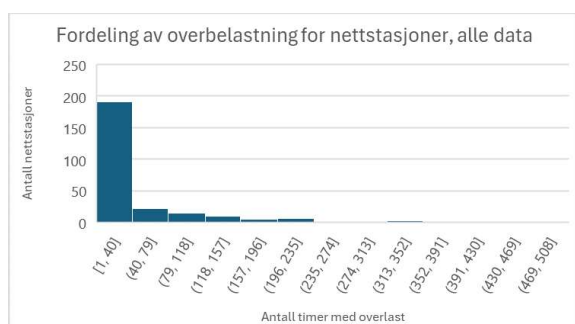
Følgende aktiviteter har blitt gjennomført:

1. Kartlegge behov og problem
I hvor stor grad forekommer overbelastning?
Hva er typiske belastningsmønstre for en nettstasjon?
2. Vurder ulike modeller
Statistiske modeller
ML-baserte modeller
3. Utvikle, trene og teste modeller
4. Evaluering og analyse
Hvordan er ytelsen på valgt modell i forhold til andre
Hvordan er kompleksitet og kostnader sett i forhold til alternative modeller
5. Sette løsning i operativ drift

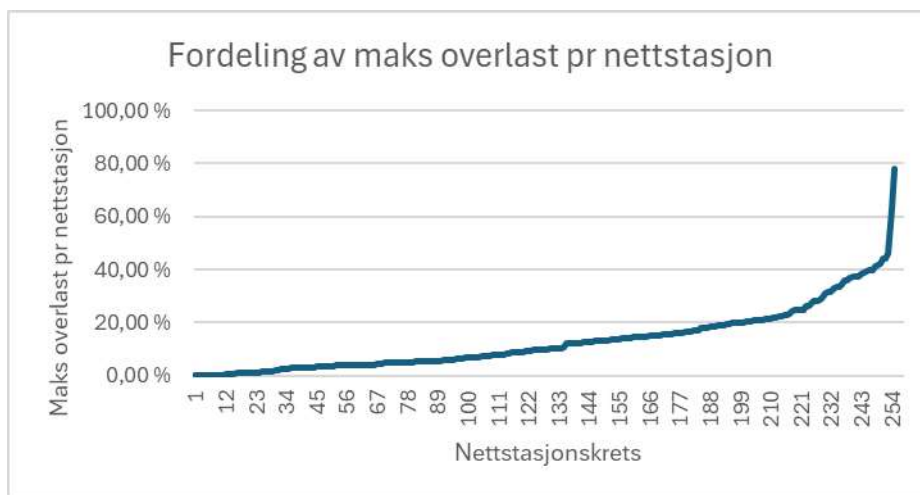
3.1 Resultater fra delaktivitet 1: Kartlegge behov og problem

Hos Glitre Nett Sør har det vært overbelastning det siste året på rundt 255 av ca 8900 nettstasjonskretser, dvs 2,9% av nettstasjonene opplever overlast. Disse tallene vil naturlig nok varierer fra år til år. Enkelte år har svært kalde vintre, og overbelastningen vil da skje i større grad.

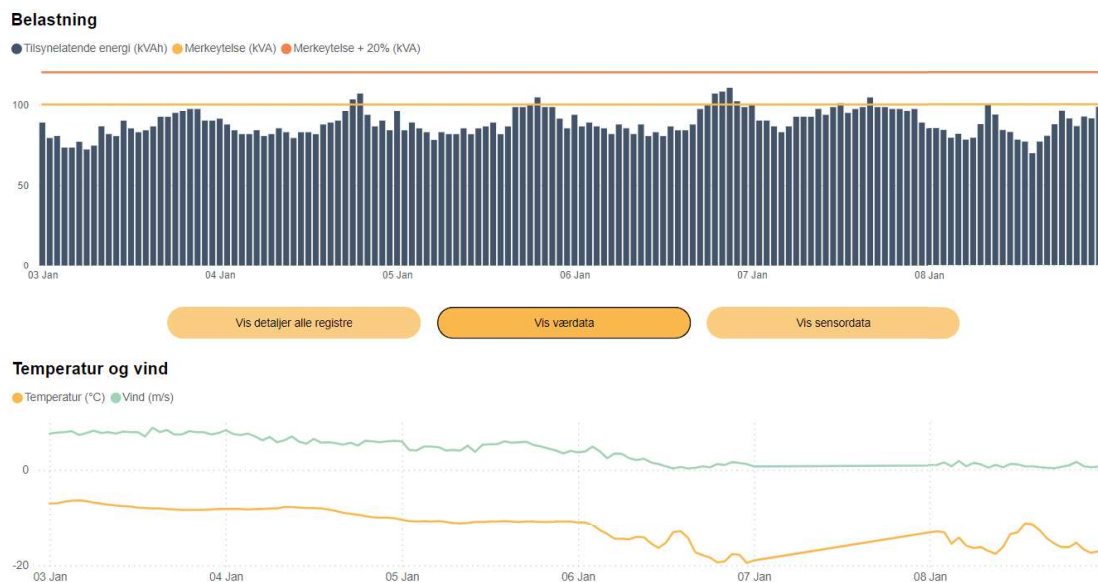
Varighet av overbelastningen summert over hele året viser at de fleste nettstasjoner kun opplever overbelastning noen få timer på årsbasis. Av 330 nettstasjoner er det 147 som har 5 eller færre timer med overbelastning i løpet av året. Omfattende overbelastning er dermed ikke et stort problem, men alvorlig nok for de som opplever det.



Omfanget varierer fra 1% til bortimot 80%. Grafen under viser maks overlast registrert pr nettstasjon, sortert fram lavest til høyest. Omtrent halvparten av nettstasjonene har overlast mellom 1% og 10%. Overlast skjer som regel på kalde dager, da er kjølingseffekten som regel god og effekten av overlast vil være uproblematisk.



Den største utfordringen skjer når man har overlast over lengre tid. For å se hvordan dette utarter seg, må man vurdere flere tidsforløp. Her vil det være mange ulike tilfeller, og eksemplene under er kun ment som illustrasjon for hvordan langvarig overbelastning kan utarte seg.



Figur 1 Belasting på nettstasjon i kald vinterperiode

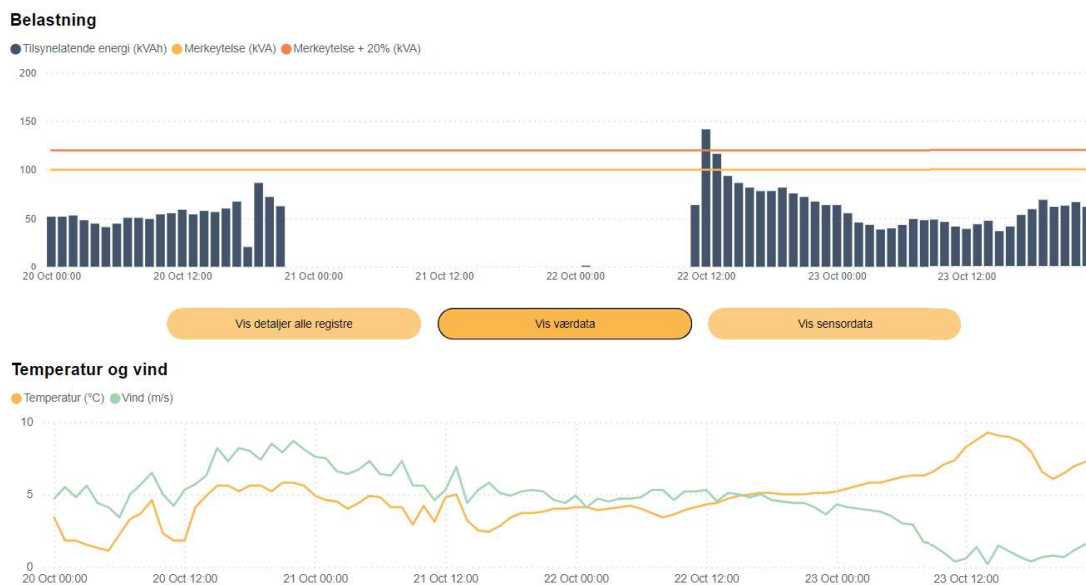
I nettstasjonskretser for husholdning (og fritid samt til dels næring) vil belastningen som regel følge kuldeperioder. Et typisk eksempel på dette vises i Figur 1, der forbruket toppe seg på ettermiddag/kveld når de fleste er hjemme og driver husholdningsaktivitet som krever strøm (lage mat, vaske klær, lade bil). I denne situasjonen er utetemperaturen lav, varigheten på overlast er kort og overlasten er begrenset, og dette vil mest sannsynlig ikke skape utfordringer for transformatoren i nettstasjonen. Det skal imidlertid lite til før forbruket blir for høyt. Om noen år vil det sannsynligvis være flere elbiler i området, og dette kan føre til dramatisk overlast.

I tilfellet i Figur 1 kan man forvente at prediksjon av forbruk skal fungere godt. Forbruksmønsteret følger en logisk utvikling, og det er god sammenheng mellom forbruk og temperatur. Det er imidlertid flere situasjoner der forbruket ikke kan predikeres like enkelt, dette vises i de følgende eksemplene.



Figur 2 Belastning på nettstasjon i alpinanlegg

Figur 2 viser belastning på nettstasjon som leverer strøm til et alpinanlegg. Dette er et forbruk det er vanskelig å predikere ettersom forbruket ikke følger et forutsigbart mønster. Det høye forbruket kommer først og fremst fra snøkanoner som er aktive flere dager i strekk. Temperaturen er lav, men dette forbruket kan likevel være problematisk og skape høy temperatur på nettstasjonen.



Figur 3 Overbelastning etter strømbrudd

Etter et strømbrudd vil det være behov for forbruk av mye strøm på kort tid. Dette kommer av at bygninger har blitt nedkjølt, og det trengs mye energi for å gjenvinne akseptabel temperatur. Et eksempel på dette vises i Figur 3. Det vil i praksis være svært krevende å predikere forbruket fremover i tid når et strømbrudd pågår. Man vil kunne anta at forbruket blir høyt, men det er ikke mulig å forutse

når strømmen kommer tilbake. Dessuten er dette en unntakssituasjon, og det vil normalt sett ikke være mye fokus på å kunne predikere denne typen situasjoner.

3.2 Resultater fra delaktivitet 2: Vurder ulike modeller

I denne piloten har det blitt vurdert statistiske modeller og maskinlæringsmetoder for predikering av belastning på nettstasjonskretser.

- Resultat viser at maskinlæringsmodeller gir bedre ytelse. Det er imidlertid viktig å understreke at maskinlæringsmetoder krever mer arbeid, både i form av forarbeid med input-data, utvalg og tilpassing av modeller, samt kontinuerlig trening. I tillegg må man være oppmerksom på "overfitting", dvs at ML-metoder fungerer bra på treningsdata, men gir vesentlig dårligere prediksjoner på nye data.
- Statistiske metoder er enklere og krever mindre forarbeid, men gir ikke like presise prediksjoner.

I det videre arbeidet med piloten har maskinlæringsmetoder blitt brukt. Metoden XGBoost viser seg å være både stabil og å gi akseptable prediksjoner.

3.3 Resultater fra delaktivitet 3: Utvikle, trene og teste modeller

Utvikling, trening og testing av modeller er dokumentert i [1].

3.4 Resultater fra delaktivitet 4: Evaluering og analyse

Resultat fra XGBoost-prediksjoner (prediksjon basert på maskinlæringsmetoden XGBoost) er evaluert opp mot faktisk målte verdier. I tillegg sammenlignes XGBoost en «naiv» prediksjonsmetode der forrige ukes belastning benyttes som prediksjon¹.

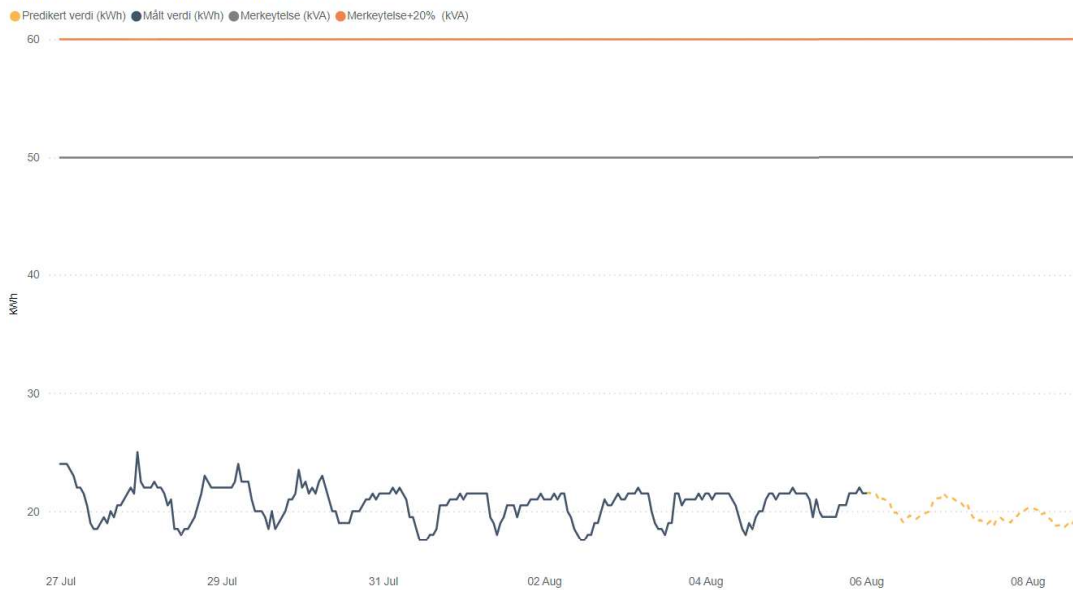
Evaluering viser at XGBoost oftest gir bedre prediksjon enn den naive prediksjonsmetoden. Det er imidlertid overraskende mange ganger at den naive metoden er best. Det kan se ut til at XGBoost har problem med tilfeller der forbruket er svært uregelmessig og tilsynelatende ikke følger noe forutsigbart mønster.

3.5 Resultater fra delaktivitet 5: Sette løsning i operativ drift

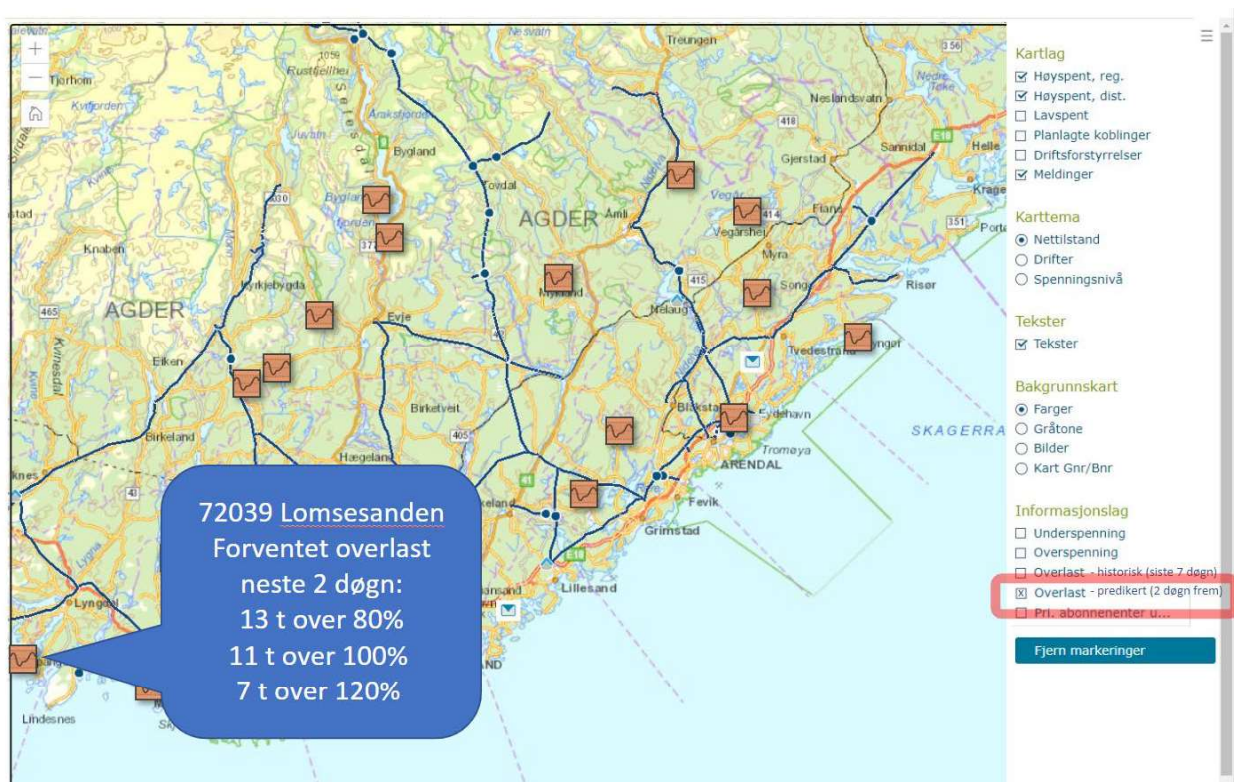
Løsningen er satt i drift hos Glitre Nett, og predikerer daglig belastning for ca 130 nettstasjonskretser. Disse er valgt på grunnlag av at de erfaringsmessig har høy belastning i løpet av et kalenderår, og det er derfor viktig å følge med på disse. Historisk og predikert belastning vises i graf for hver enkelt nettstasjonskrets, se Figur 4. Det er også laget et kartlag som viser nettstasjoner med predikert belastning over 80%, se Figur 5.

¹ Ved stipulering av manglende måleverdier for slutt kunder er det vanlig praksis å bruke timesforbruk fra samme tidspunktforrige uke. Denne viser seg å fungere overraskende bra ettersom forbruket til en vilkårlig kunde ofte har et repeterende ukemønster. Denne metoden brukes her som en enkel prediksjonsmetode for belastning på nettstasjon, dvs som målt belastning fra time x i forrige uke brukes som prediksjon for belastning i time x i inneværende uke.

Målt og predikert verdi på register A- aktiv



Figur 4 Prediksjon av belastning (stiplet oransje linje)



Figur 5 Kart med oversikt over predikert belastning. Status for hver enkelt nettstasjon vises ved når muspeker føres over.

3.6 Innovasjoner fra Piloten

Tabell 3 Beskrivelse av innovasjoner i forskningsrådets kategorier

Forskningsrådets kategorier	Beskrivelse	Antall
Ferdigstilte nye/bedre metoder/modeller/ prototyper	Ny metode for prediksjon av belastning	1
Bedrifter utenfor FMEen som har innført nye/forbedrede metoder eller modeller eller teknologi		
Bedrifter innenfor FMEen som har innført nye/forbedrede arbeidsprosesser		
Bedrifter innenfor FMEen som har innført nye/ forbedrede metoder eller modeller eller teknologi		
Inngåtte lisensieringskontrakter		
Registrerte patenter		
Ferdigstilte nye/forbedrede produkter	Løsning for prediksjon av belastning er satt i operativ drift hos Glitre Nett	1
Ferdigstilte nye/forbedrede prosesser		
Ferdigstilte nye/forbedrede tjenester		
Nye foretak som følge av FME'en		
Nye forretningsområder i eksisterende bedrifter		

4 Tekniske/faglige erfaringer fra Piloten

God tilgang på data er viktig ved utvikling av ML-modeller. Piloten har nytt godt av at det eksisterte et omfattende og komplett datagrunnlag ved oppstart. Glitre Nett har montert måling på alle nettstasjoner (over 8000) i Agder, og har derfor et stort datasettsom strekker seg flere år bakover i tid. Dette har gitt et solid grunnlag for å bygge ML-modeller.

Selve utviklingen av egnede ML-modeller har vist seg å være krevende. Dette kommer delvis av at nettstasjoner kan ha svært ulike mønster og dynamikk for energiflyt. Enkelte kan ha stabile, repetitive mønster basert på sesongvariasjon. Andre kan ha tilsynelatende stokastisk oppførsel. Nettstasjoner med mange kunder vil ha en mer stabil og forutsigbar kurve, mens en nettstasjon med få kunder (1-5) typisk vil ha en belastningskurve med tydelig stokastisk karakter.

Som konsekvens av dette ble det konkludert med at det ikke er mulig å lage en enkelt modell som kan trenes for å predikere belastning for alle nettstasjoner. Det ble derfor besluttet å trene en separat ML-modeller for hver enkelt nettstasjon. Alle modellene er basert på samme teknologi (XGBoost), men trening skjer separat basert på data for hver enkelt nettstasjon.

Idriftsetting av pipeline (driftsmiljø) har også tatt en del tid. Dette har ikke vært spesielt komplisert, men ettersom dette var første gang vi satte opp denne typen infrastruktur, var det en del nye løyper som måtte gås opp. Selve prediksjonen går nå daglig for ca 130 nettstasjoner.

4.1 Oppsummering

Utvikling av ML-modeller for prediksjon har gitt mye innsikt i arbeidsmetodikk og kompleksitet ved denne typen analyser.

5 Kost-/nyttevurderinger basert på resultatene for Piloten

5.1 Kostnader

Kostnader er i stor grad knyttet til

- Etablering av ML-modeller og rutiner/prosesser
- Vedlikehold (ny trening, feilretting etc)
- Lagring og prosessering av data

Det første punktet er en engangsjobb, punkt to og tre medfører regelmessige driftskostnader. Kostnadene her er svært begrenset sett i forhold til gevinstene i neste delkapittel.

5.2 Nyttevurderinger

Prediksjon av lastutvikling er viktig for å kunne se hvor våre nettstasjonstransformatorer kan komme til å gå i overlast. Dette er essensielt for å kunne drifte et nett dynamisk, dvs ta aksjon når man ser at en overlastsituasjon er forventet. Man kan dermed unngå overlast og i verste fall driftsavbrudd. Nytten her er svært stor:

- Man unngår eller reduserer sannsynlighet for driftsavbrudd, og unngår dermed ulempe og potensiell skade hos kunder
- KILE-kostnader bli lavere ettersom man får færre og/eller kortere avbrudd. Dette gir dermed bedret økonomisk resultat hos nettselskapet
- Ved å unngå overlast sikrer man samtidig lenger levetid på komponenter, som igjen er bra for økonomi og opetid.

Dagen prediksjon strekker seg ca 2 døgn fremover i tid. Det er spesielt to tiltak som er aktuelle når man ser at overlast blir predikert:

- Preventivt tiltak: Bruke fleksibilitet til å ta ned lastpådraget, og dermed unngå / redusere sannsynlighet for overlast. I dag er fleksibilitet tilgjengelig på følgende måter:
 - Kunder som har inngått tariffavtalen «Fleksibelt forbruk»
 - Kunder som er tilknyttet på vilkår
 - Kunder som deltar i fleksibilitetspiloter
- Preventivt tiltak: Oppgradere transformatoren i den berørte nettstasjonen. Dette krever imidlertid at man tar aksjon raskt ettersom det går med noe arbeidstid for å bytte en transformator. I tillegg er det en forutsetning at nødvendig materiell og personell er tilgjengelig.

- Reaktivt tiltak (beredskap): Etablere beredskap slik at man er klar til å rykke ut og ta aksjon dersom avbrudd skjer.

6 Referanser

- [1] Adrian Langemyr, masteroppgave UiA 2020: *Predicting Electrical Power Consumption on Yearly Events for Substations: Algorithm Design and Performance Evaluations*. URL: <https://uia.brage.unit.no/uia-xmlui/handle/11250/2683504>
- [2] Per-Oddvar Osland og Adrian Langemyr, Agder Energi Nett: *Prediksjon av belastning*. Presentasjon på CINELDI Webinar 20.11.2020
- [3] Per-Oddvar Osland, Agder Energi Nett: *Maskinlæring – overgang frå pilot til drift*. Presentasjon på CINELDI-dagene 10.11.2020
- [4] Per-Oddvar Osland, Agder Energi Nett: *The potential of ML for Smart Grids*. Presentasjon på AI&ML for the Smart Grid 2020 – 9. Sept. 2020

FME CINELDI

Host: SINTEF Energy Research in cooperation with NTNU
Visiting address: Sem Sælands vei 11, N-7034 Trondheim
Post address: P.O.Box 4761 Torgarden, N-7465 Trondheim
Telephone: +47 454 56 000*
E-mail: cineldi@sintef.no
Enterprise/VAT No: NO 939 350 675 MVA
<http://www.cineldi.no>



Prediksjon av belastning

Webinar CINELDI 20.11.2020

Per-Oddvar Osland og Adrian Langemyr, Agder Energi Nett

Agenda

Intro

Prediksjon av belastning

Oppsummering og vidare arbeid

Hypotese: Avansert analyse (maskinlæring, statistikk) har stort potensiale!

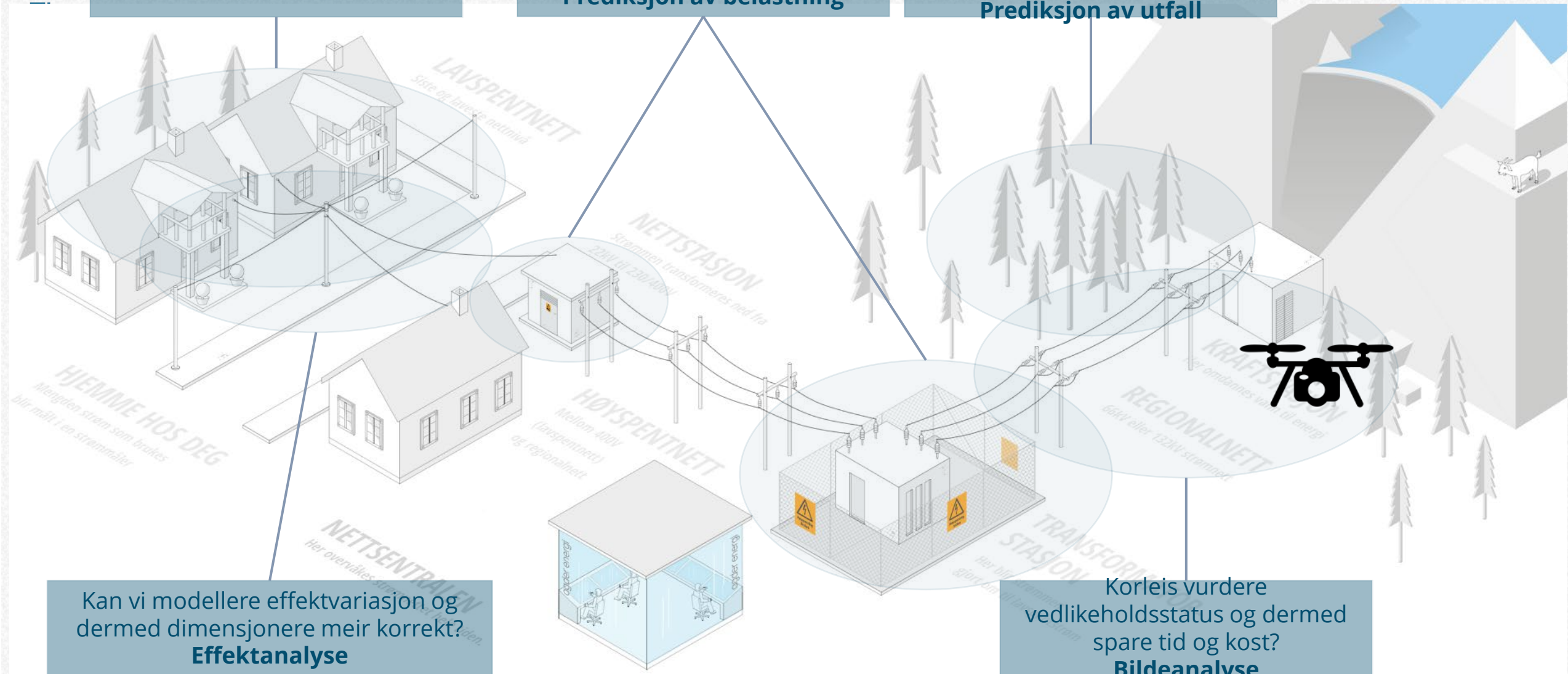
Korleis forstå kunden?
Clustering og klassifisering

Korleis forutsjå belastning på vitale komponentar i nettet?
Prediksjon av belastning

Kan vi forutsjå utfall basert på ver, vegetasjon og utfallshistorikk?
Prediksjon av utfall

Kan vi modellere effektvariasjon og dermed dimensjonere meir korrekt?
Effektanalyse

Korleis vurdere vedlikeholdsstatus og dermed spare tid og kost?
Bildeanalyse



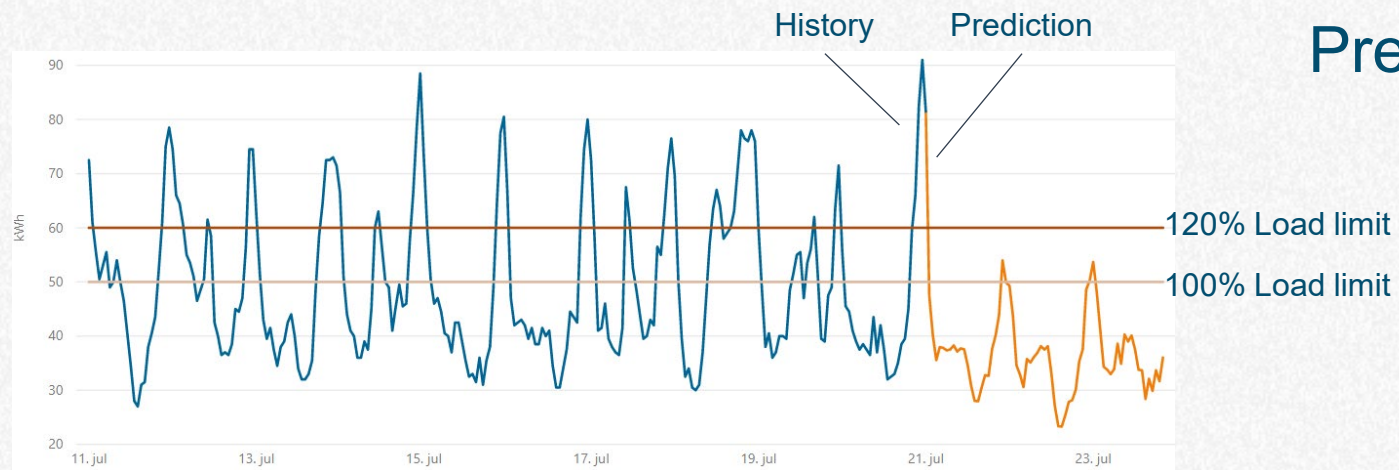


Effektanalyse

Prediksjon av belastning

Bildeanalyse

Prediksjon av nettstasjonsbelastning



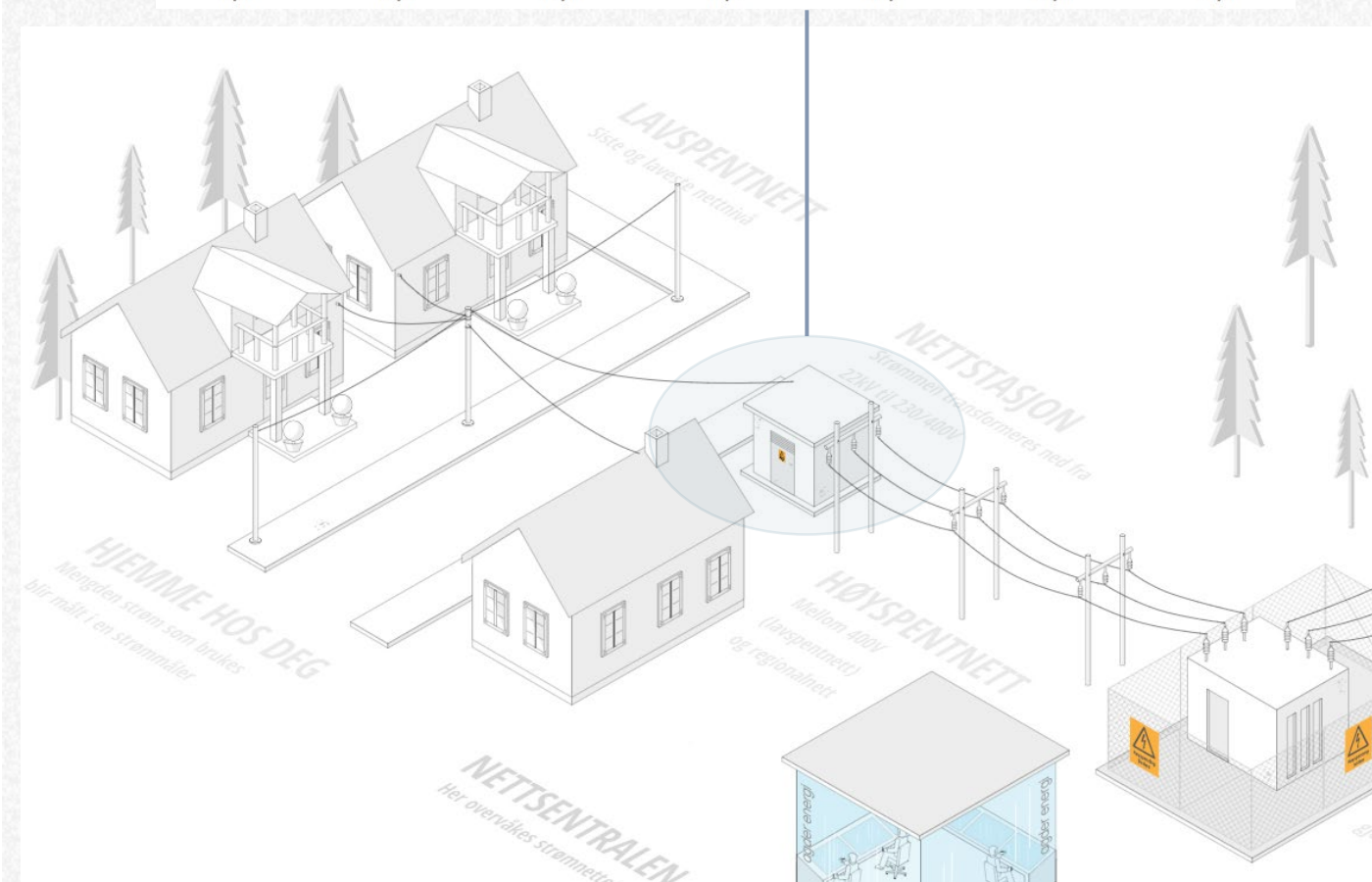
Enkelte nettstasjoner vert overbelasta i topplast-situasjon. Dette medfører

- Auka behov for vedlikehold
- Redusert levetid
- Avbrudd

I sum: Auka kostnad, redusert stabilitet i nettet

Prediksjon av belastning viktig for å

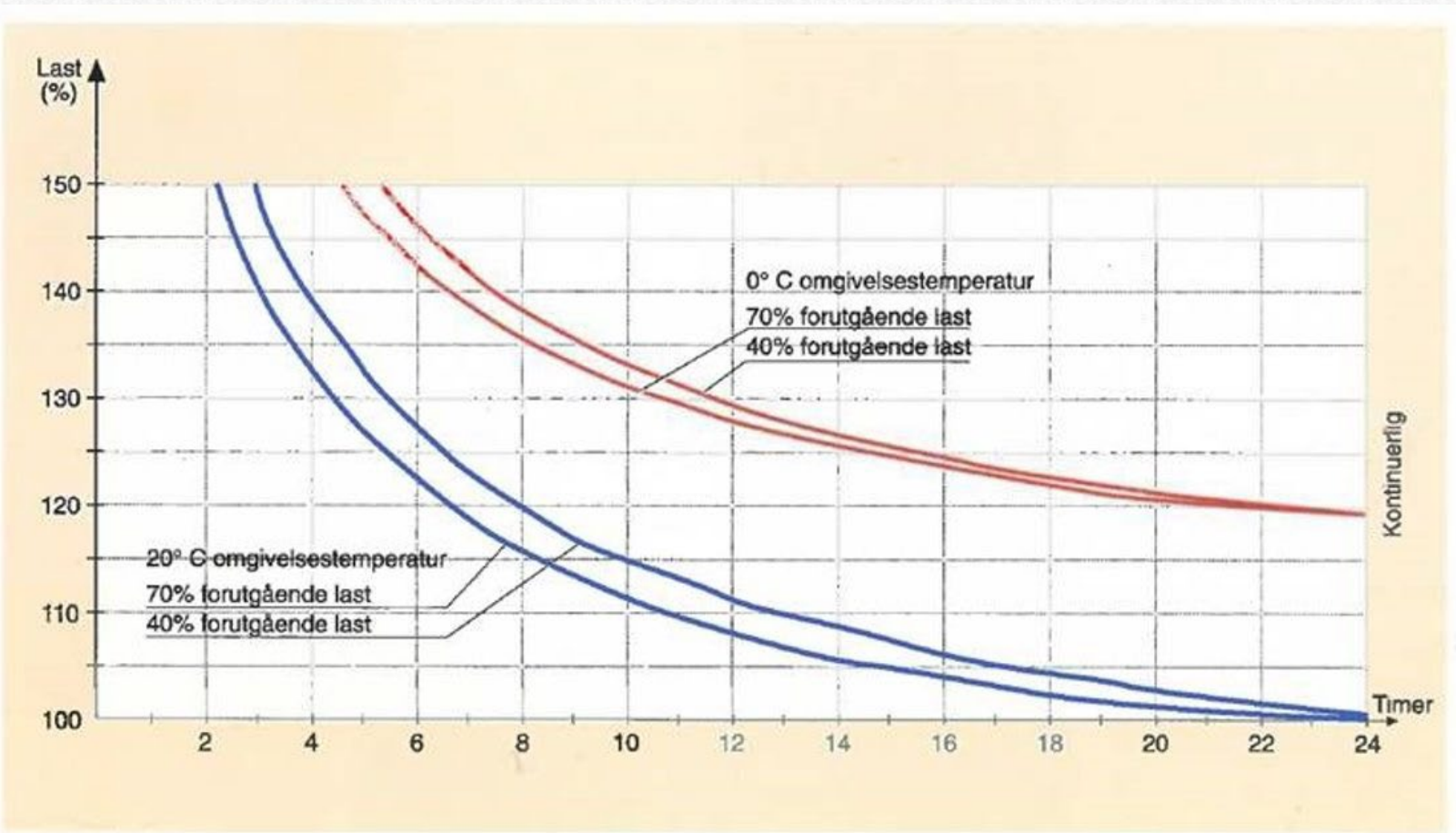
- Redusere last hos kundar (tariff «Fleksibelt forbruk» eller fleksibilitetsmarked)
- Etablere beredskap
- Reparere / oppgradere



Relasjon mellom overlast og redusert levetid

ABB: Indikert nivå for maks belastning

Belastning over kurvene medfører redusert levetid



Agenda

Intro

Prediksjon av belastning

Oppsummering og vidare arbeid

Substation load

- 121 substations
- Hourly metered load (kWh)
- Collected every 6 hours
- 2-3 years of historical data

Weather

- Temperature, humidity, rain/snow etc.
- Source: Norwegian Meteorological Institute
- Location precision: 2,5 km

Time stamps

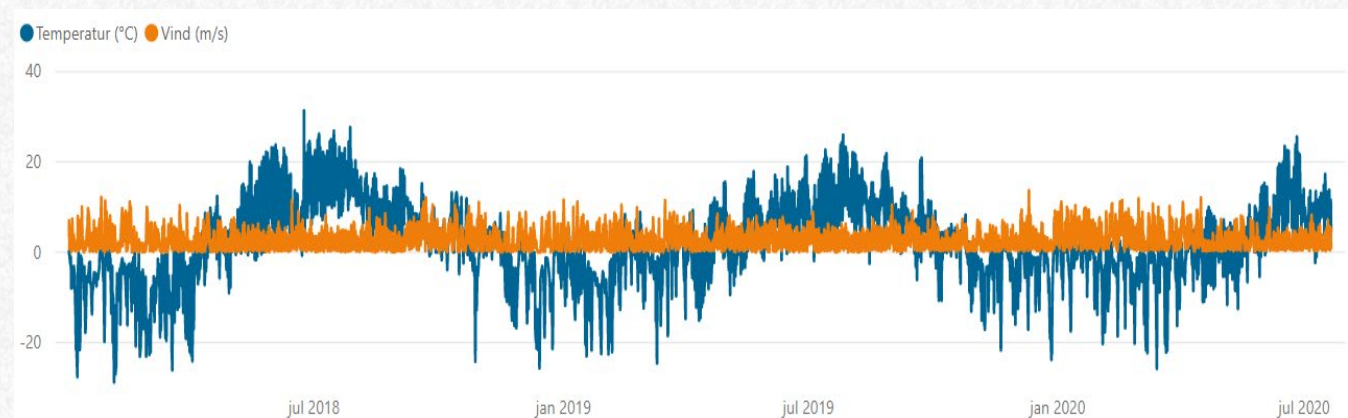
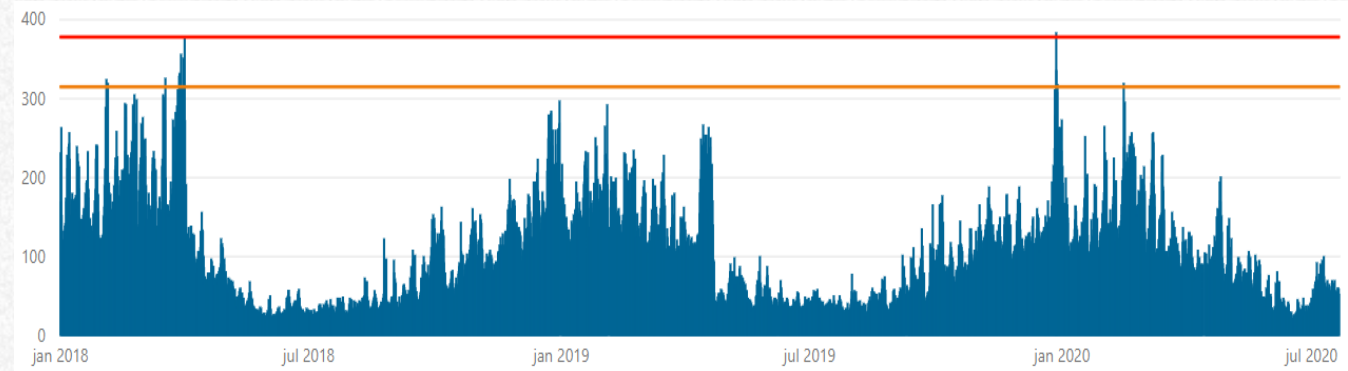
- Month, week of year, day of week, etc.

Derived features

- Diffs and lags
- Holidays

Data we did not use (yet):

- Customer classification (residential, industry, vacation dwelling, etc.)
- Demographic data
- EV penetration



Data frå AMS-utrulling (2015-2018)

Belastning målt på alle nettstasjonar

460 - Regional grid

Substations, production plants, and large customers (High Voltage)
El. Meter: Cewe prometer 100
Communication: Fixed broadband or Mobile (4G)

8300 – Secondary Substations

El. Meter: Kamstrup Ominpower
Multi-instrument IME Nemo D4-Le
Ground fault sensor
Additional I/O ports for sensors
Communication: Mobile (3G/2G)

200 000 - Customer premises

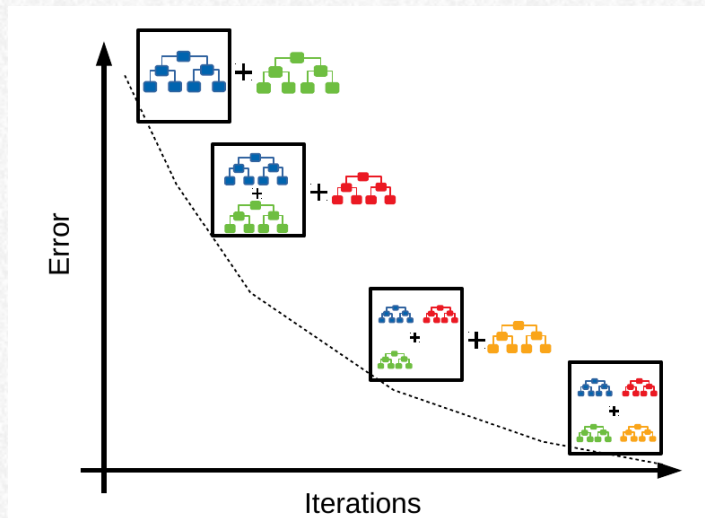
El. Meter: Kamstrup Ominpower
Communication: Radio Mesh



Gradient Boost Trees

XGBoost - Extreme Gradient Boost

LGBM – Light Gradient Boosting Machine

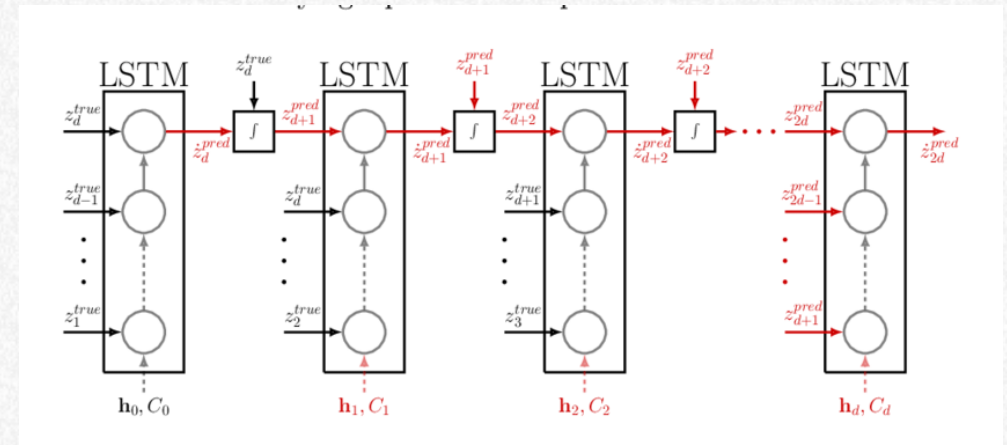


Gradient Boost Trees are an ensemble of trees. At each iteration a tree is added, trained on the *gradient of the errors* from the previous iteration.

Source: Theodore Vasiloudis (<http://tvas.me/>)

Recurrent Neural Network

LSTM - Long short-term memory



Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video).

Source: Wikipedia, Pantelis Rafail Vlachas

One common model or several individual models?

A. One common prediction model for all 2ndary substations

- Benefit: Can train one model on all available data
- or

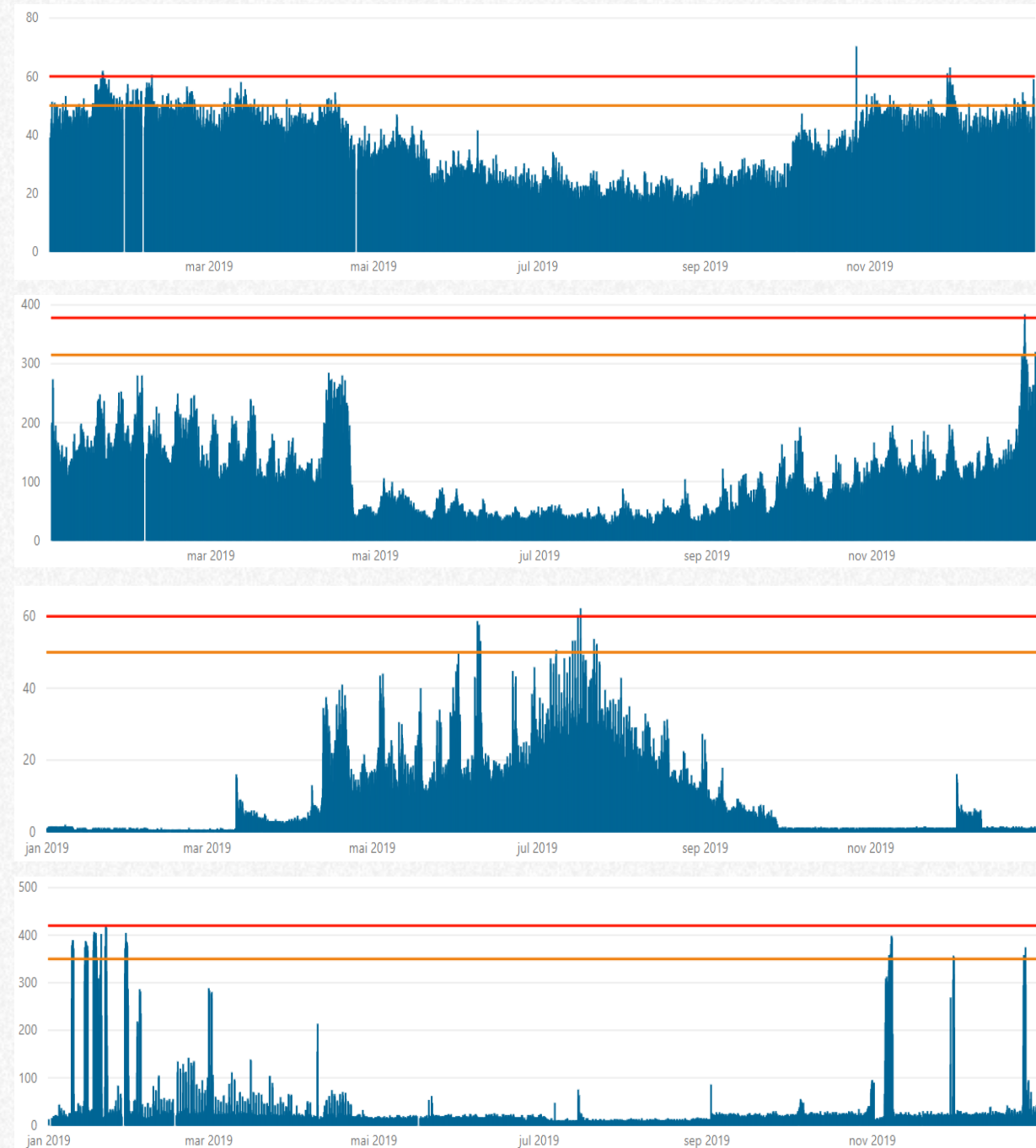
B. Train one individual model for each 2ndary substation

- Benefit: Model better suited for each substation
 - Benefit: More flexible in terms of retraining
- or

C. Make clusters of 2ndary substations, one prediction model for each cluster

Conclusion:

- B (individual models) work best in general due to very different load profiles
- A (common model) sometimes gives best result with LGBM
- C (clusters) not evaluated in order to keep the model simple and robust. Potential further work.



Statistical method:

SARIMAX – Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors is an extension of the well-known ARIMA method. It takes seasonal trends into account when making predictions and can use information from additional time series.

Naïve methods:

Methods based on guessing the forecasted load to be the same as the observed load some time in the past:

naive_24h_mean – Average over the 24 past hours

naive_24h_median – Median over the 24 past hours

naive_seasonal – Same value as observed last year

naive_last_period – Repeated values from the past 3 days

naive_last_week – Same value as observed last week

RMSE - Root Mean Square Error

MAPE - Mean Absolute Percentage Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where y_i = actual load value at hour i , \hat{y}_i = predicted value, n = number of hours

We prefer using **MAPE** as it allows for comparison between time series where the load values (the y_i 's) vary in size.

RMSE and MAPE are calculated over an 80-day period during February – April 2020 for 26 substations.

ML prediction methods

XGBoost
(individual models)**LGBM**
(one common model)**LGBM**
(individual models)LSTM
(dropped)

LSTM requires a long training period, and was dropped early in the process

compared to

Baseline methods

SARIMAXnaive_24h
meannaive_24h
mediannaive
seasonalnaive last
periodnaive last
week

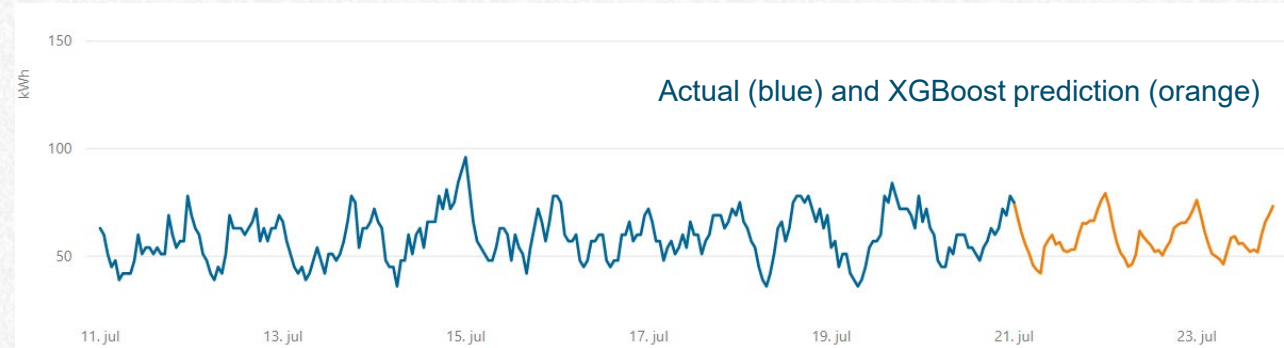
measured with

$$MAPE = \frac{1}{n} \sum_{I=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Much of the electricity consumption is depending on temperature. There is a large number of residential customers, and the load variation is fairly stable.

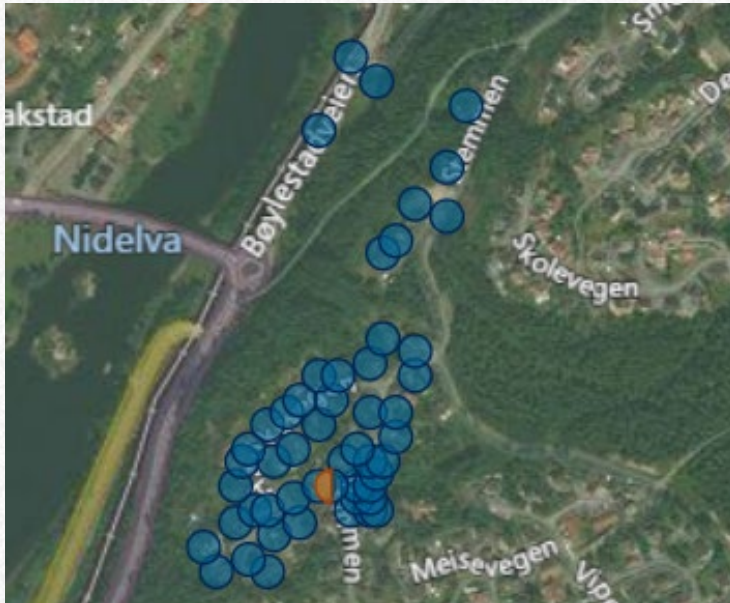
Best model: LGBM (individual models)

Example 1: Substation in a residential area

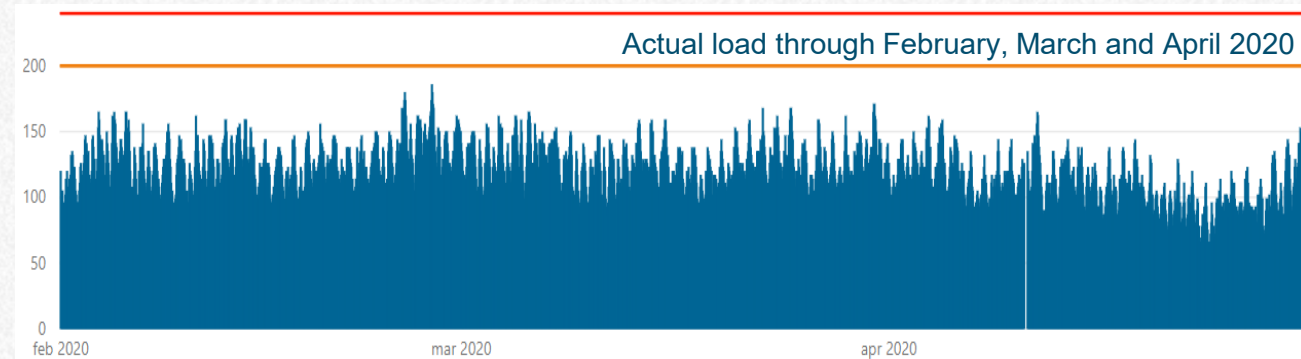


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
7.908	7.746	6.763	12.921	12.597	12.973	13.802	12.595	12.926



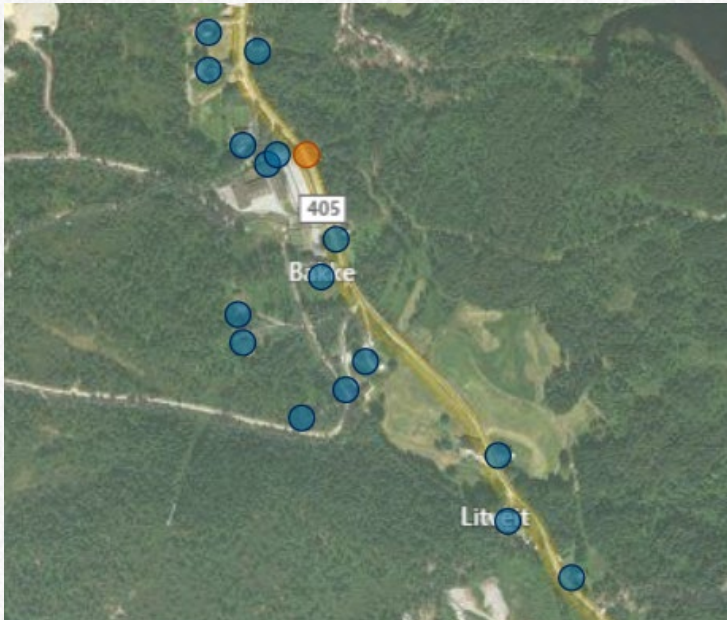
Number of customers: 60
Type of area: Residential



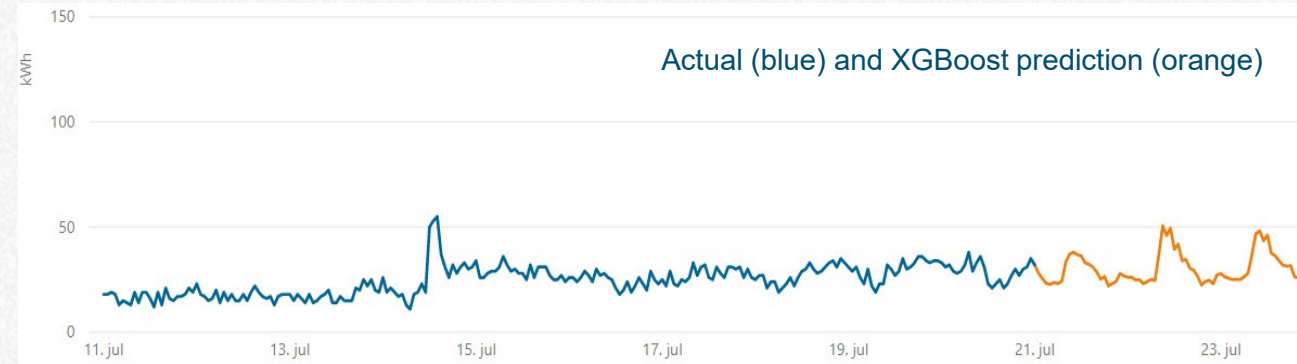
Example 2: Substation in a rural area

A school building dominates the consumption in the area. This gives a clear seasonal load profile, and simple models such as SARIMAX and “naïve last week” perform well.

Best model: SARIMAX

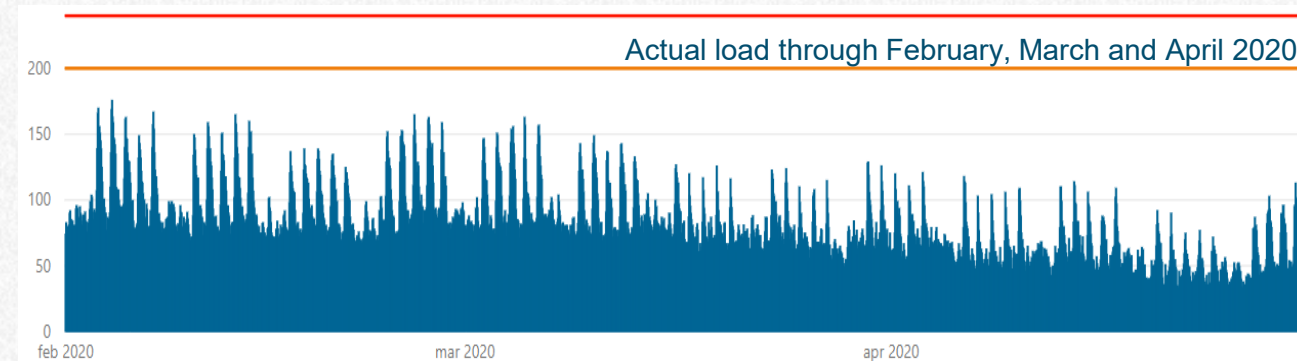


Number of customers: 16
Type of area: Dwellings, school, small industry



MAPE scores based on prediction for February, March and April 2020

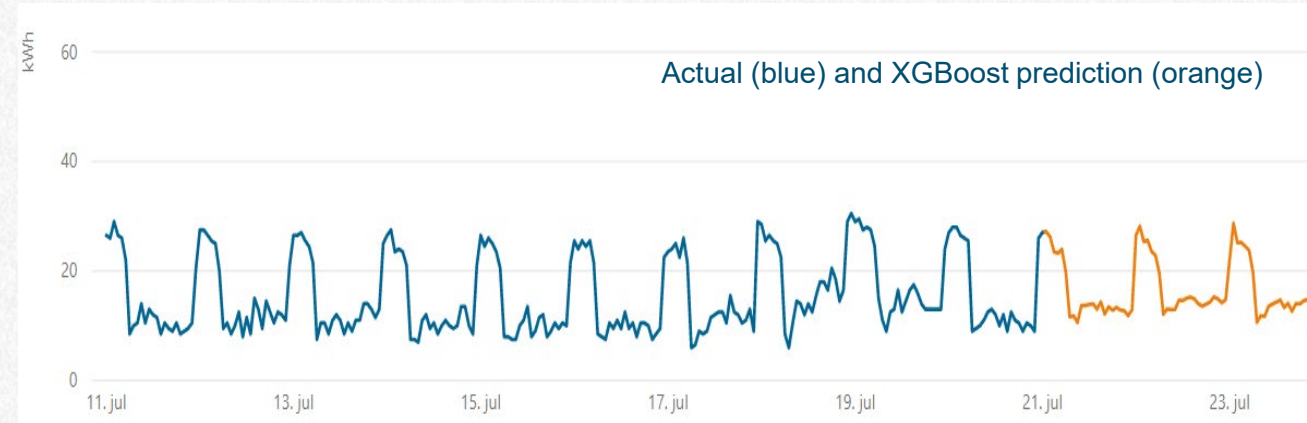
XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
15.395	14.282	15.31	13.088	21.62	19.576	30.433	20.73	16.119



Example 3: Substation in rural area with periodic load

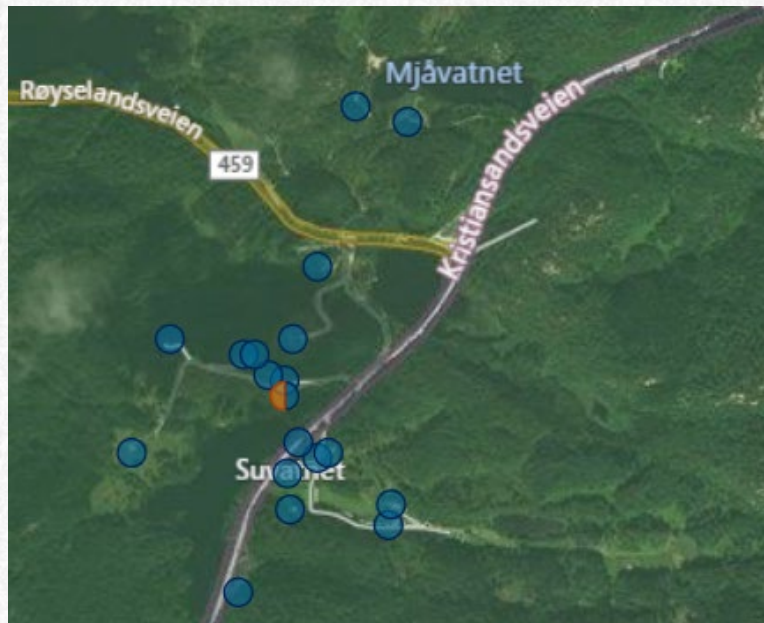
The substation load is very regular and periodic, and simple models such as SARIMAX and “naïve last period” perform well.

Best model: Naïve last period

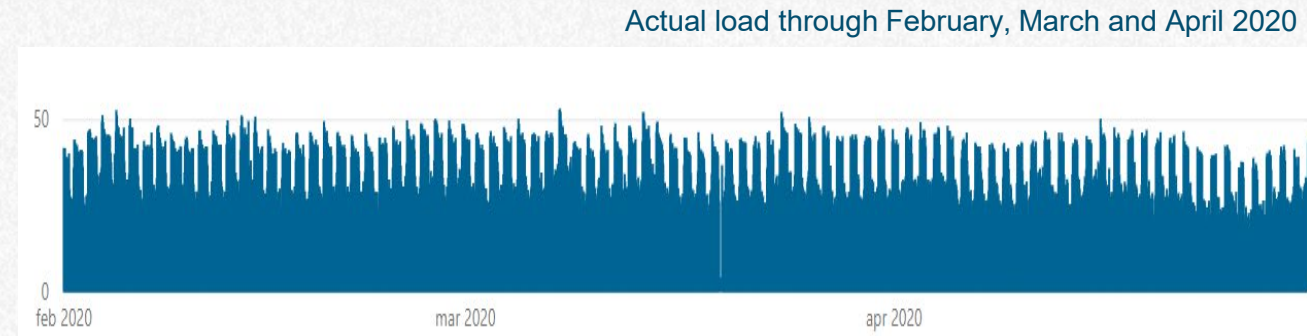


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
10.609	9.883	9.22	9.131	20.236	19.435	16.057	8.925	10.268



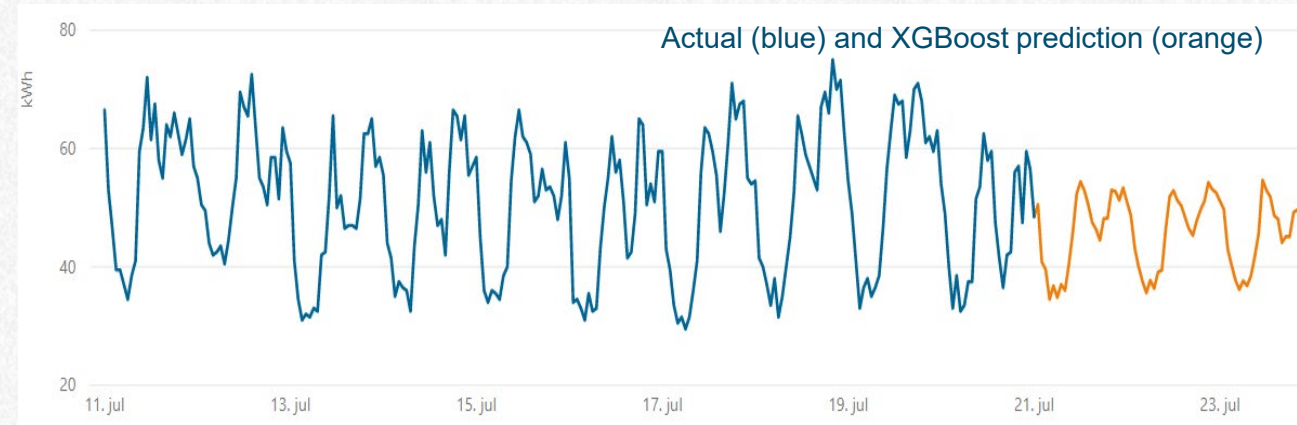
Number of customers: 19
 Type of area: Dwellings, small industry



Example 4: Residential area

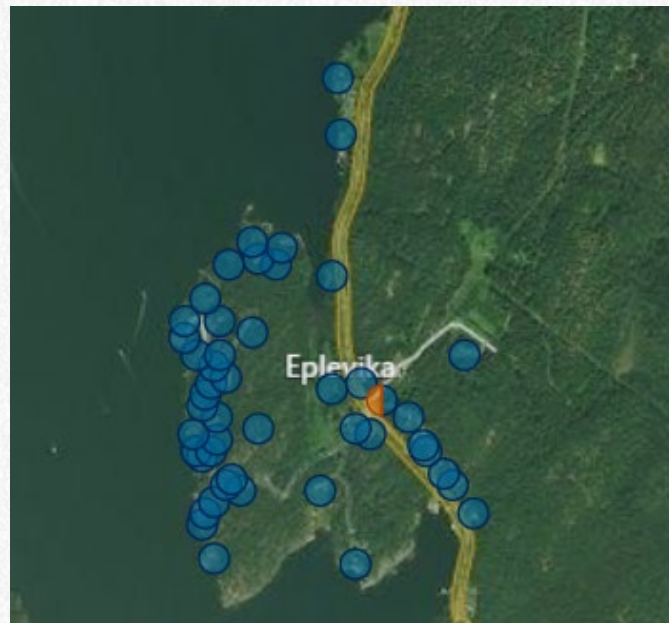
Residential area with stable, temperature-dependent consumption. ML-methods perform best, with some of the naïve methods close behind.

Best model: LGBM (based on one common model)

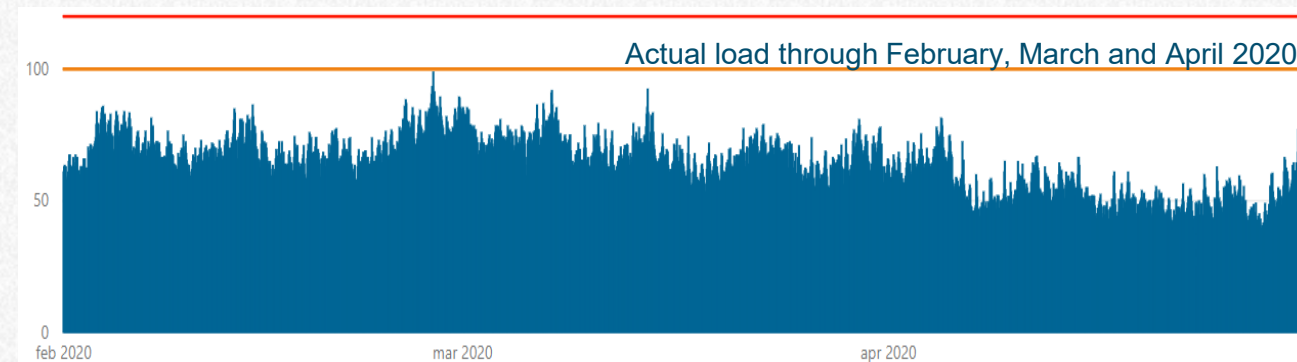


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
7.57	7.232	7.528	11.434	9.57	9.749	17.555	12.278	12.666



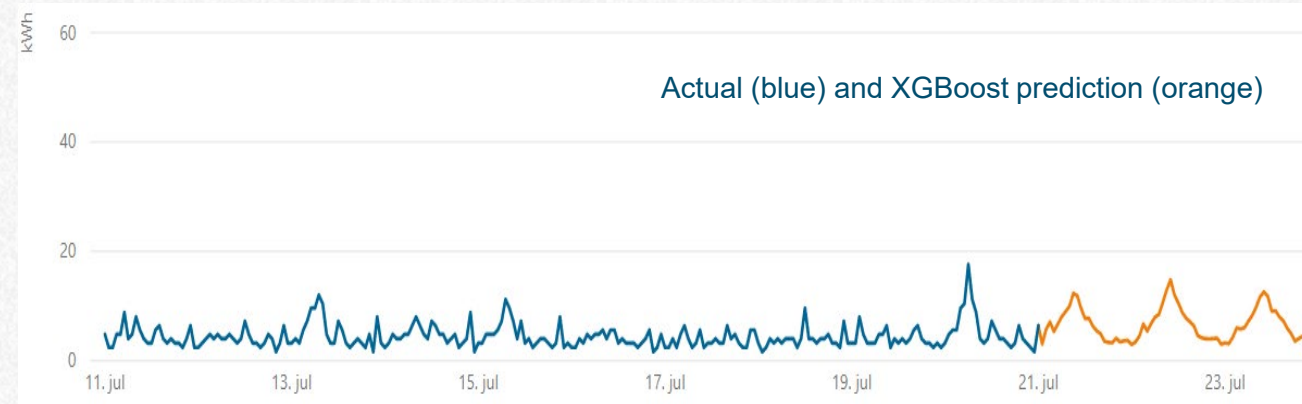
Number of customers: 50
 Type of area: Dwellings, some small industry.



Example 5: Substation load dominated by school building

The substation serves a school complex. Consumption during the prediction period (Feb-April 2020) was heavily influenced by temperature and the corona situation. In addition there is a period with missing data. Most models perform poorly.

Best model: XGBoost

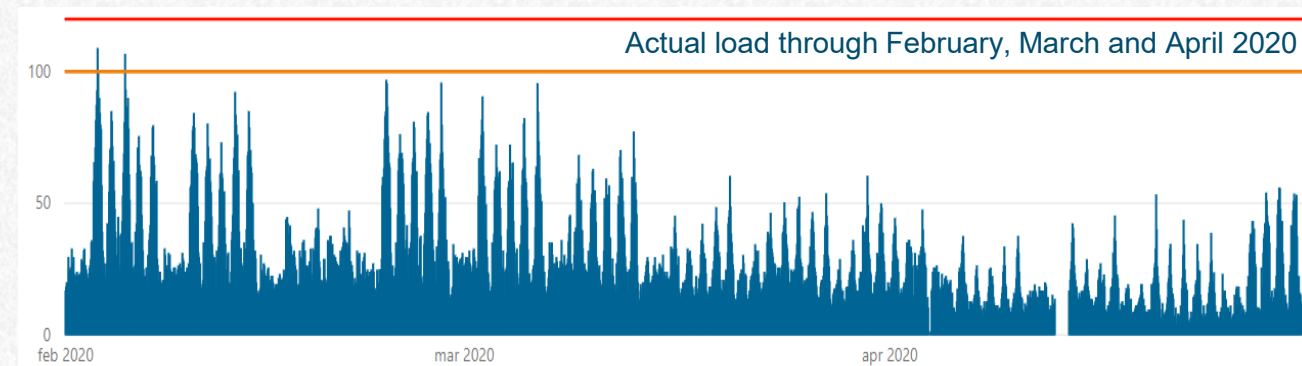


MAPE scores based on prediction for February, March and April 2020

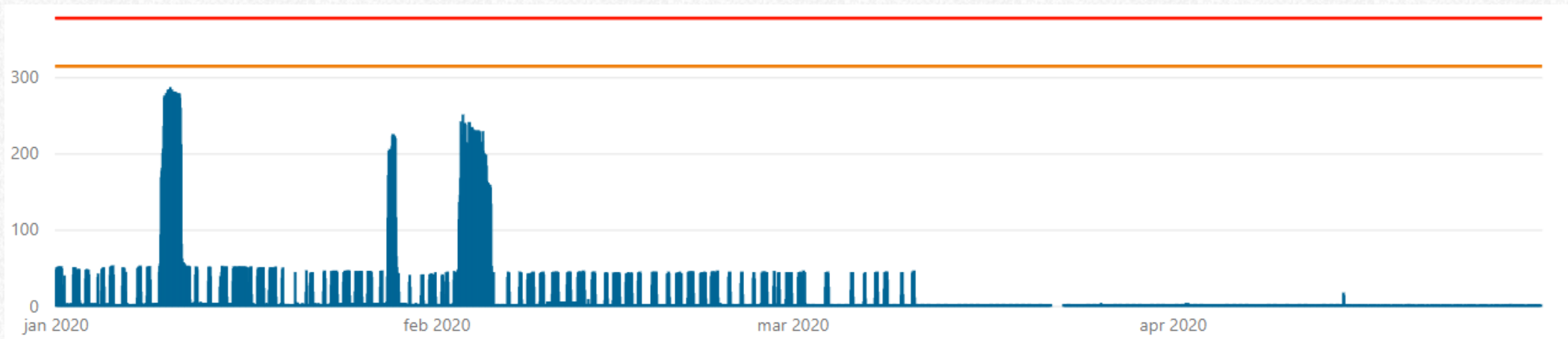
XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
26.144	32.688	33.481	32.509	46.281	42.475	58.008	40.542	37.976



Number of customers: 6
Type of area: School.



Some things are hard to predict ...



Summary: Which forecasting model is best?

There is no «one model best in all cases».

LGBM points out as the model with best score (lowest error measure) most often.

LGBM with **individual models** (i.e. model trained individually for each substation load profile) emerges as the best option.

XGBoost perform well in most cases where LGBM is accurate, but usually ends up with a slightly higher error.

SARIMAX and the naive models perform well in cases where the load is repetitive.

Count of «best model score» achieved

Model score (error measure)	XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
RMSE	2	6	12	2	2	0	0	1	0
MAPE	1	7	12	1	0	2	0	2	0

Full-automatisert ML verdikjede i produksjon

Predikerer belastning for 123 nettstasjonar



67 t
Forecast
frå met.no

Innsamling
målt
belastning

**Data-
inn-
samling**

Data prepp

Prediksjon
(XGBoost, ein
modell pr nettstasjon)

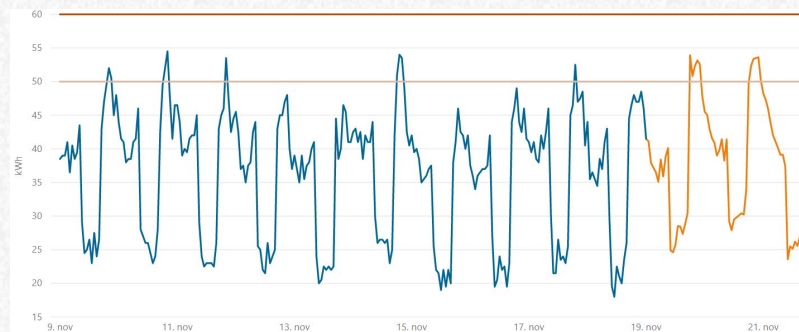
Presentasjon

~ 60 t
prediksjon

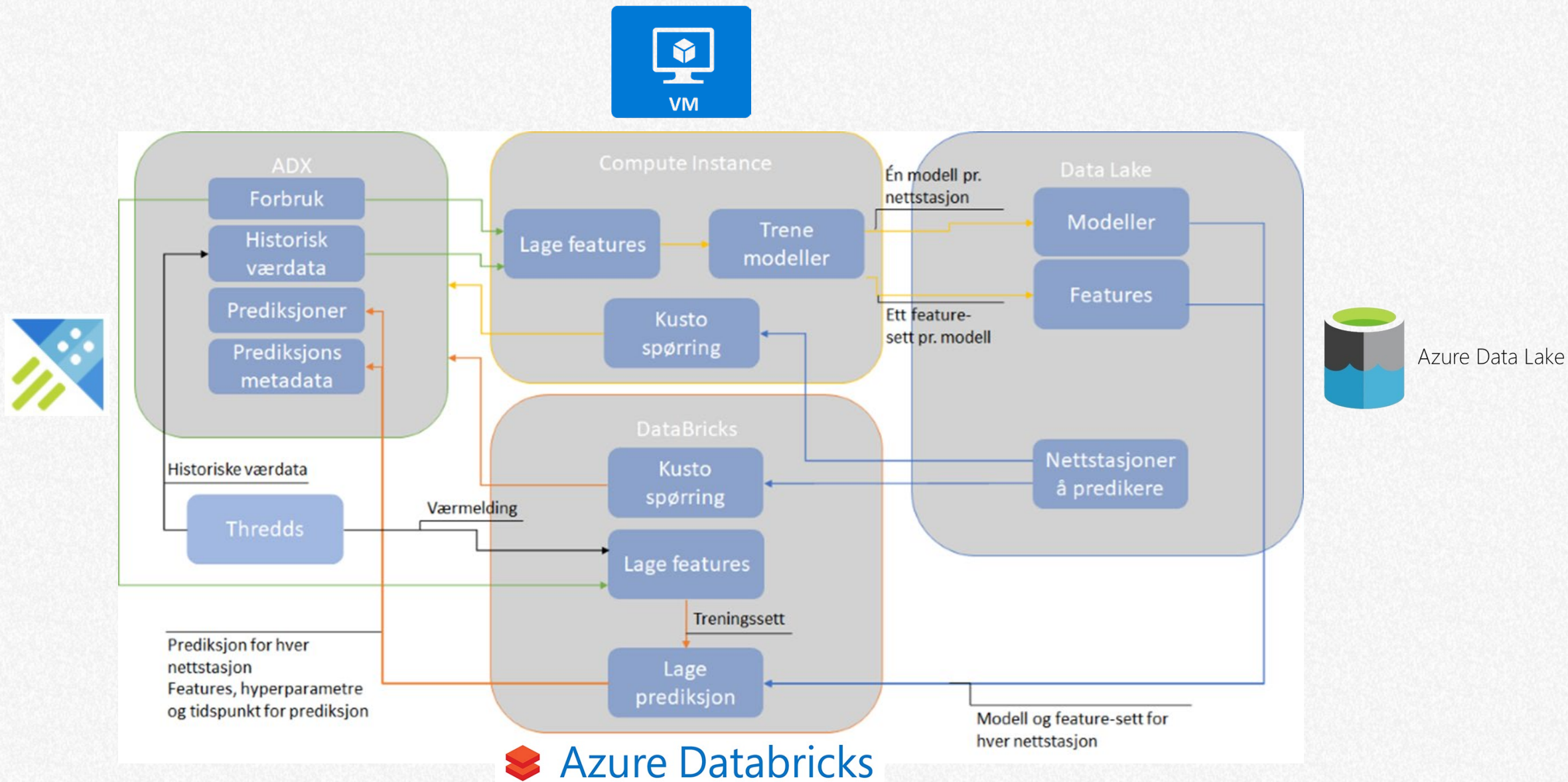
Trening
Trene modell
basert på nye
data

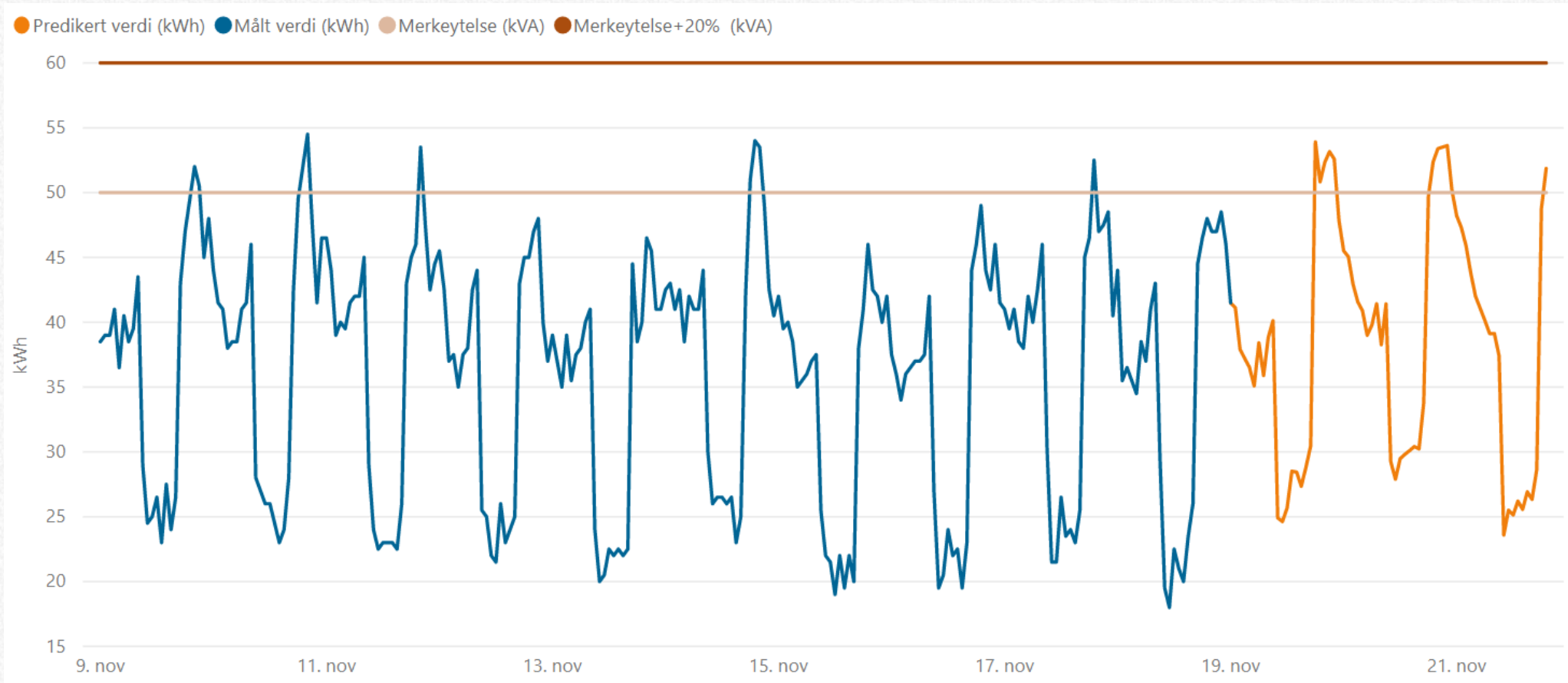
**2x / år:
Trene modell**

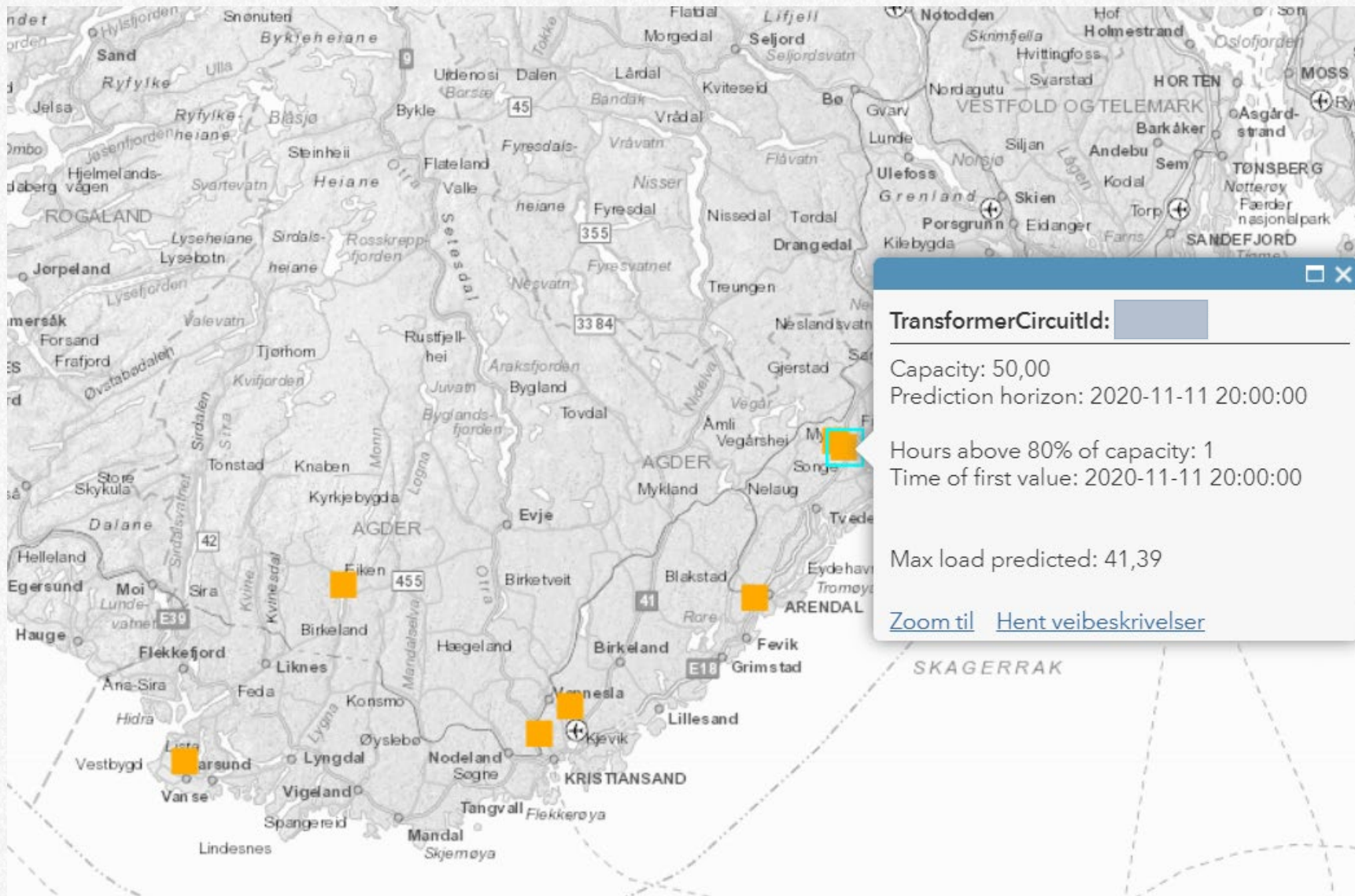
Dagleg



Trening og prediksjon i Microsoft Azure







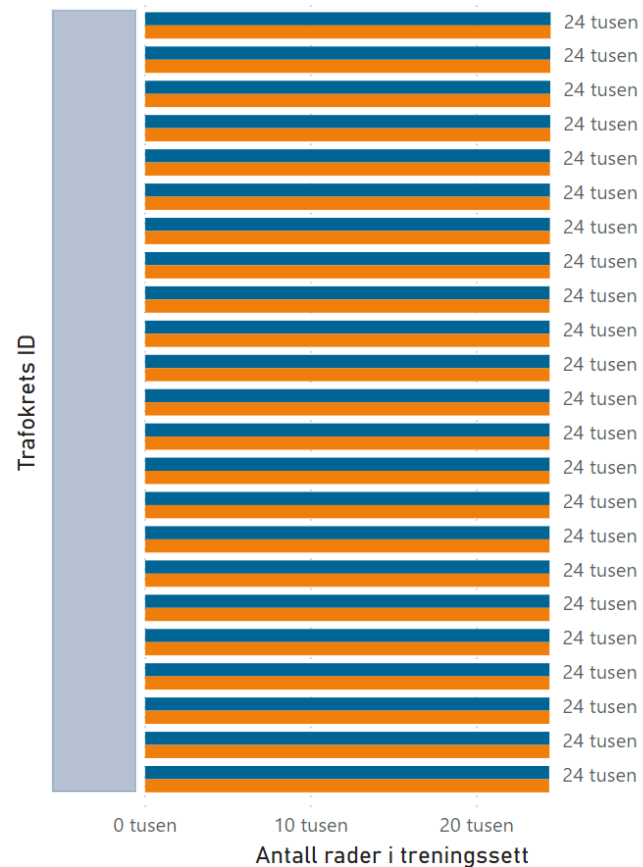
Siste prediksjon

Prediksjonsmonitorering v1.0

onsdag 11. novemb...

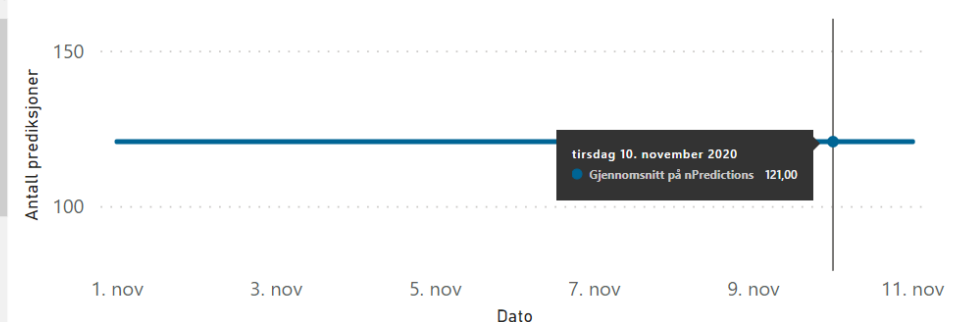
Antall rader i treningssett per trafokrets

● Ved siste komplette sammenligning ● Ved siste prediksjon

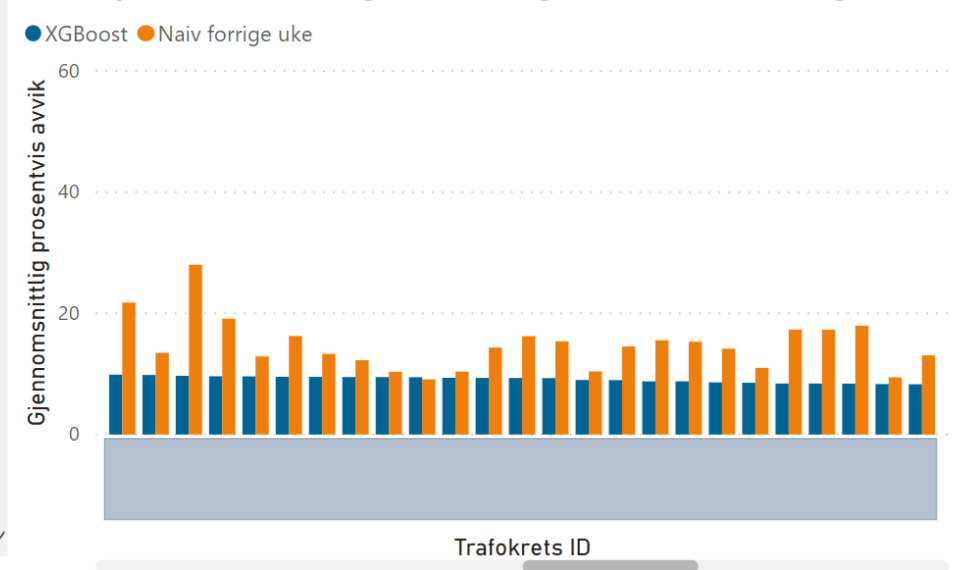


Trafokrets ID	Antall predikerte verdier
1	68
2	68
3	68
4	68
5	68
6	68
7	68
8	68
9	68
10	68
11	68
12	68
13	68
14	68
15	68
16	68
17	68
18	68
19	68
20	68
21	68
22	68
23	68
24	68
25	68
26	68
27	68
28	68
29	68
30	68
31	68
32	68
33	68
34	68
35	68
36	68
37	68
38	68
39	68
40	68
41	68
42	68
43	68
44	68
45	68
46	68
47	68
48	68
49	68
50	68
51	68
52	68
53	68
54	68
55	68
56	68
57	68
58	68
59	68
60	68
61	68
62	68
63	68
64	68
65	68
66	68
67	68
68	68

Antall prediksjoner per dag



Prediksjonsavvik sammenlignet med forrige ukes forbruk for 3 dager



Agenda

Intro

Prediksjon av belastning

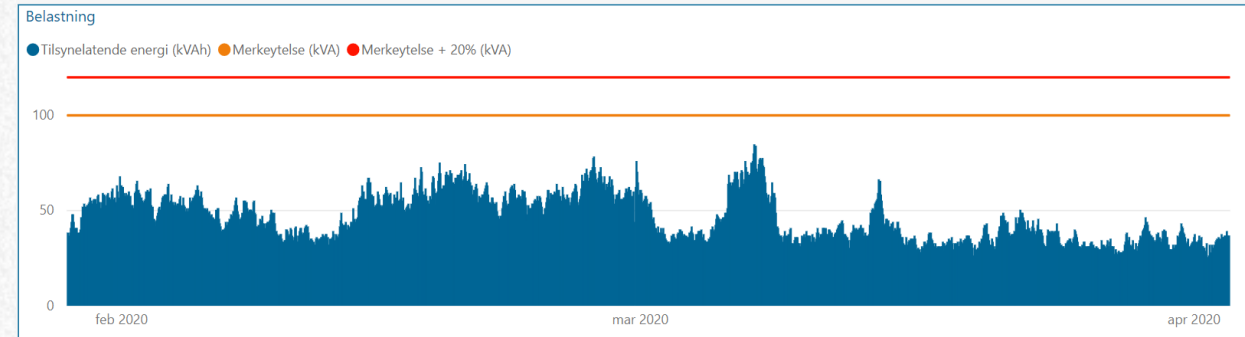
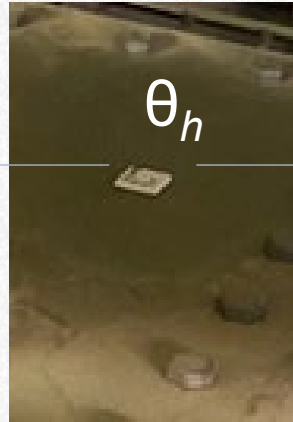
Oppsummering og vidare arbeid

- Involver fagressursar for å bekrefte nytteverdi
- Prediksjon vil aldri treffe 100% - husk å overvåke og formidle kvalitet
- XGBoost gir akseptable resultat
- Etabler stabil, fullautomatisert verdikjede (tidkrevande!)
- Tenk skalering

θ_h – «hot spot temperature» - er viktigare indikator enn belastning

Neste steg:

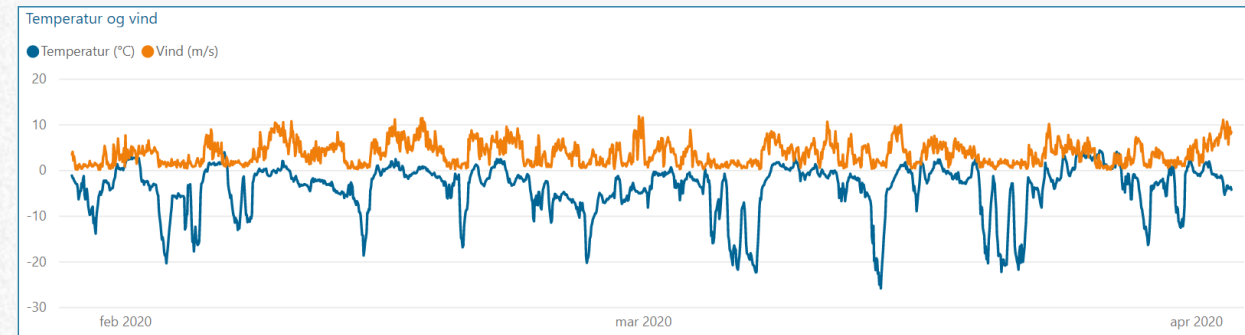
- Lære utvikling av θ_h - hot spot temp basert på ML + belastning + verdata + målt temperatur (*Disruptive Technologies*)
- Predikere utvikling av θ_h på alle nettstasjonar.



Vis detaljer alle registre

Vis værdata

Vis sensordata



Temperatur (C)

Sensornavn ● 04.0204Temp rom kiosk ● 04.0204Temp trafo bunn ● 04.0204Temp trafo topp ● 04.0204Temp ute



per-oddvar.osland@ae.no
adrian.langemyr@ae.no

Maskinlæring – overgang frå pilot til drift

CINELDI 10.11.2020

Per-Oddvar Osland, Agder Energi Nett
Programleiar, AMS
Avdelingsleiar, Analyse

Hypotese: Avansert analyse (maskinlæring, statistikk) har stort potensiale!

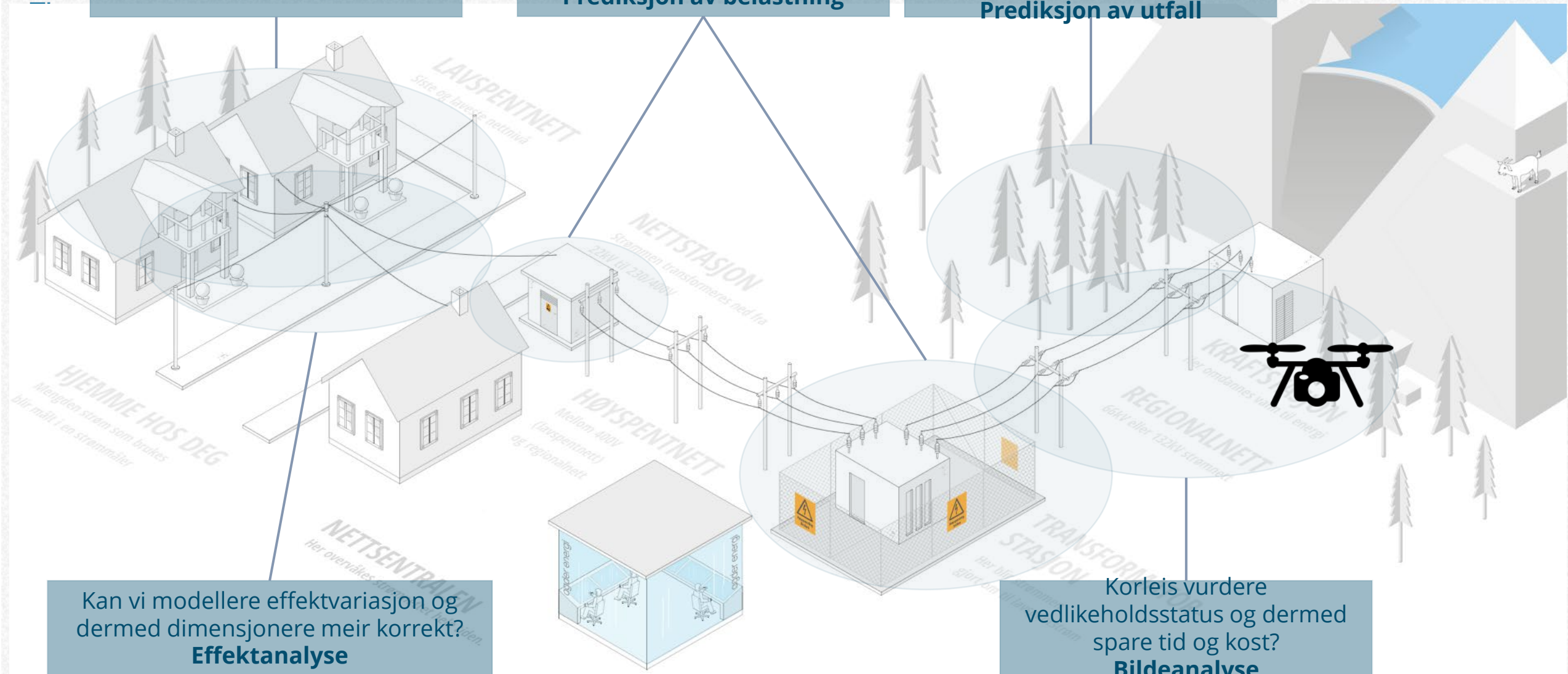
Korleis forstå kunden?
Clustering og klassifisering

Korleis forutsjå belastning på vitale komponentar i nettet?
Prediksjon av belastning

Kan vi forutsjå utfall basert på ver, vegetasjon og utfallshistorikk?
Prediksjon av utfall

Kan vi modellere effektvariasjon og dermed dimensjonere meir korrekt?
Effektanalyse

Korleis vurdere vedlikeholdsstatus og dermed spare tid og kost?
Bildeanalyse





Effektanalyse

Prediksjon av belastning

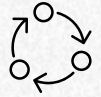
Bildeanalyse

FoU / Pilot

Maskinlæring – Frå pilot til drift



Ide



Valg av metode



Datainnsamling



Datavask



Feature engineering



Trening



Testing



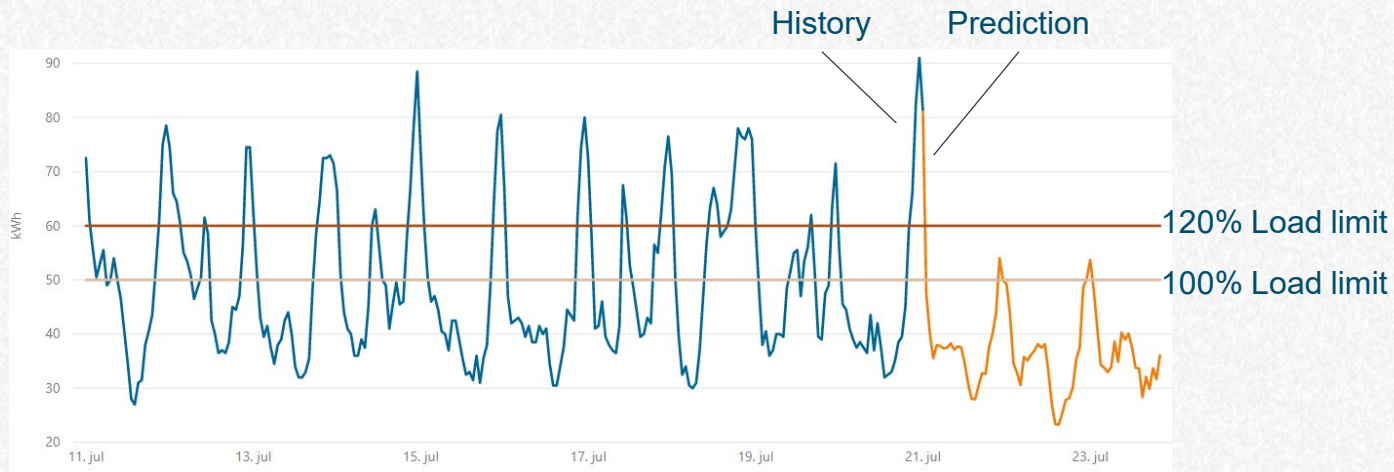
Ferdig testa modell

Kva må til for idriftsetting?

- Solid løysing
 - Automatisk datainnsamling
 - Autonom modell (med minimal overvåkning)
- Skalerbar
- Må vere nyttig (positiv gevinst)
- Implementert i driftsprosess
- Må vedlikeholdast – krev kompetanse

1010
1010
1010
1010





Forecasting load on secondary substations

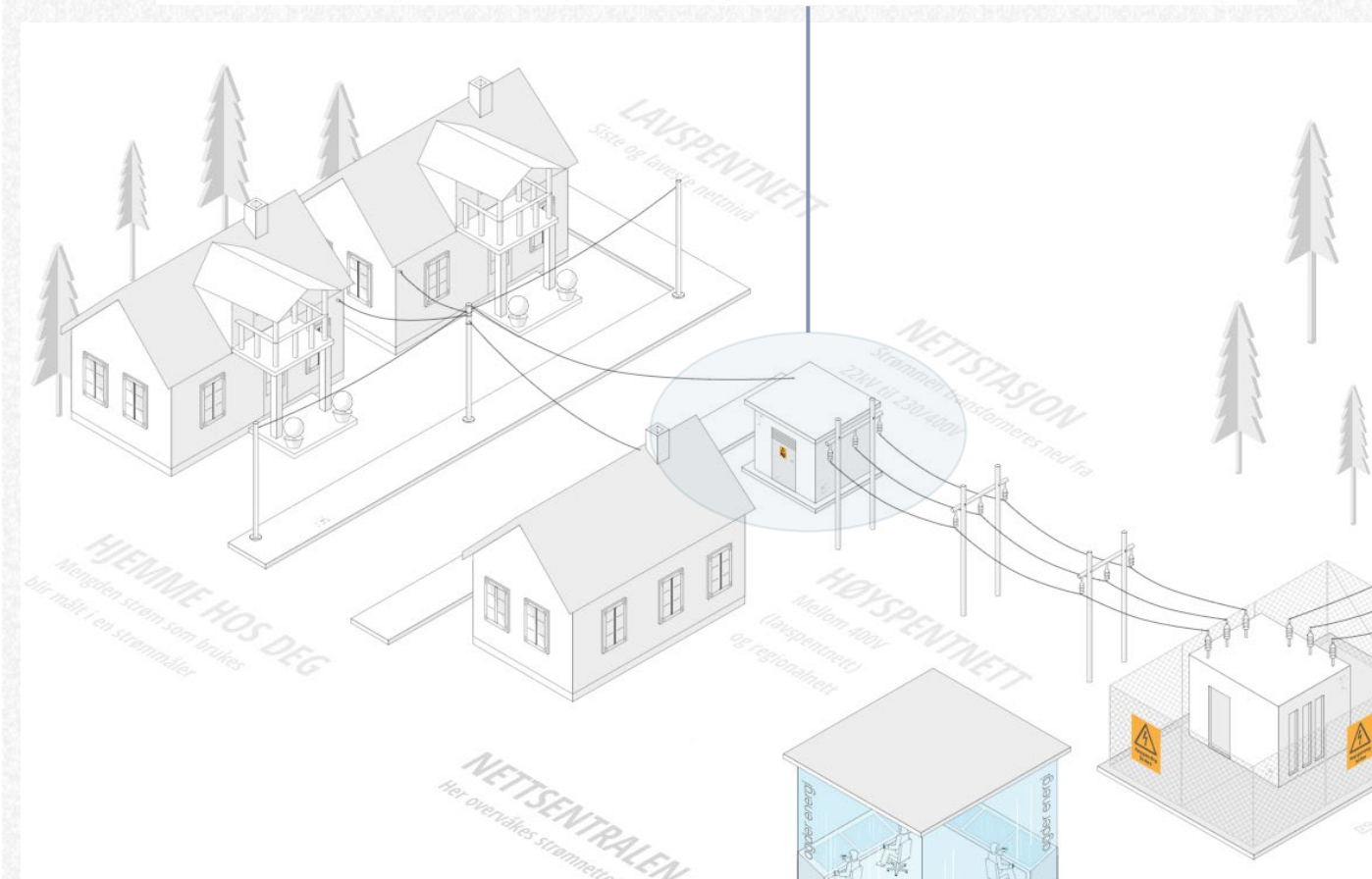
Some substations get overloaded in peak load situations. This may lead to

- Need for maintenance
- Reduced lifetime
- Outage

In sum: Increased cost, reduced grid stability

Load Forecasting can be used to

- Avoid overload using Demand Response (price incentives, customer flexibility)
- Operational readiness
- Repair or upgrade



Substation load

- 26 substations
- Hourly metered load (kWh)
- Collected every 6 hours
- 2-3 years of historical data

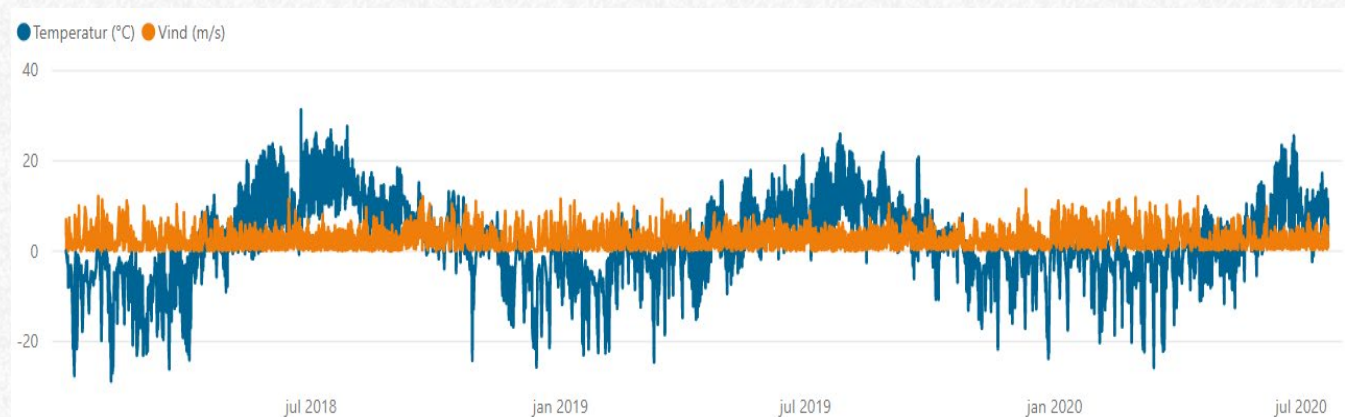
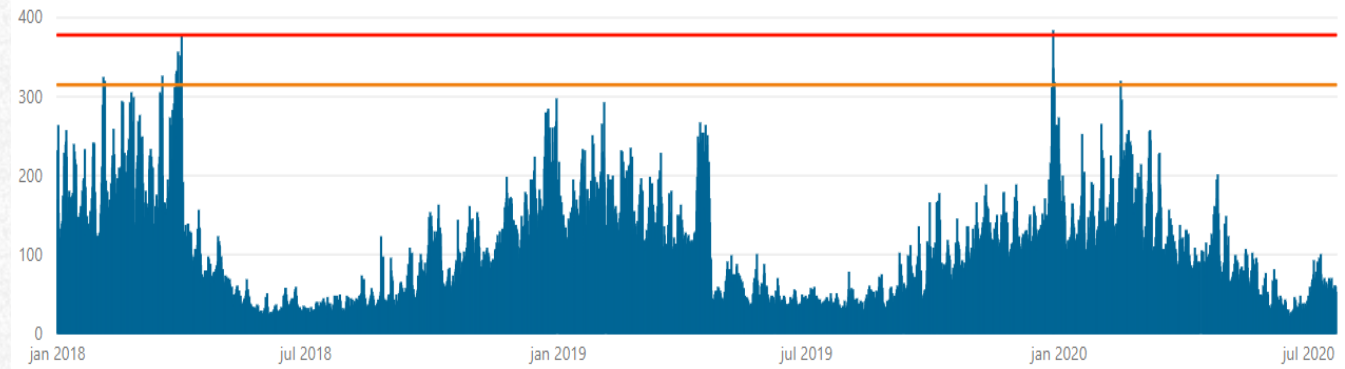
Weather

- Temperature, wind, rain/snow, cloud coverage...
- Source: Norwegian Meteorological Institute
- Location precision: 2,5 km

Pluss derived features ...

Data we did not use (yet):

- Customer classification (residential, industry, vacation dwelling, ...)
- Demographic data
- EV penetration



One common model or several individual models?

A. One common prediction model for all 2ndary substations

- Benefit: Can train one model on all available data
- or

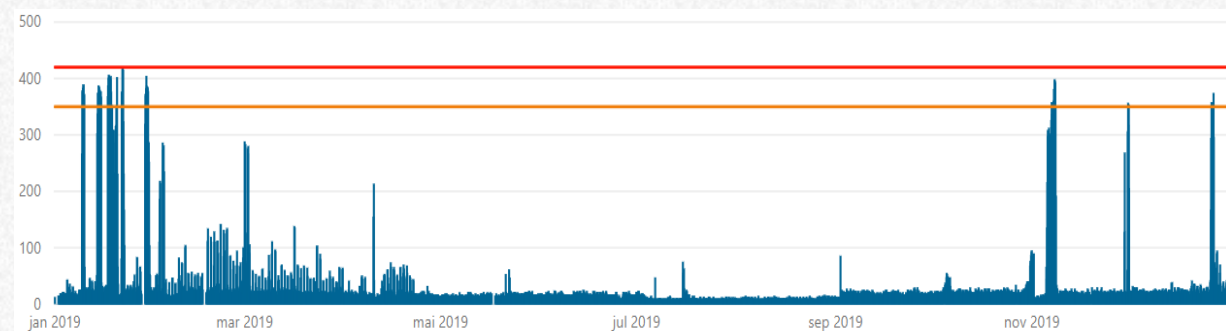
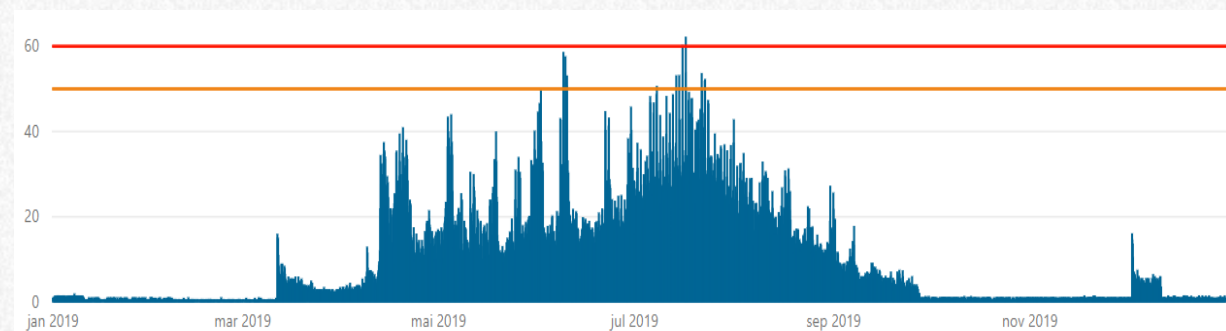
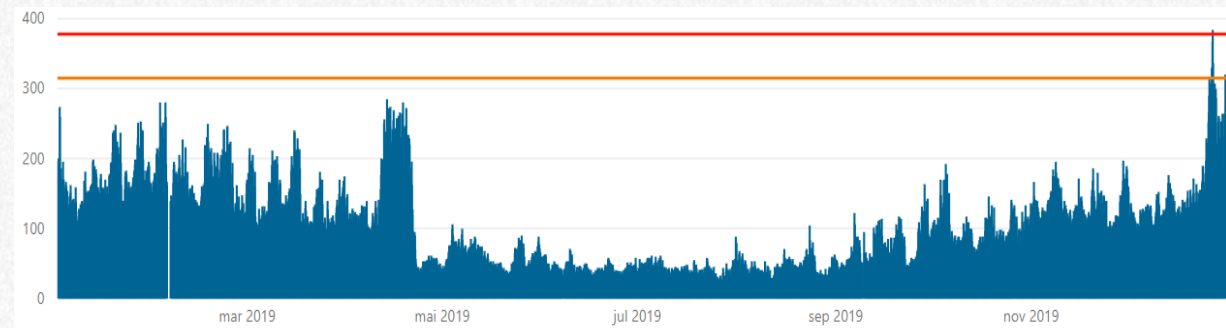
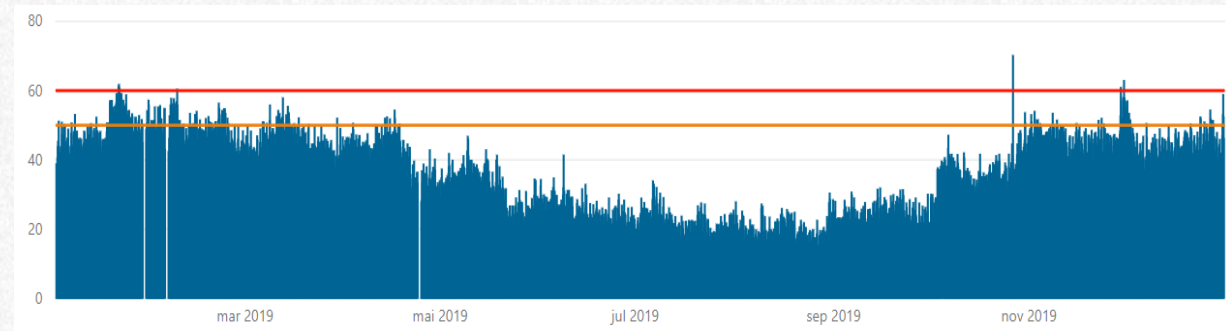
B. Train one individual model for each 2ndary substation

- Benefit: Model better suited for each substation
- or

C. Make clusters of 2ndary substations, one prediction model for each cluster

Conclusion:

- B (individual models) work best in general due to very different load profiles
- A (common model) sometimes gives best result with LGBM
- C (clusters) not evaluated in order to keep the model simple and robust. Potential further work.



ML prediction methods

XGBoost
(individual models)

LGBM
(one common model)

LGBM
(individual models)

LSTM
(dropped)

LSTM requires a long training period, and was dropped early in the process

compared to

Baseline methods

SARIMAX

naive_24h mean

naive_24h median

naive seasonal

naive last period

naive last week

measured with

$$MAPE = \frac{1}{n} \sum_{I=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Substation in a residential area

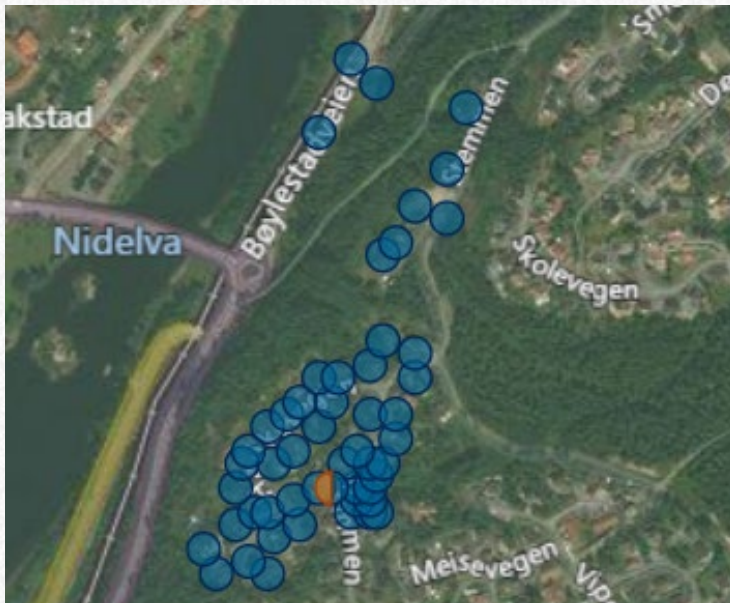
Much of the electricity consumption is depending on temperature. There is a large number of residential customers, and the load variation is fairly stable.

Best model: LGBM (individual models)

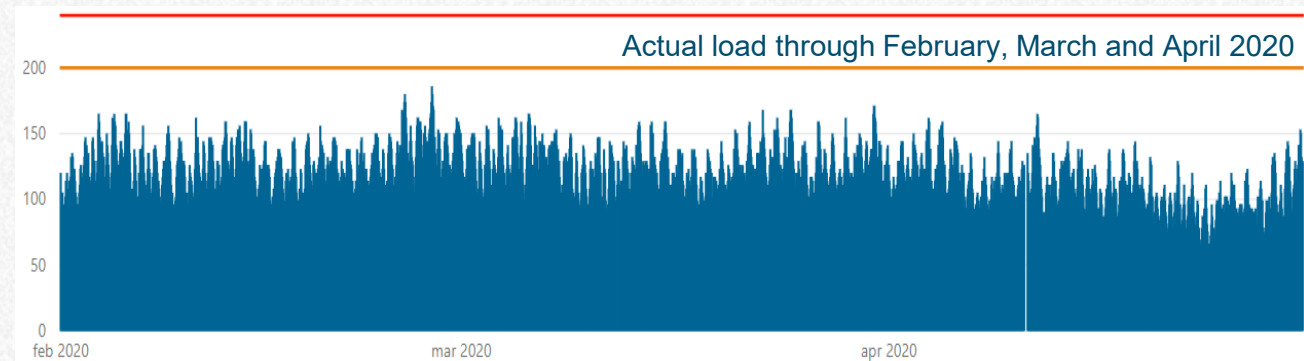


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
7.908	7.746	6.763	12.921	12.597	12.973	13.802	12.595	12.926



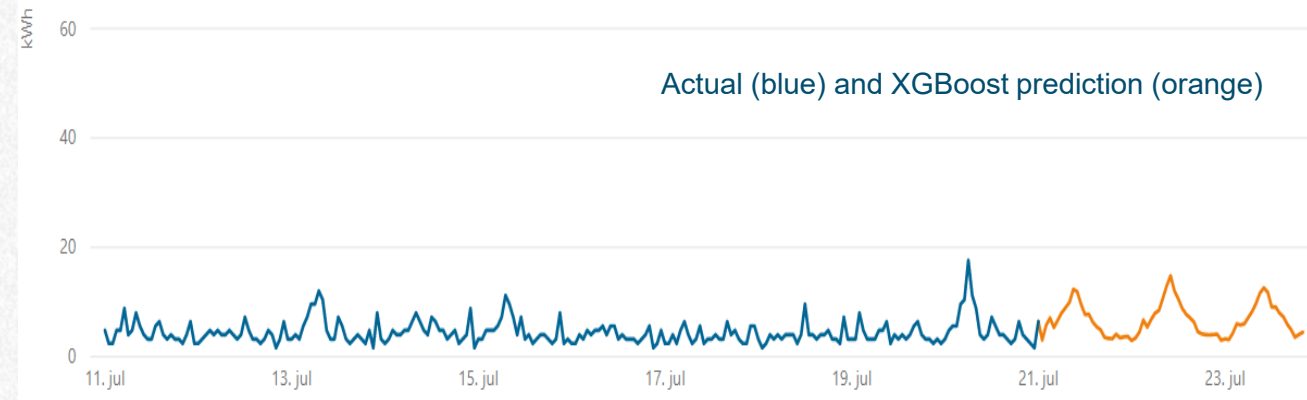
Number of customers: 60
Type of area: Residential



Substation load dominated by school building

The substation serves a school complex. Consumption during the prediction period (Feb-April 2020) was heavily influenced by temperature and the corona situation. In addition there is a period with missing data. Most models perform poorly.

Best model: XGBoost

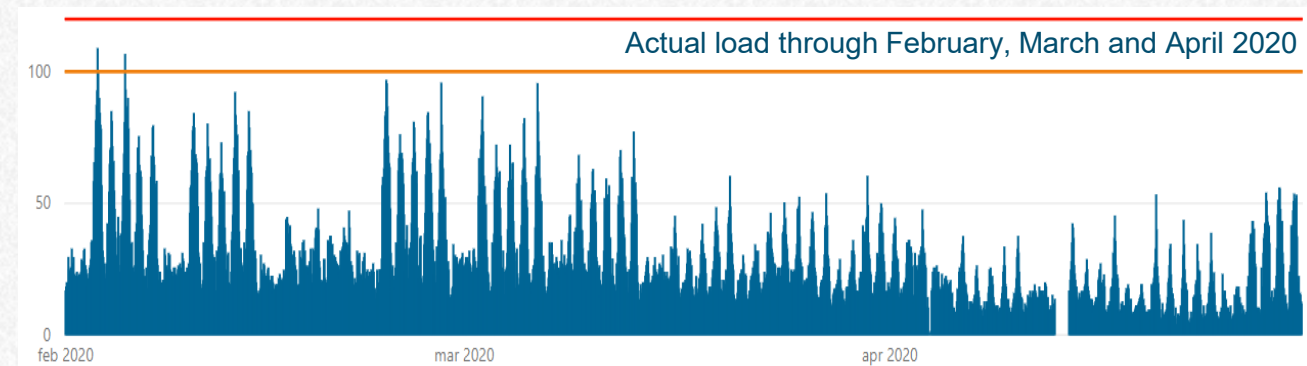


MAPE scores based on prediction for February, March and April 2020

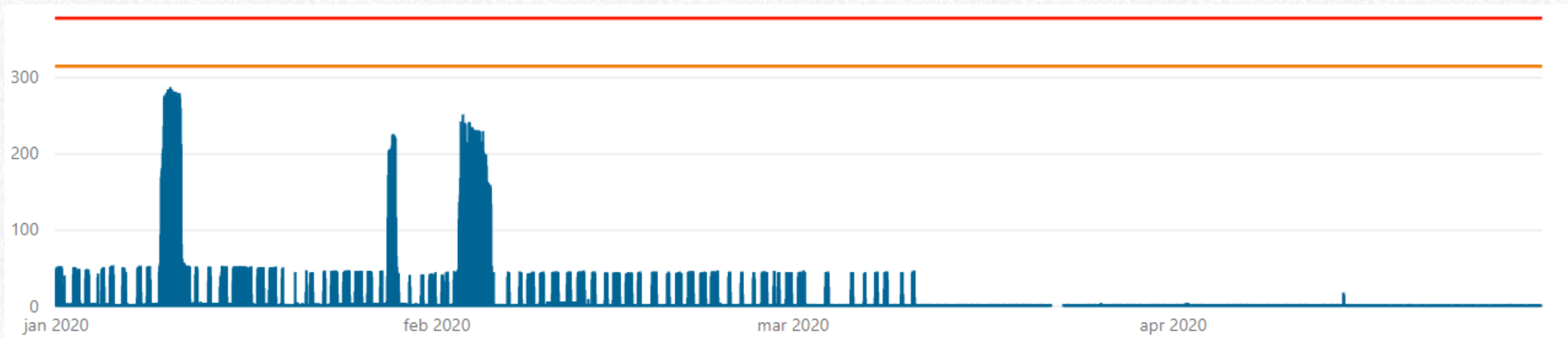
XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
26.144	32.688	33.481	32.509	46.281	42.475	58.008	40.542	37.976



Number of customers: 6
Type of area: School.



Some things are hard to predict ...



Summary: Which forecasting model is best?

There is no «one model best in all cases».

LGBM points out as the model with best score (lowest error measure) most often.

LGBM with **individual models** (i.e. model trained individually for each substation load profile) emerges as the best option.

SARIMAX and the naive models perform well in cases where the load is repetitive.

Count of «best model score» achieved

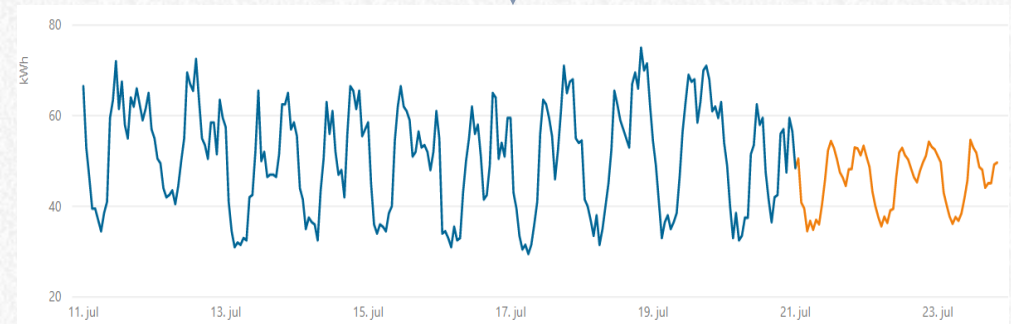
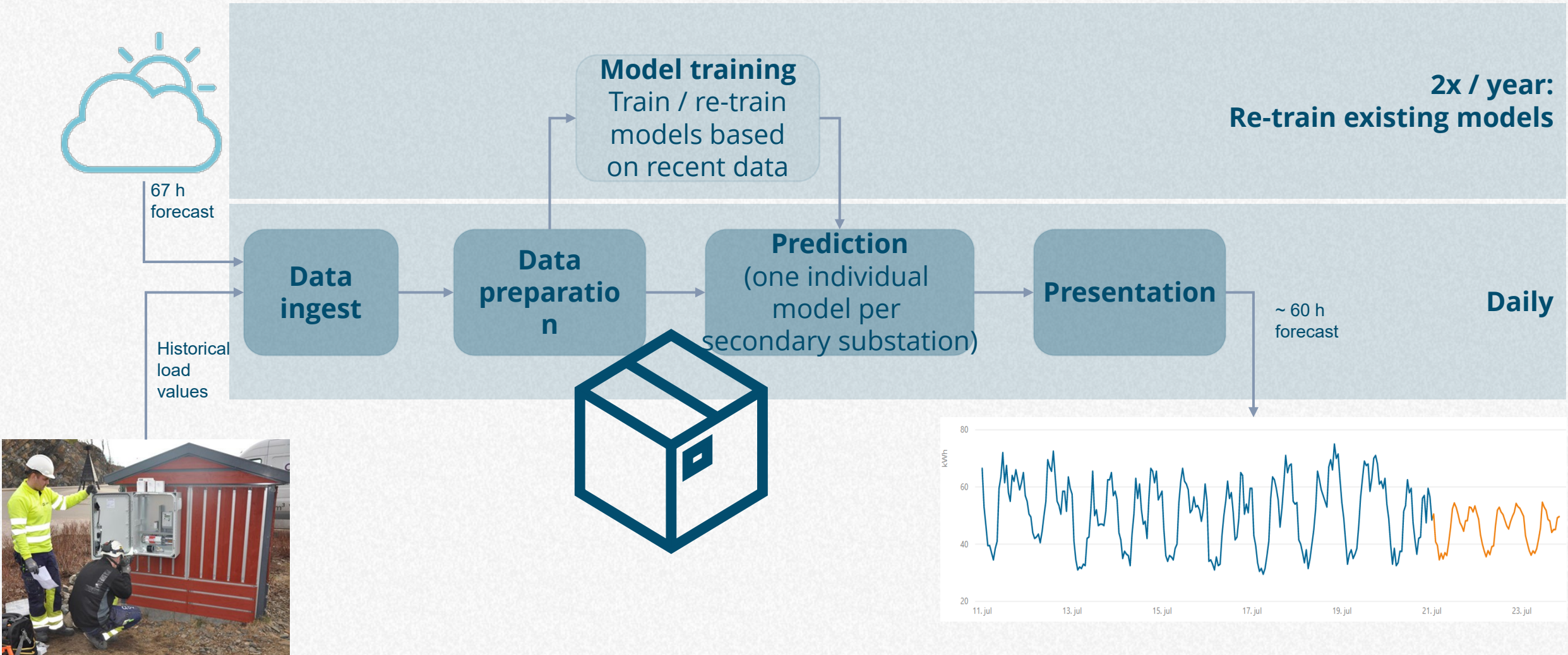
Model score (error measure)	XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
RMSE	2	6	12	2	2	0	0	1	0
MAPE	1	7	12	1	0	2	0	2	0

Overgang til drift

Fully automated ML pipeline in production

Currently predicting load for 120+ secondary substations

Model based on XGBoost





1. Søk etter nettstasjonstrafo

Trafokretsld

Alle

Nettstasjonsnavn

Alle

Status (belastning)

Flere valg

Status (balanse)

Flere valg

Kategori (balanse)

Alle

↶ Nullstill alle valg

2. Velg en nettstasjonstrafo (høyreklikk i en av tabellene og velg **ekstraher**)

Trafokretsld	Nettstasjonsnavn	Konsentrator	Prediksjon	Målepunktld	Avgang
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Ja	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		
		Nei	Ja		

Vis som tabell
 Inkluder
 Ekskluder
 Ekstraher
 Kopier

Nettstasjonsinfo
 Belastning og Værdata
 Balanse
 Stør, Spenning, Cos φ
 Hendelser og Alarmer
Prediksjon
 Momentanspørring

Trafokretser med høyest belastning (siste år)

Trafo	Nettstasjonsnavn	Tid siste maksverdi	Overlast	Timer	Merkeyt.	Status
		20.07.2020 23:00:00	82,33 %	308	50,00	Ubehandlet
		06.07.2020 10:00:00	70,07 %	212	30,00	Ubehandlet
		22.05.2020 19:00:00	57,55 %	335	50,00	Overbelastet
		23.11.2019 18:00:00	39,14 %	84	30,00	Ubehandlet
		26.12.2019 14:00:00	37,09 %	198	50,00	Ubehandlet
		30.04.2020 23:00:00	35,00 %	116	50,00	Ubehandlet
		02.12.2019 19:00:00	33,38 %	186	50,00	Overbelastet
		01.12.2019 13:00:00	31,10 %	551	50,00	Ubehandlet
		27.12.2019 17:00:00	31,01 %	170	100,00	Ubehandlet
		24.12.2019 18:00:00	30,06 %	144	50,00	Ubehandlet
		03.02.2020 09:00:00	29,33 %	56	350,00	Ubehandlet
		02.12.2019 20:00:00	28,57 %	1011	50,00	Ubehandlet
		14.07.2020 20:00:00	26,25 %	33	50,00	Ubehandlet
		30.11.2019 18:00:00	26,15 %	291	50,00	Ubehandlet
		19.07.2020 14:00:00	26,05 %	36	50,00	Ubehandlet

Trafokretser med størst ubalanse (siste 3 mnd)

Trafo	Avvik (%)	Avvik (kWh)	Kategori	Status
	14271,26 %	39 959,52	K > T	Ubehandlet
	8454,04 %	14 371,86	K > T	Ubehandlet
	7190,46 %	31 242,55	K > T	Ubehandlet
	5461,67 %	352 933,10	K > T	Ubehandlet
	3448,83 %	52 680,91	K > T	Ubehandlet
	3232,86 %	111 533,81	K > T	Ubehandlet
	2885,60 %	26 042,58	K > T	Ubehandlet
	2875,95 %	20 031,01	K > T	Ubehandlet
	888,15 %	2 926,46	K > T	Ubehandlet
	709,71 %	72 922,20	K > T	Ubehandlet
	421,25 %	300 350,38	T > K	Ubehandlet
	399,53 %	81 900,09	K > T	Ubehandlet
	278,41 %	133 940,77	K > T	Ubehandlet
	258,54 %	117 724,67	K > T	Ubehandlet
	198,96 %	5 135,47	K > T	Ubehandlet

Machine Learning - assessing the potential of machine learning and building out experimental applications to aid carbon reduction initiatives that stimulate the development of smart cities

The potential of ML for Smart Grids

AI&ML for the Smart Grid 2020 – 8-10 Sept. 2020

Per-Oddvar Osland, Agder Energi Nett
Program Manager, Smart Metering
Dept Manager, Analytics

Agenda

Introduction

Application of Machine Learning

Summary

One of the largest energy companies in Norway
Based in southern Norway, operations in 4 countries
A strong focus on the value chain

*) EMT – Energy Management & Trading



Agder Energi Nett

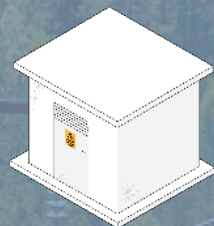
Task: Build, operate and maintain the electricity grid



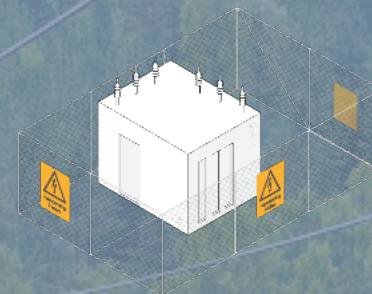
204,000
customers



21,900 km
lines and cables



8,300
distribution
substations



78
transmission
substations

Can AI/ML assist in solving key problems? Need data, tools and competence!



Voltage quality

Discover areas / periods with capacity constraints. Choose the right operational means to meet challenges. Better planning when building / extending grid.



Energy and power balance

Discover loss in low voltage grid – technical vs non-technical loss.

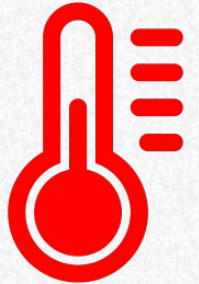
Ground fault/interruptions

Quick and cost-effective failure correction.
Improved safety.



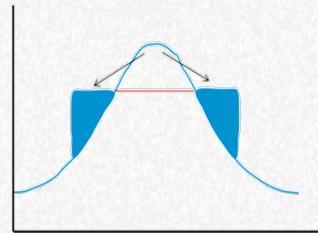
Substation surveillance

Temperature, moisture, switch position, open doors, ...



Demand response

Motivate customers to distribute grid load.
Results: Better power quality and grid stability, reduced infrastructure investments.



Data from Smart Metering roll out (2015-2018)

Currently collecting 200 M values daily

460 - Regional grid

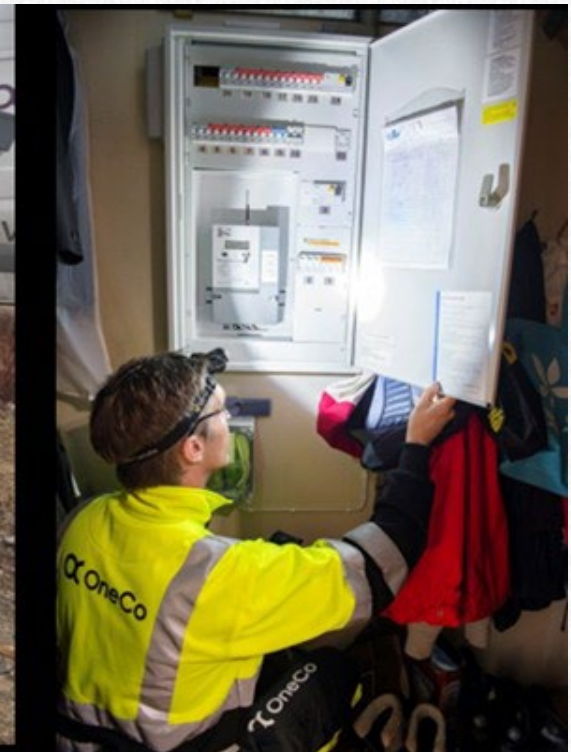
Substations, production plants, and large customers (High Voltage)
El. Meter: Cewe prometer 100
Communication: Fixed broadband or Mobile (4G)

8300 – Secondary Substations

El. Meter: Kamstrup Ominpower
Multi-instrument IME Nemo D4-Le
Ground fault sensor
Additional I/O ports for sensors
Communication: Mobile (3G/2G)

200 000 - Customer premises

El. Meter: Kamstrup Ominpower
Communication: Radio Mesh



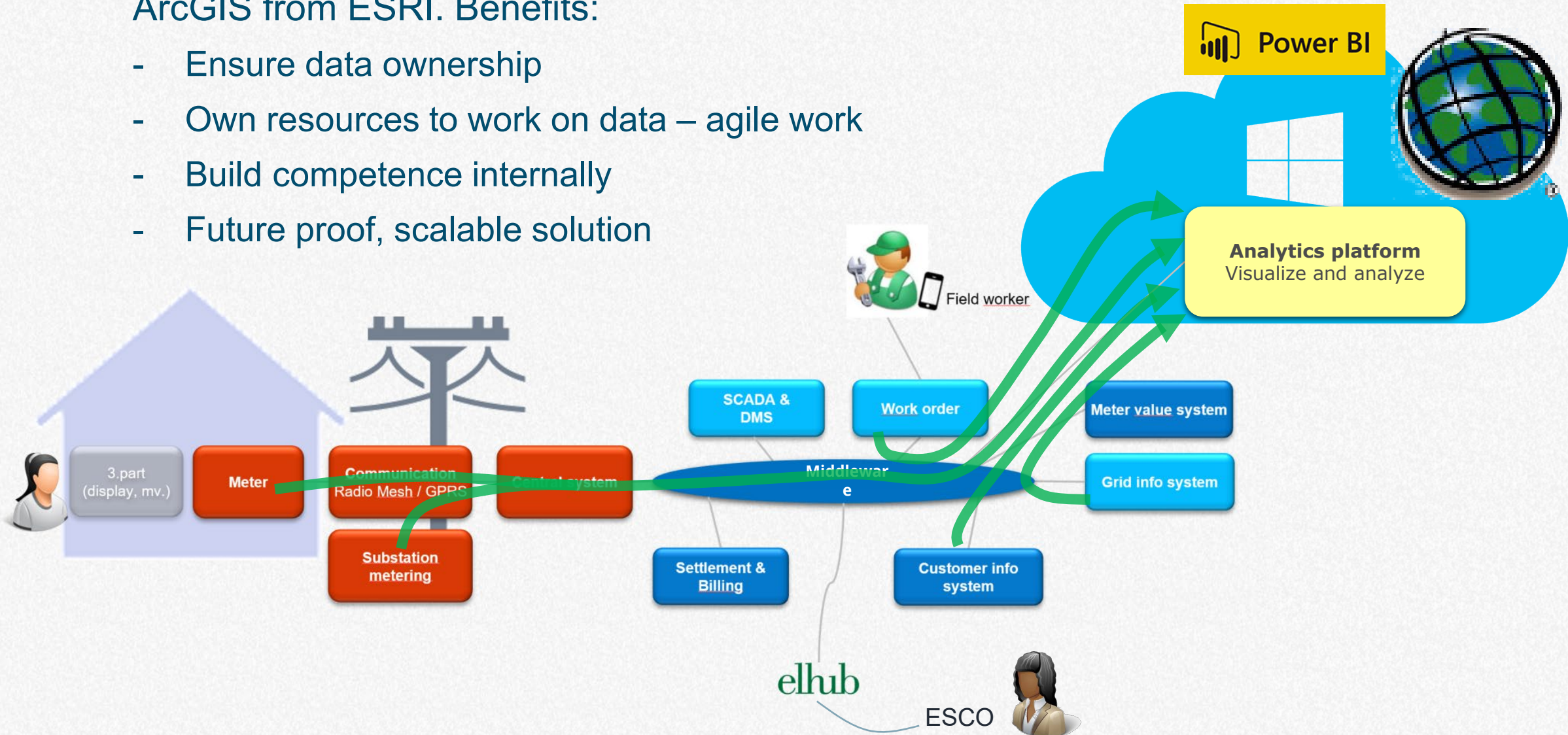


Data from Grid Inspection
~ 0,5 – 6 M images/year



Self-developed cloud-based solution based on MS Azure and ArcGIS from ESRI. Benefits:

- Ensure data ownership
- Own resources to work on data – agile work
- Build competence internally
- Future proof, scalable solution



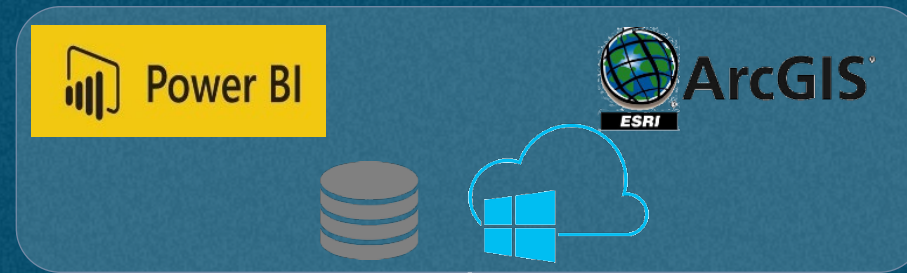
Competence – success criteria

Competence

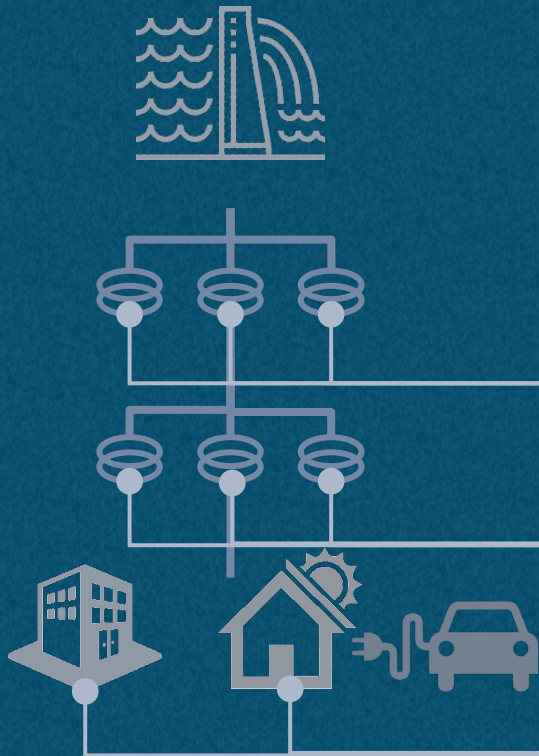
Domain Expert



Data Scientist



Tools



Data

Agenda

Introduction

Application of Machine Learning

Summary

Hypotesis: ML has a great potential for utilities

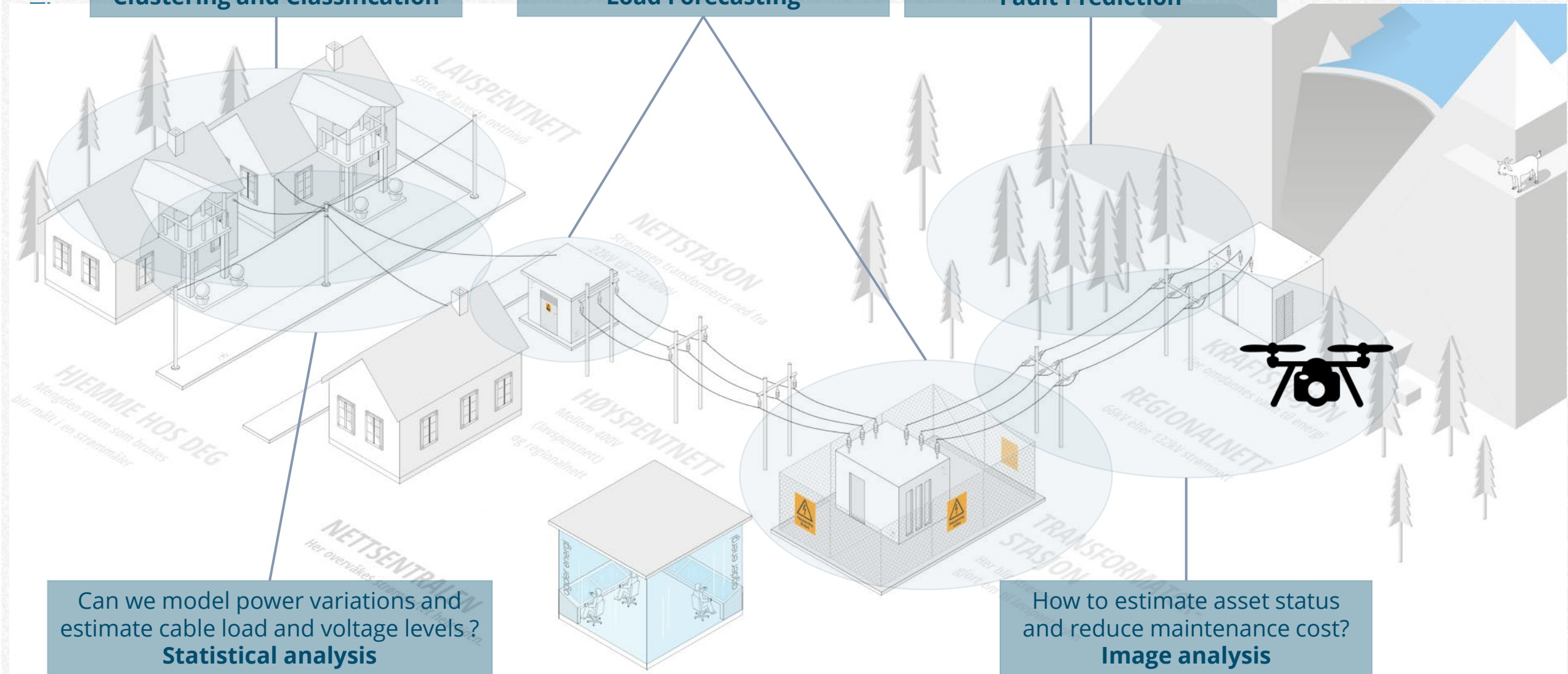
How can we better understand customer behavior?
Clustering and Classification

How do we predict load on vital grid components?
Load Forecasting

Can we predict outages based on weater and vegetation?
Fault Prediction

Can we model power variations and estimate cable load and voltage levels?
Statistical analysis

How to estimate asset status and reduce maintenance cost?
Image analysis



The logo for CINELDI, where 'CIN' is in dark blue, 'E' is a stylized green lightning bolt, and 'LDI' is in dark blue.

Centre for Intelligent Electricity Distribution
www.cineldi.no

How can we better understand
customer behavior?
Clustering and Classification

How do we predict load on vital
grid components?
Load Forecasting

Can we predict outages based
on weater and vegetation?
Fault Prediction

Can we model power variations and
estimate cable load and voltage levels?
Statistical analysis

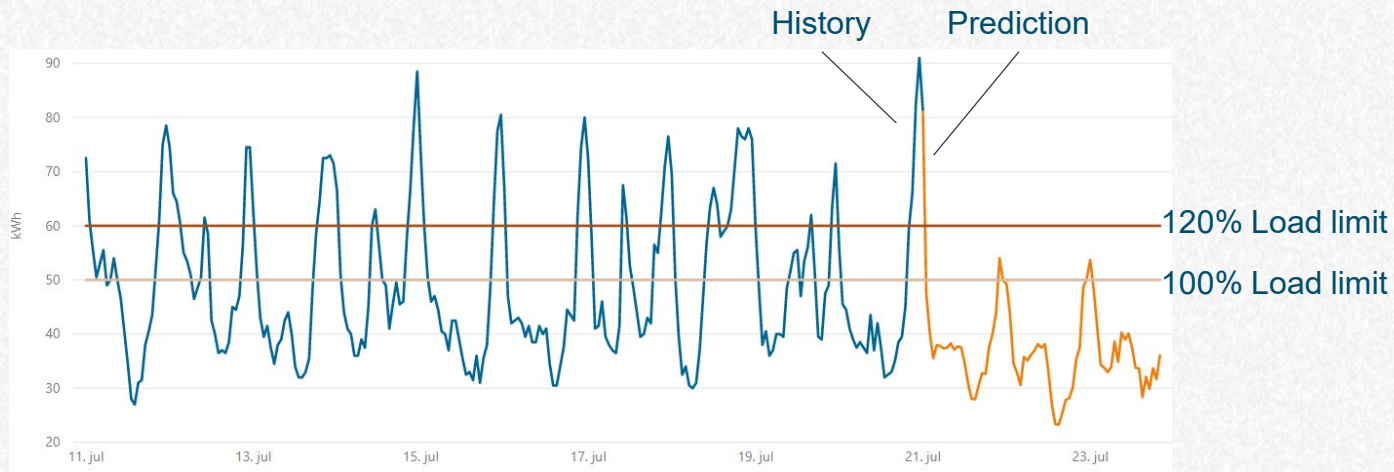
How to estimate asset status
and reduce maintenance cost?
Image analysis

We focus on

- Key principles, methods used
- What works, what does not (and why)
- Benefits and issues

This presentation does not contain

- A lesson in Machine Learning
- Details on technology
- Code



Forecasting load on secondary substations

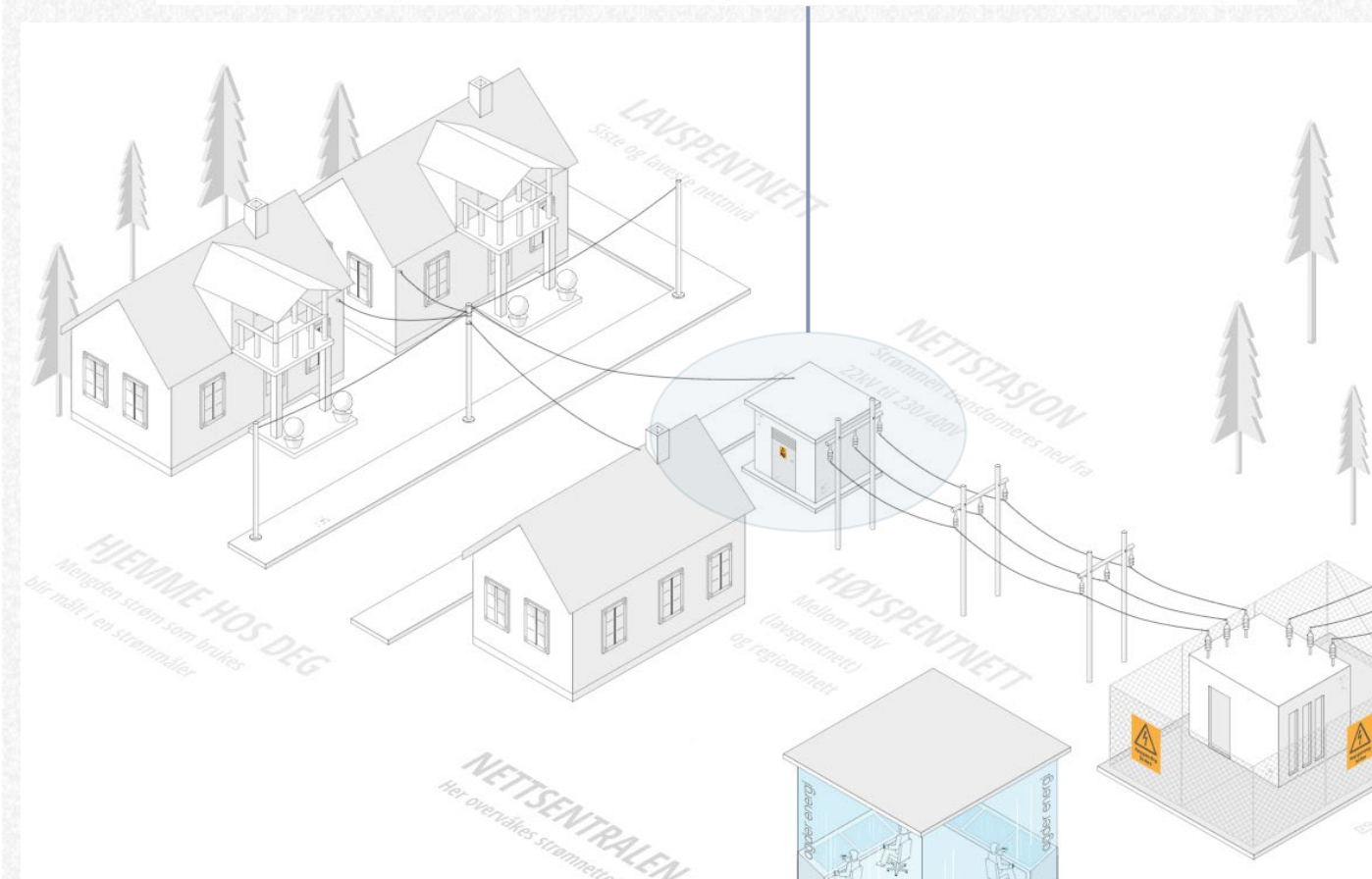
Some substations get overloaded in peak load situations. This may lead to

- Need for maintenance
- Reduced lifetime
- Outage

In sum: Increased cost, reduced grid stability

Load Forecasting can be used to

- Avoid overload using Demand Response (price incentives, customer flexibility)
- Operational readiness
- Repair or upgrade



Substation load

- 26 substations
- Hourly metered load (kWh)
- Collected every 6 hours
- 2-3 years of historical data

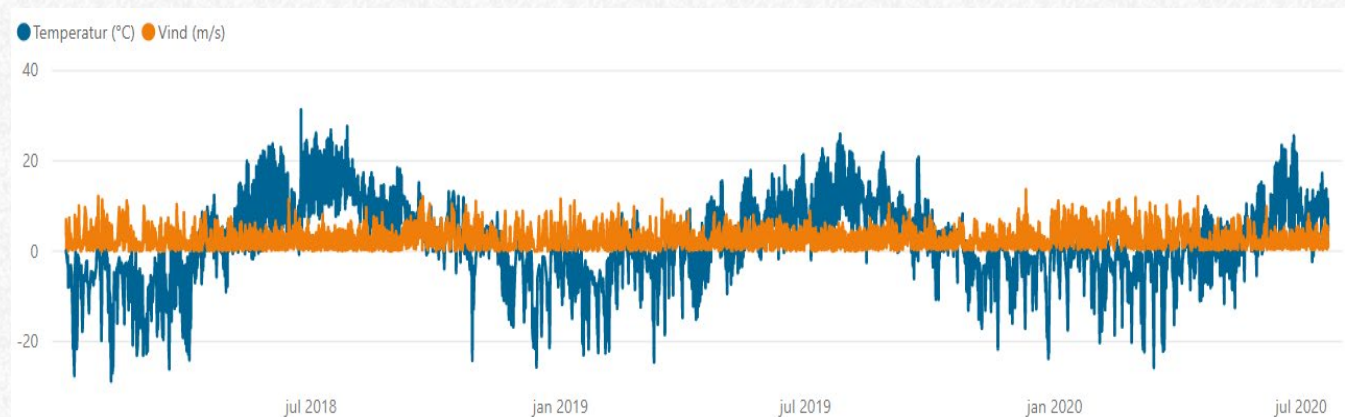
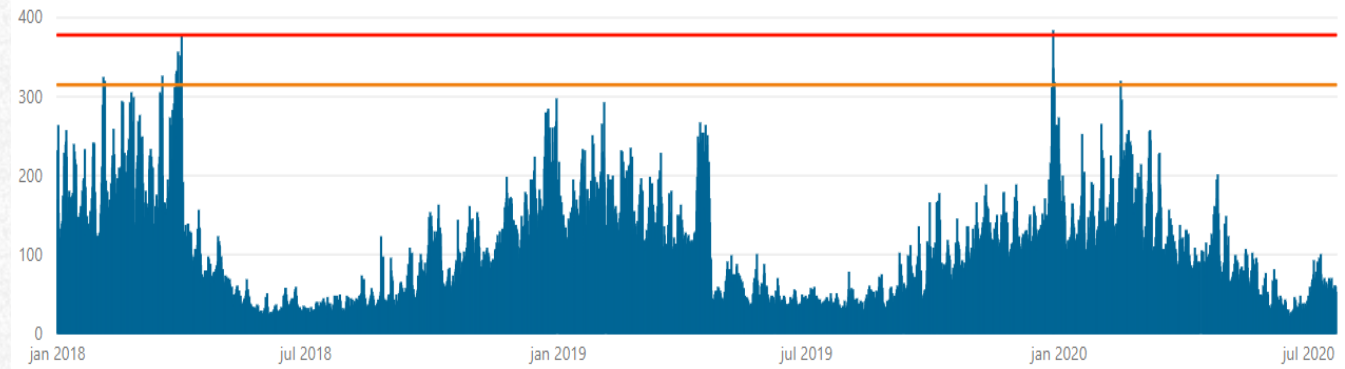
Weather

- Temperature, wind, rain/snow, cloud coverage...
- Source: Norwegian Meteorological Institute
- Location precision: 2,5 km

Pluss derived features ...

Data we did not use (yet):

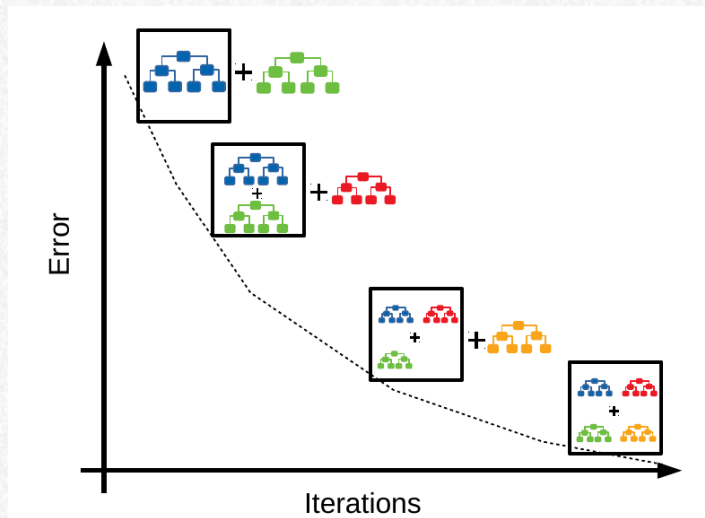
- Customer classification (residential, industry, vacation dwelling, ...)
- Demographic data
- EV penetration



Gradient Boost Trees

XGBoost - Extreme Gradient Boost

LGBM – Light Gradient Boosting Machine

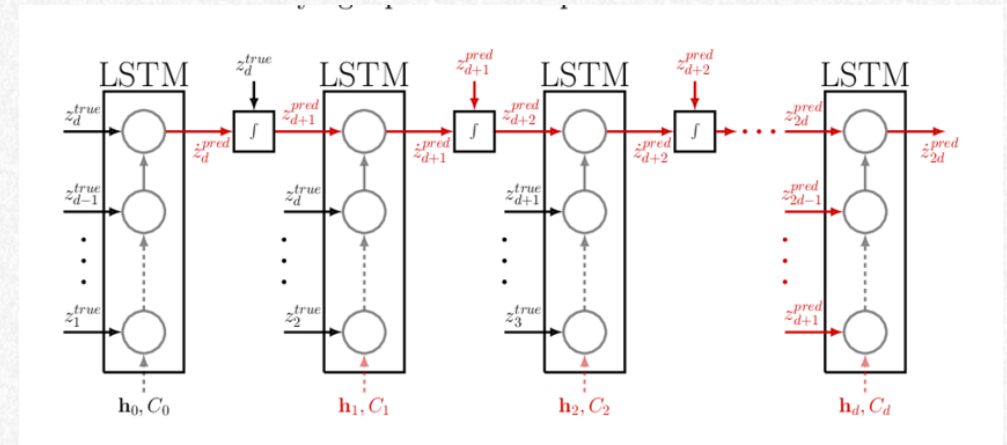


Gradient Boost Trees are an ensemble of trees. At each iteration a tree is added, trained on the *gradient of the errors* from the previous iteration.

Source: Theodore Vasiloudis (<http://tvas.me/>)

Recurrent Neural Network

LSTM - Long short-term memory



Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video).

Source: Wikipedia, Pantelis Rafail Vlachas

One common model or several individual models?

A. One common prediction model for all 2ndary substations

- Benefit: Can train one model on all available data
- or

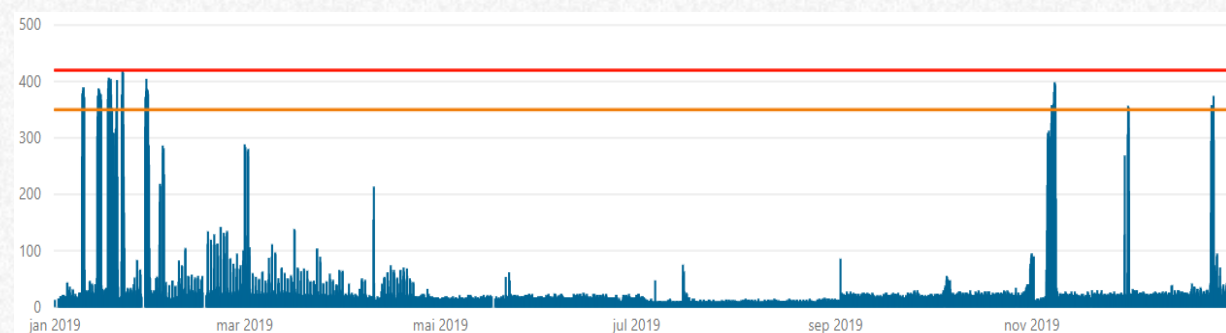
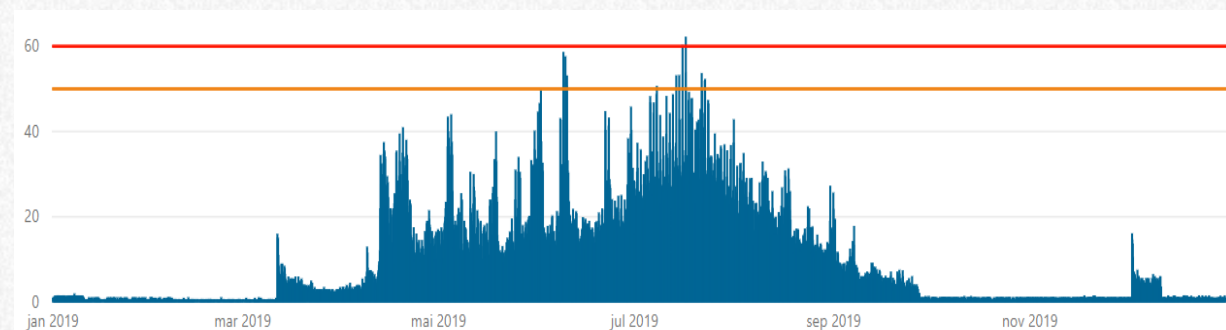
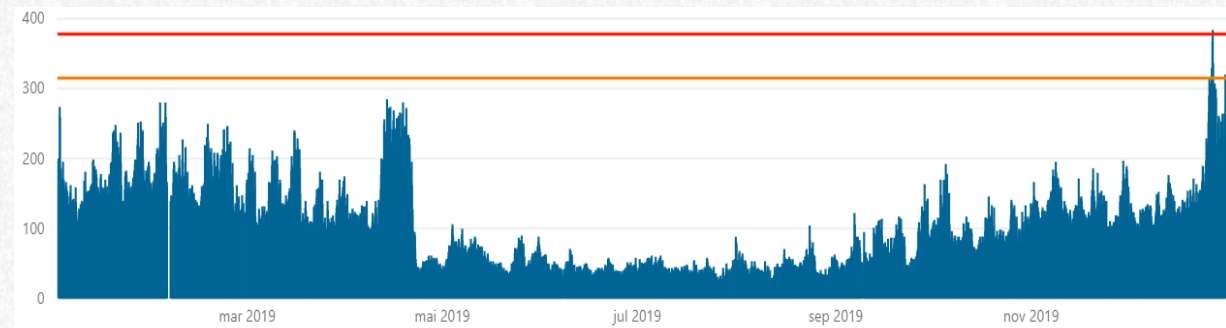
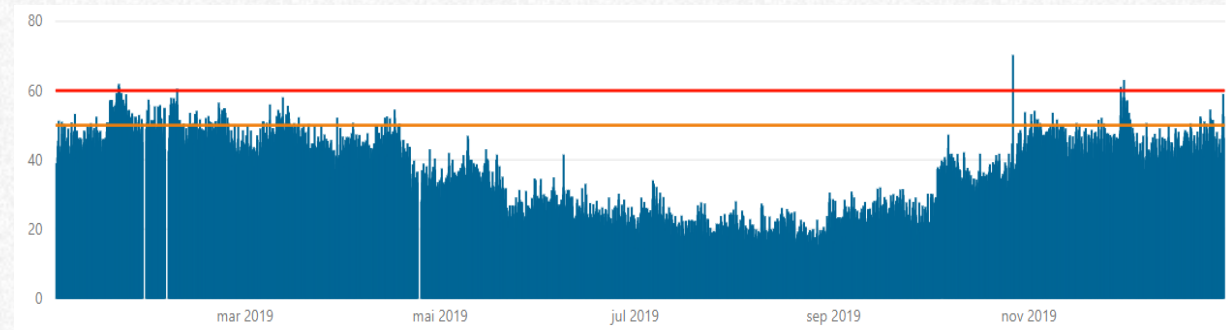
B. Train one individual model for each 2ndary substation

- Benefit: Model better suited for each substation
- or

C. Make clusters of 2ndary substations, one prediction model for each cluster

Conclusion:

- B (individual models) work best in general due to very different load profiles
- A (common model) sometimes gives best result with LGBM
- C (clusters) not evaluated in order to keep the model simple and robust. Potential further work.



Statistical method:

SARIMAX – Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors. SARIMAX allows for seasonal variation and external variables and is an extension of the well-known ARIMA method.

Naïve methods:

Methods based on guessing the forecasted load to be the same as the observed load some time in the past:

naive_24h_mean – Average over the 24 past hours

naive_24h_median – Median over the 24 past hours

naive_seasonal – Same value as observed last year

naive_last_period – Repeated values from the past 3 days

naive_last_week – Same value as observed last year

RMSE - Root Mean Square Error

MAPE - Mean Absolute Percentage Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where y_i = actual load value at hour i , \hat{y}_i = predicted value, n = number of hours

We prefer using **MAPE** as it allows for comparison between time series where the load values (the y_i 's) vary in size.

RMSE and MAPE are calculated over a 80 day period during February – April 2020 for 26 substations.

ML prediction methods

XGBoost
(individual models)**LGBM**
(one common model)**LGBM**
(individual models)LSTM
(dropped)

LSTM requires a long training period, and was dropped early in the process

compared to

Baseline methods

SARIMAXnaive_24h
meannaive_24h
mediannaive
seasonalnaive last
periodnaive last
week

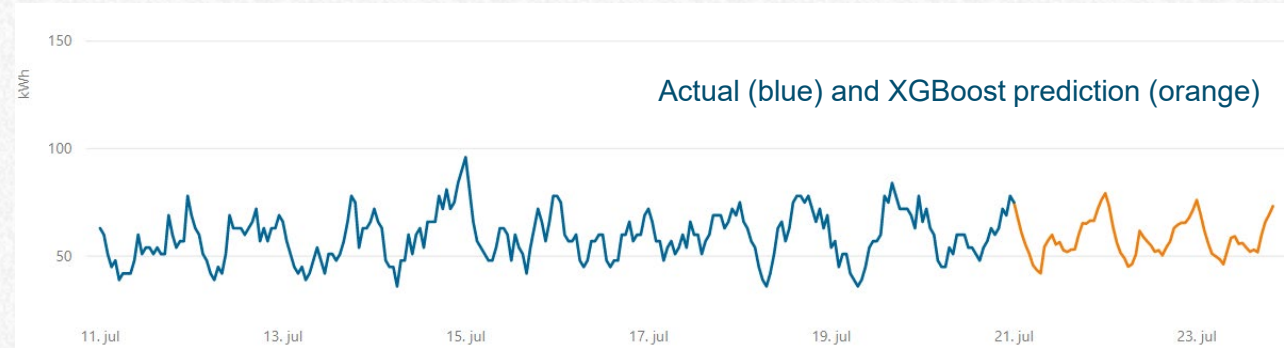
measured with

$$MAPE = \frac{1}{n} \sum_{I=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Much of the electricity consumption is depending on temperature. There is a large number of residential customers, and the load variation is fairly stable.

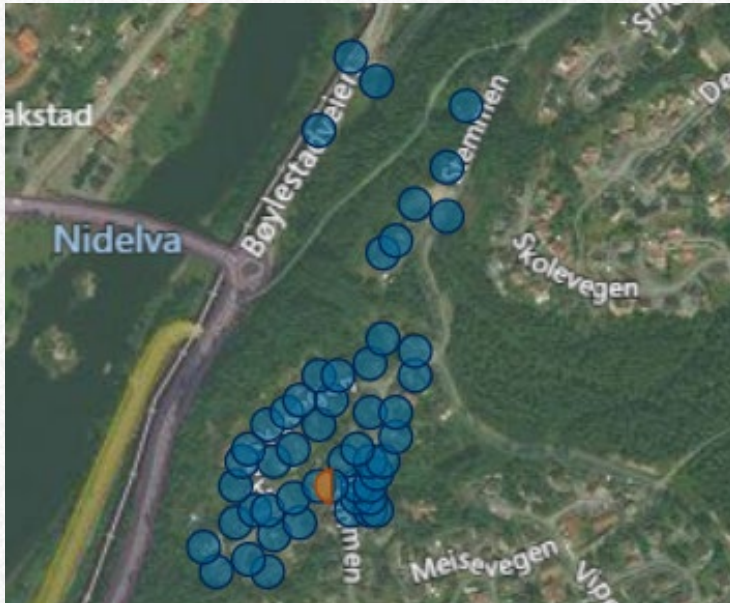
Best model: LGBM (individual models)

Example 1: Substation in a residential area

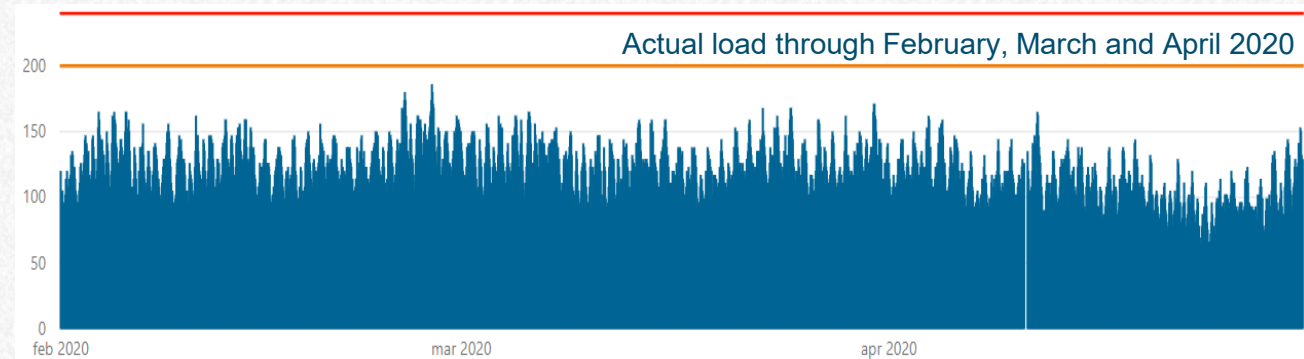


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
7.908	7.746	6.763	12.921	12.597	12.973	13.802	12.595	12.926



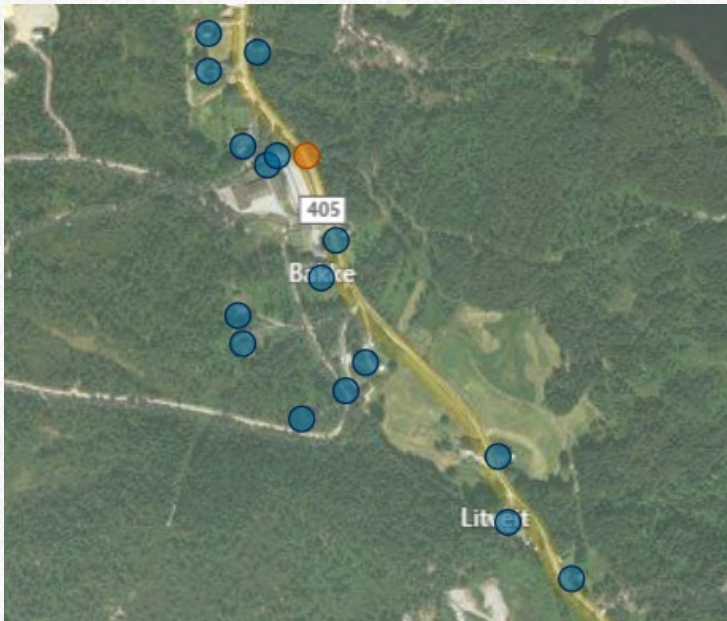
Number of customers: 60
Type of area: Residential



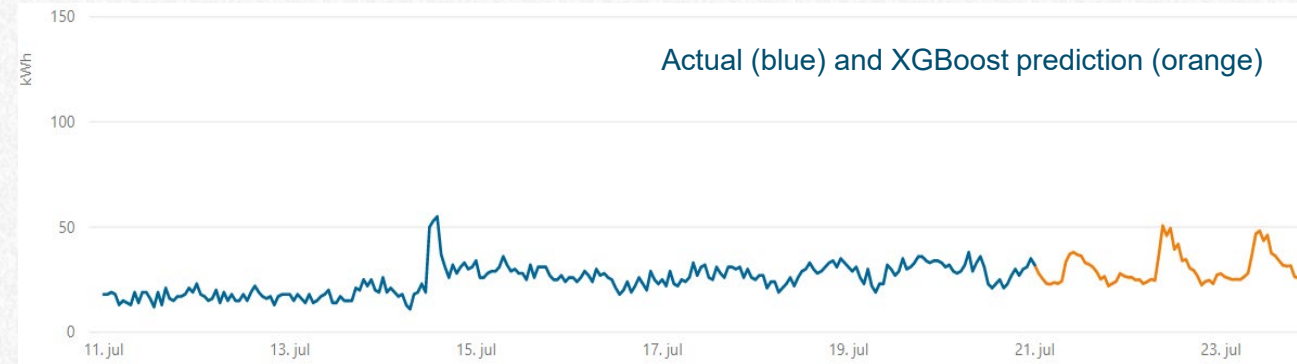
Example 2: Substation in a rural area

A school building dominates the consumption in the area. This gives a clear seasonal load profile, and simple models such as SARIMAX and “naïve last week” perform well.

Best model: SARIMAX

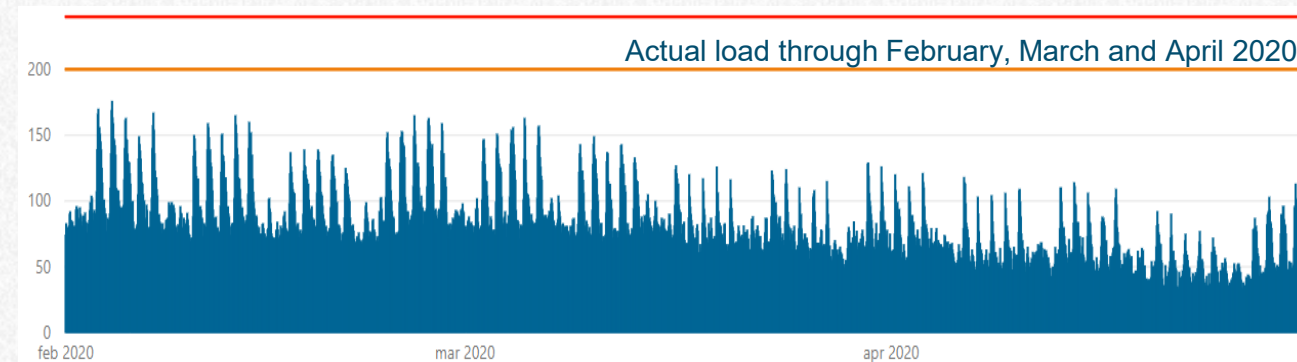


Number of customers: 16
Type of area: Dwellings, school, small industry



MAPE scores based on prediction for February, March and April 2020

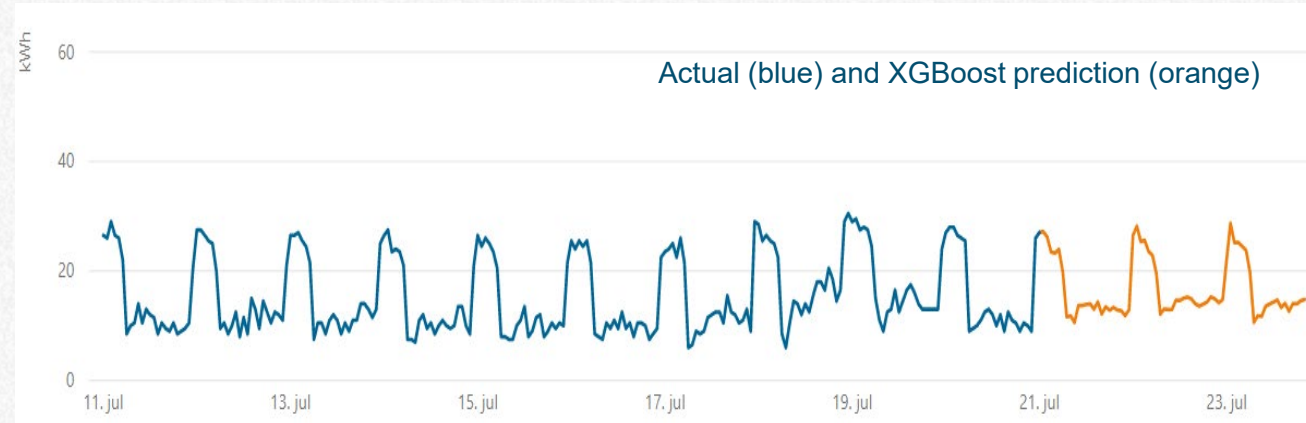
XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
15.395	14.282	15.31	13.088	21.62	19.576	30.433	20.73	16.119



Example 3: Substation in rural area with periodic load

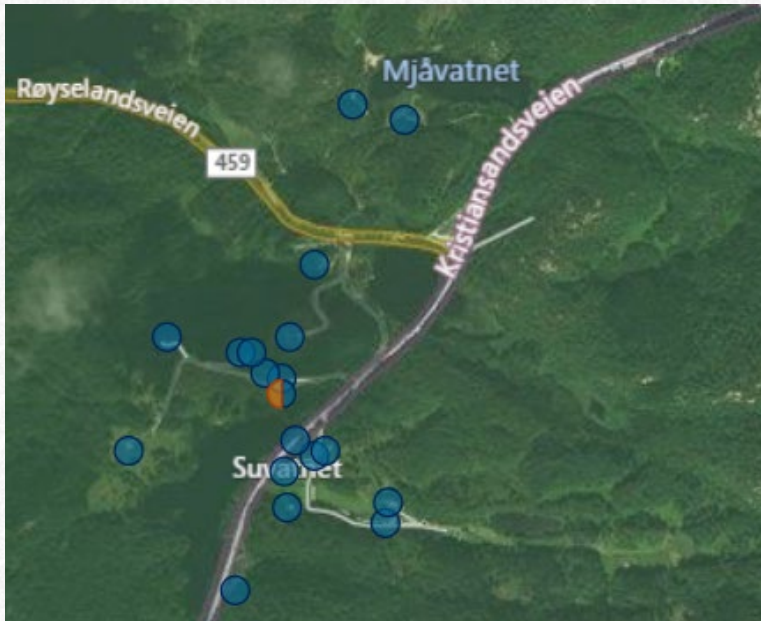
The substation load is very regular and periodic, and simple models such as SARIMAX and “naïve last period” perform well.

Best model: Naïve last period



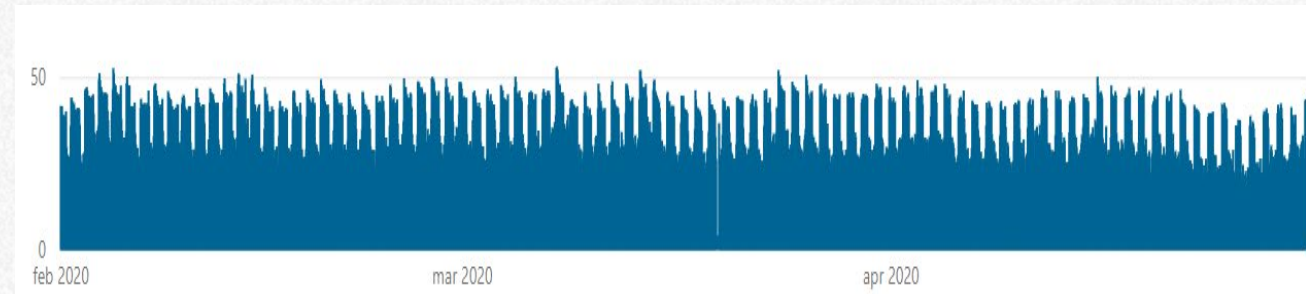
MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
10.609	9.883	9.22	9.131	20.236	19.435	16.057	8.925	10.268



Number of customers: 19
 Type of area: Dwellings, small industry

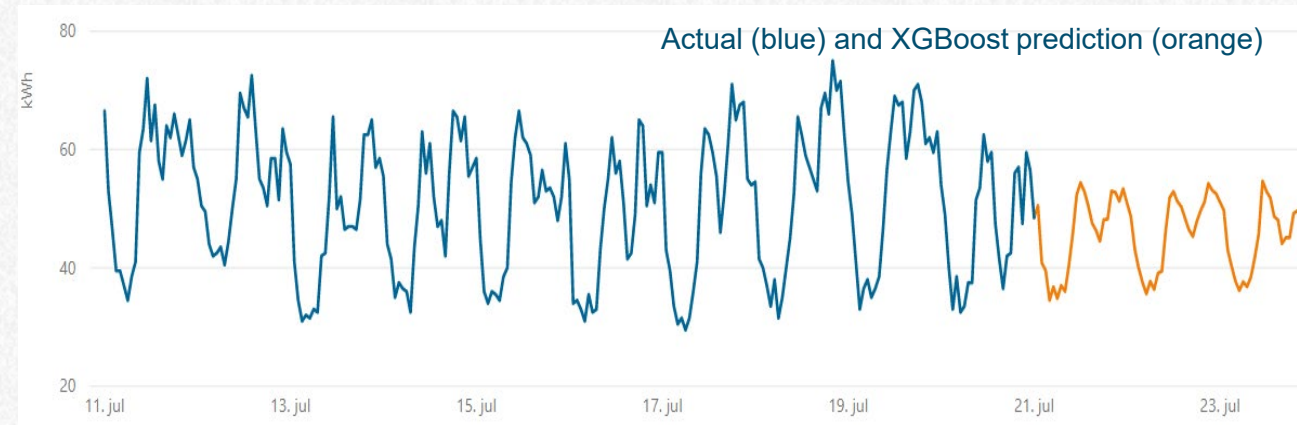
Actual load through February, March and April 2020



Example 4: Residential area

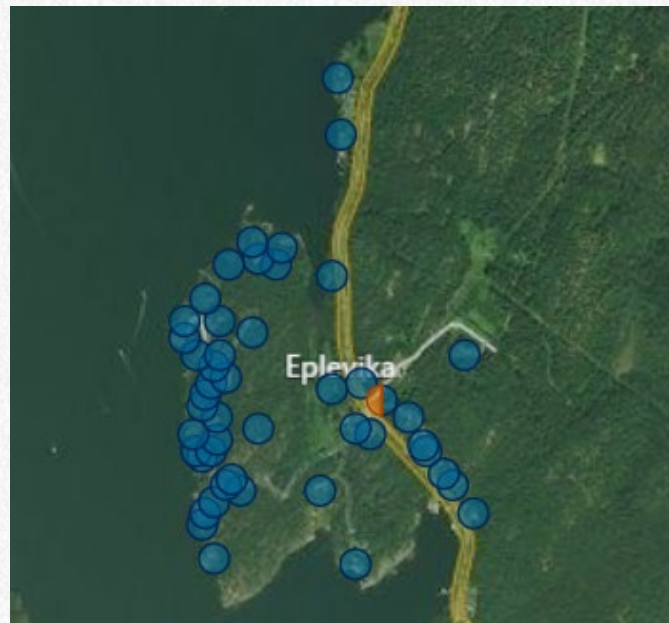
Residential area with stable, temperature-dependent consumption. ML-methods perform best, with some of the naïve methods close behind.

Best model: LGBM (based on one common model)

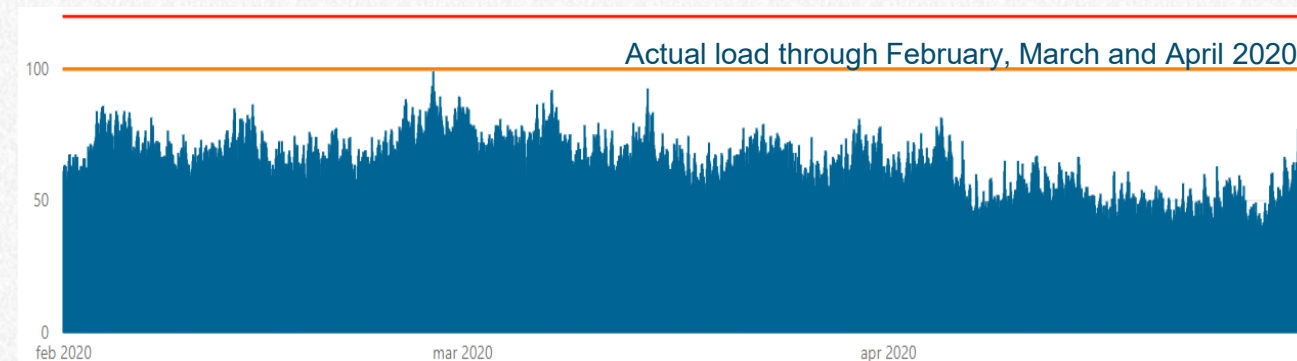


MAPE scores based on prediction for February, March and April 2020

XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
7.57	7.232	7.528	11.434	9.57	9.749	17.555	12.278	12.666



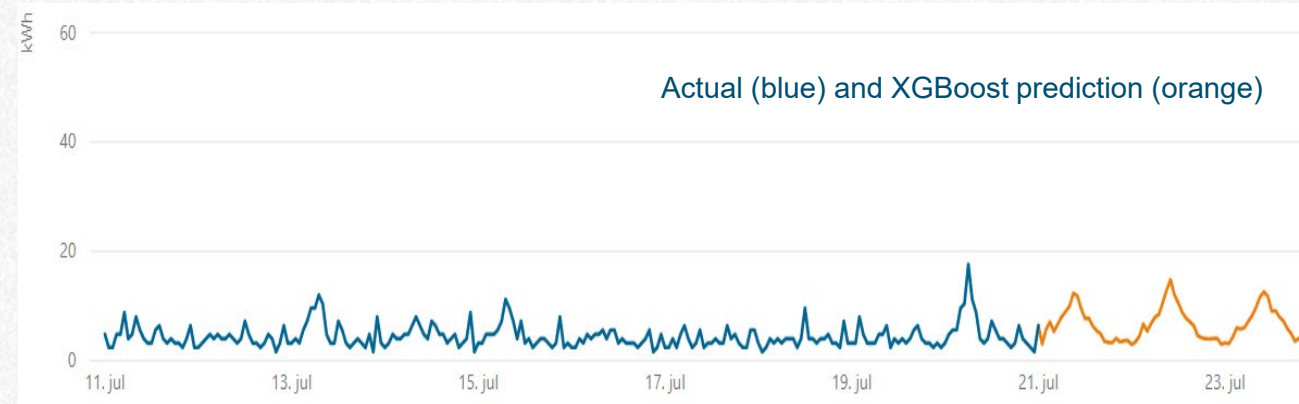
Number of customers: 50
 Type of area: Dwellings, some small industry.



Example 5: Substation load dominated by school building

The substation serves a school complex. Consumption during the prediction period (Feb-April 2020) was heavily influenced by temperature and the corona situation. In addition there is a period with missing data. Most models perform poorly.

Best model: XGBoost

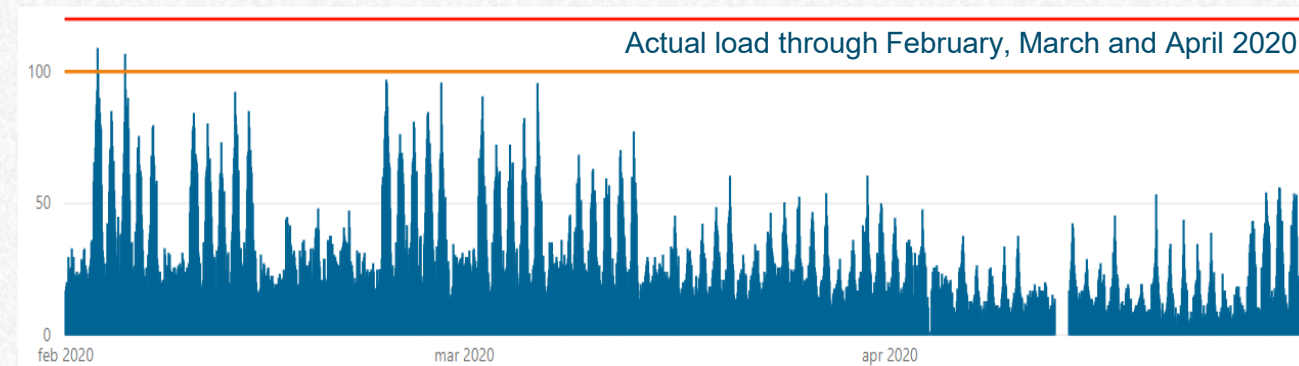


MAPE scores based on prediction for February, March and April 2020

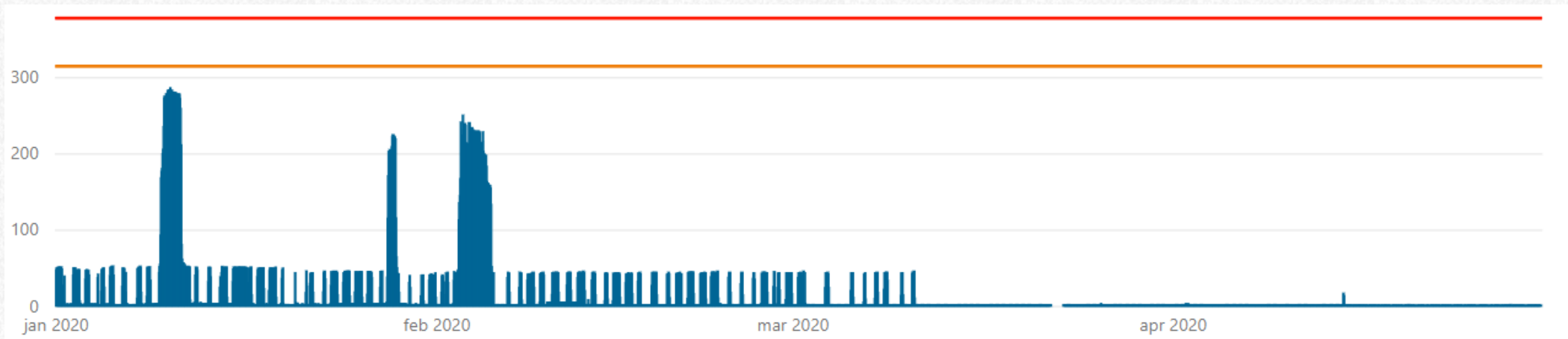
XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
26.144	32.688	33.481	32.509	46.281	42.475	58.008	40.542	37.976



Number of customers: 6
Type of area: School.



Some things are hard to predict ...



Summary: Which forecasting model is best?

There is no «one model best in all cases».

LGBM points out as the model with best score (lowest error measure) most often.

LGBM with **individual models** (i.e. model trained individually for each substation load profile) emerges as the best option.

SARIMAX and the naive models perform well in cases where the load is repetitive.

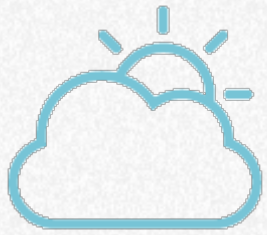
Count of «best model score» achieved

Model score (error measure)	XGBoost (individual models)	LGBM one common model	LGBM (individual models)	SARIMAX	naive_24h_mean	naive_24h_median	naive_seasonal	naive_last_period	naive_last_week
RMSE	2	6	12	2	2	0	0	1	0
MAPE	1	7	12	1	0	2	0	2	0

Fully automated ML pipeline in production

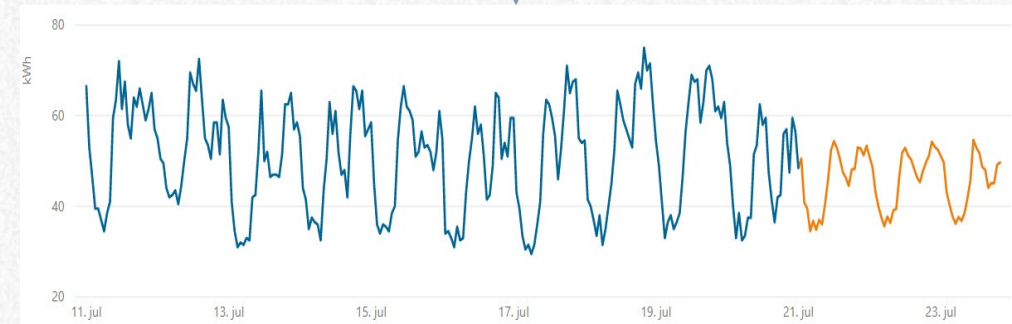
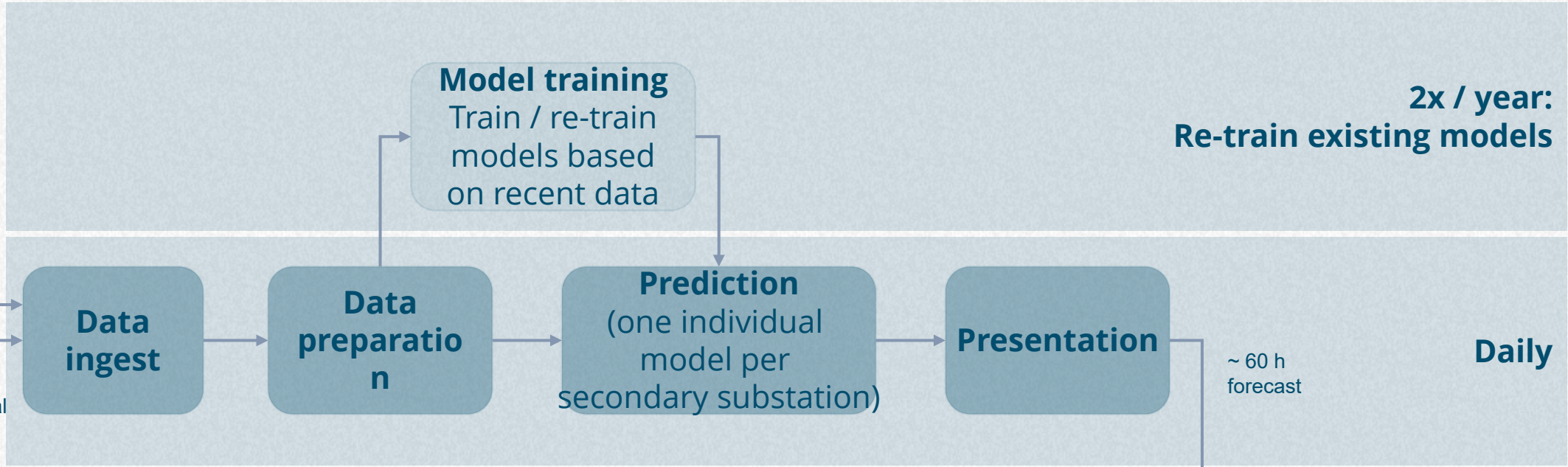
Currently predicting load for 123 secondary substations

Model based on XGBoost



67 h forecast

Historical load values



Agenda

Introduction

Application of Machine Learning

Summary

Hypotesis: ML has a great potential for utilities

Conclusion: **Yes**

Our investigations on load forecasting, image analysis, fault prediction, clustering and classification show that ML has a huge potential for utilities.

But

- Competence is needed (data scientists, data engineers)
- Deployment of new methods require periodic maintenance of ML pipeline
- The economical results need to be verified
- Data quality plays a vital role
- GDPR and security must be taken into account
- It takes time and effort to incorporate the methods into the organization

Thank you!

Per-Oddvar Osland
Agder Energi Nett
perosl@ae.no

