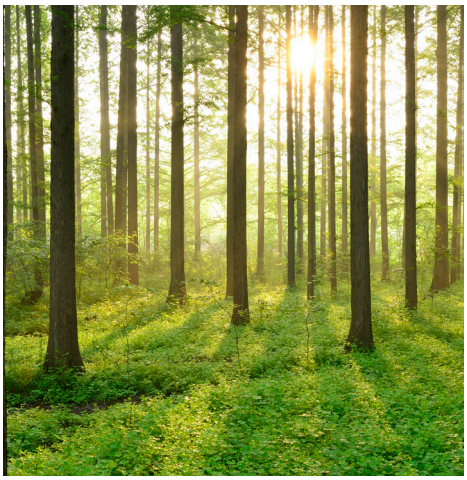
PREDIKTERING AV BEGYNNANDE FEL I ELEKTRISKA NÄT

RAPPORT 2022:889



ELNÄTENS DIGITALISERING
OCH IT-SÄKERHET

Energiforsk

Prediktering av begynnande fel i elektriska nät

En maskininlärningsbaserad metod för prediktering, lokalisering och identifiering av PQ-händelser

EBRAHIM BALOUJI
GUZAL BULATOVA
KARL BÄCKSTRÖM
VIKTOR OLSSON
JOHAN RÅDEMAR

ISBN 978-91-7673-889-4| © Energiforsk september 2022

Energiforsk AB | Telefon: 08-677 25 30 | E-post: kontakt@energiforsk.se | www.energiforsk.se

Förord

Projektet *Prediktering av begynnande fel i elektriska system* tillhör programmet Elnätens digitalisering och IT-säkerhet och syftar till att verifiera och vidareutveckla Eneryields algoritmer med data från Göteborg Energi samt ta fram ett proof-of-concept för produkten. Projektets maskininlärningsalgoritm kan förutsäga tiden till nästa störning, vilken typ av störning och hur allvarligt det är.

Ett proaktivt förhållningssätt till övervakning, innebär en bättre överblick av elnätet. I praktiken innebär det att störningar och fel kan bemötas mer effektivt, dvs att avbrott i större utsträckning kan förhindras samt att producerade kWh kan utnyttjas mer optimalt.

Johan Rådemar från Eneryield har varit projektledare för projektet och han har arbetat med Ebrahim Balouji, Karl Bäckström, Viktor Olsson, Guzal Bulatova också på Eneryield samt Ferruccio Vuinovich på Göteborg Energi.

Följande referensgrupp har på ett förträffligt sätt engagerat sig i projektet:

- Ferruccio Vuinovich, Göteborg Energi
- David Lundback, Vattenfall
- Martin Holmstrand, Ellevio

Stort tack också till programstyrelsen för programmet som består av ledamöterna:

- Kristina Nilsson, Ellevio (ordförande)
- Arne Berlin, Vattenfall Eldistribution
- Hampus Bergquist, Svenska kraftnät
- Ferruccio Vuinovich, Göteborg Energi
- Teddy Hjelm, Gävle Energi (Elinorr)
- Torbjörn Solver, Mälarenergi (vice ordförande)
- Magnus Sjunnesson, Öresundskraft
- Thorsten Handler, Jämtkraft
- Magnus Brodin, Skellefteå Kraft
- Johan Örnberg, Umeå Energi Elnät
- Peter Ols, Tekniska Verken i Linköping
- Jesper Bjärvall, Karlskoga Energi
- Peter Addicksson, HEM
- Malin Wallenberg, VB Energi
- Claes Wedén, Hitachi Energy Sweden
- Katarina Porath, ABB
- Björn Ållebrand, Trafikverket
- Patrik Björnström, Sveriges Ingenjörer (Miljöfonden)
- Matz Tapper, Energiföretagen Sverige (adjungerad)

Energiforsk framför ett stort tack för värdefulla insatser till samtliga intressenter:

Ellevio
Vattenfall Eldistribution
Svenska kraftnät
Göteborg Energi
Mälarenergi Elnät
Öresundskraft Elnät
Tekniska Verken i Linköping
Skellefteå Kraft Elnät
Umeå Energi Elnät
Jämtkraft Elnät
Elinorr ekonomisk förening
Eskilstuna Strängnäs E&M
Karlstads El- och Stadsnät
Borås Elnät
Halmstad E&M Nät
Luleå Energi Elnät
Hitachi Energy
ABB
Trafikverket
Sveriges ingenjörer (MF)
Forumet Swedish Smartgrid
Teknikföretagen
Exeri
Evado
Huawei Sverige
Borlänge Energi
Nacka Energi
Västerbergslagens Elnät
PiteEnergi
Södra Hallands Kraftförening
Karlskoga Elnät
Bergs Tingslags Elektriska
Blåsjön Nät
Härjeåns Nät
Sandviken Energi Nät
Sundsvall Elnät
Dala Energi Elnät
Elektra Nät
Gävle Energi
Hamra Besparingsskog
Hofors Elverk
Härnösand Elnät
Ljusdal Elnät
Malungs Elnät
Söderhamn Elnät
Åsele Elnät
Årsunda Kraft & Belysningsfören
Övik Energi Nät

Stockholm i augusti 2022
Susanne Stjernfeldt

Sammanfattning

Det har vuxit fram en ökad efterfrågan för stabila och pålitliga elnät, som samtidigt är miljömässigt hållbara. Nya innovativa lösningar behövs därmed för att möta uppsatta klimatmål utan att stabiliteten i elleveransen äventyras. Men med detta kommer redan idag stora utmaningar, som dessutom blir allt mer komplexa i och med samhällets pågående elektrifiering och den uppkoppling av nya enheter som medföljer.

Fokus i det här projektet är framtagandet av ett *Proof-of-Concept* för maskininlärnningsbaserad prediktering av begynnande fel i elnät, och därmed belysa fördelarna med en proaktiv approach i relation till O&M. I det här projektet används data från PQ-mätare stationerade i Göteborg Energis elnät, utan att ny hårdvara/mätutrustning installeras. I stället handlar projektet om att maximera nyttan av den data som redan idag spelas in och sparas.

I projektet testas, anpassas och vidareutvecklas Eneryields maskininlärningsmetod för prediktering och lokalisering av begynnande fel i elektriska nät. Tjänsten kan exempelvis förutspå allvarliga kabelfel dagar eller veckor innan de sker. Den tidiga varning som genereras ger nätägaren möjlighet till att vidta nödvändiga åtgärder för att undvika ett avbrott.

I själva analysen genomförs först klassificering av enskilda störningar med avseende på händelsetyp och lokalisering. Detta ger en mer djupgående förståelse för enskilda störningar och därmed elnätets generella hälsa. Med detta som grund utformas därefter algoritmer för maskininlärningsbaserad prediktering och lokalisering av begynnande kabelfel. Modellen visar goda resultat, med en uppmätt precision på 97% och en känslighet på 29%. Rent praktiskt innebär det att nästan 1/3 av alla oplanerade avbrott kan predikteras, med mycket litet antal falska positiva.

I och med detta har Eneryield framgångsrikt genomfört ett proof-of-concept för proaktiv analys i Göteborg Energis elnät med gott resultat.

Nyckelord

Maskininlärning, prediktering, klassificering, lokalisering, distributionsnät, elkvalitet,

Summary

An increased demand has emerged for stable and reliable power grids, which are at the same time environmentally sustainable. New innovative solutions are thus needed to meet set climate goals without compromising the stability of electricity supply. But with this there are already major challenges, which are also becoming increasingly complex with society's ongoing electrification and the connection of new units that follows.

The focus of this project is the development of a Proof-of-Concept for machine learning-based prediction of incipient faults in power grids, and thus highlight the benefits of a proactive approach in relation to O&M. This project uses data from PQ meters stationed in Göteborg Energi's power grid, without installing any new hardware/measuring equipment. Instead, the project is about maximizing the benefits of the data that today is already being recorded and saved.

The project tests, adapts and further develops Eneryield's machine learning method for predicting and locating incipient faults in power grids. The service can, for example, predict serious cable faults days or weeks before they occur. The early warning that is generated gives the grid owner the opportunity to take the necessary measures to avoid an outage.

In the analysis itself, the classification of individual disturbances with regard to event type and location is first carried out. This provides a more in-depth understanding of individual disturbances and thus the general health of the power grid. With this as a basis, algorithms are then designed for machine learning-based prediction and localization of incipient cable faults. The model shows good results, with a measured precision of 97% and a sensitivity of 29%. In practical terms, this means that almost 1/3 of all unplanned interruptions can be predicted, with a very small number of false positives.

As a result, Eneryield has successfully implemented a proof-of-concept for proactive analysis in Göteborg Energi's power grid with good results.

Innehåll

1 Introduktion

I samband med utbyggnaden av nya, distribuerade kraftkällor, samt fler uppkopplade enheter, förändras elnätet radikalt. Elnätet står inför stora utmaningar, med ett ökat energi- och effektbehov från redan existerande såväl som tillkommande användare. Vidare behöver elnätet klara en ökad flexibilitet, genom exempelvis lagring av överskottsenergi, detta parallellt med en försäkran om hög leveranssäkerhet. Rådande kapacitet kommer inom en relativt snar framtid inte att räcka till, därmed behövs nya innovativa lösningar för att mer effektivt undvika strömavbrott, ta vara på producerade kWh, underlätta implementering av intermittenta energikällor med mera.

Det har med detta vuxit fram en ökad efterfrågan för stabila och pålitliga elnät, som samtidigt är miljömässigt hållbara. Nya lösningar behövs därmed för att möta klimatmål, såsom 100% förnybar elproduktion till 2040, nettonollutsläpp av växthusgaser 2045 etcetera, utan att stabiliteten i elleveransen äventyras.

Strömavbrott är en incident med allvarliga konsekvenser som följd. Exempelvis kan det innebära produktionsstopp, energiförluster, hårdvaruskador, trafikstörningar etcetera. Nätägares huvudansvar är således att upprätthålla och underhålla ett stabilt och pålitligt nät. Men med detta kommer redan idag stora utmaningar, som dessutom blir alltmer komplexa i och med samhällets pågående elektrifiering och den uppkoppling av nya enheter som medföljer. I ett åldrande elnät världen över, ser vi således hur avbrotten ökar sett till antal och tid. Redan idag kostar strömavbrott bara i USA i snitt $33 miljarder per år (Energyprofessionals, 2022).

Digitalisering av elnät och stationer har länge varit ett pågående projekt, men ännu saknas sofistikerade verktyg för att underlätta nätägares dagliga arbete. Till exempel saknar nätägare insikter kring när ett strömavbrott riskerar att inträffa. Dessutom, om ett avbrott väl sker, så kan det vara svårt att veta var det inträffat. Ibland upptäcks till och med felet inte förrän en kund hör av sig. Nuvarande arbetssätt medför en begränsad överblick och förståelse för elnätets status, och är följaktligen reaktivt, snarare än proaktivt.

Idag samlar nätägare in mycket spännings- och strömdata. Både mindre och allvarligare störningar registrerade av PQ-mätare och reläskydd sparas. Dessvärre saknas metoder för att extrahera relevant information ur den redan tillgängliga datan.

Genom Eneryields innovation, kan nuvarande data användas för att förebygga strömavbrott, samt lokalisera känsliga områden i nätet. I och med sin insats verkar Eneryield således för digitalisering och automatisering av elnät, och därmed till stabil elleverans, förhöjd energieffektivitet samt underlättad implementering av förnyelsebara energikällor.

I den här rapporten redovisas resultaten från det projekt som genomfördes för Energiforsks räkning tillsammans med Göteborg Energi. Göteborg Energi bidrog med nödvändig data och det är således på denna som analysen genomförts. Eneryield har sedan tidigare utvecklat metoder för prediktering, lokalisering och

identifiering av PQ-störningar. Projektet var således en god möjlighet för att testa dessa metoder i ett nytt nät, med annan typ av uppsättning av mätutrustning och protokoll för datainsamling. Projektet innebär därmed att en adaptering/anpassning av Eneryields metoder för att prestera specifikt på data från Göteborg Energis elnät.

2 Bakgrund

Eneryields innovation baseras på aktuella forskningsresultat inom analys av elkvalitet (power quality, PQ) och elektriska system med hjälp av maskininlärning. Eneryields teknologi grundar sig i nära ett decenniums forskning utförd av företagets grundare (Balouji et al., 2017: Balouji et al., 2018: Balouji et al., 2018, Balouji et al., 2019: Balouji et al., 2020, Balouji et al., 2020: Balouji er al., 2022). Genom denna grundläggande forskning och fortsatt vidareutveckling av state-of-the-art maskininlärning har Eneryield tagit fram unika algoritmer för proaktiv analys av elnät.

Eneryield tillhandahåller en mjukvarubaserad maskininlärningsmetod för prediktering av begynnande kabelfel i elektriska nät. Tjänsten kan förutspå allvarliga kabelfel timmar, dagar eller veckor innan de sker. Den tidiga varning som genereras ger nätägaren möjlighet till att vidta nödvändiga åtgärder för att undvika strömavbrott.

Tjänsten använder historisk PQ-data (vågform och RMS-värden) för att prognostisera framtida status i nät. Genom utveckling av sofistikerade maskininlärninsalgoritmer har Eneryield tagit fram en lösning som genom att först djupgående analysera individuella mindre störningar, sedan kan identifiera mönster och trender bland dessa. På så vis identifierar metoden hur en viss typ av mindre störningar i spänning och ström över tid utvecklas till allvarliga kabelfel. Således kan tjänsten notifiera kontrollrummet att ett avbrott kommer att ske inom en viss tidpunkt, vid en viss plats till en viss sannolikhet.

Det förstås att tillförlitlig prediktering av förestående avbrott bygger på initial noggrann analys av individuella störningar. I Eneryields pipeline för analys ingår därmed klassificering av störningstyp, grundorsaksanalys och lokalisering. På så vis kan en nätägare också förstå vad som orsakat en störning samt var i nätet den skett.

Eneryields utveckling och förståelse för marknadens behov har vuxit fram ur flera års tid nära samarbete med flertalet olika nätägare. Till exempel genomfördes för Energiforsks räkning ett projekt under 2020 (Balouji et al., 2020). Flera nätägare deltog i projektet, där Eneryield på verklig och kundspecifik data, testade sina metoder. Till exempel gjordes analys för klassificering av störningstyp, grundorsak, lokalisering (spänningsnivå) etcetera. Arbetet som utfördes i projektet var framgångsrikt, och la i viss mån grunden för detta projekt, där analytiska insikter tillsammans med förväntad kundnytta, styrde Eneryield mot att tillhandahålla en proaktiv approach för nätanalys genom prediktering av kabelfel.

Sedan tidigare finns det en del aktörer som arbetar med prediktering av fel i relation till kablar. Gällande detta finns ett flertal beprövade metoder, där den kanske mest välkända är analys av partiella urladdningar (Faisal et al., 2012: Dashti et al., 2021: Balouji et al., 2019). Nödvändig utrustning för att analysera PD (Partial Discharge) är dock dyr, och kan vara utmanande i mätmässigt brusiga situationer. Vidare fungerar inte detta tillvägagångssätt för nedgrävda kablar.

Andra metoder för att för förutsäga kabelfel använder data från exempelvis temperatursensorer (Chen et al, 2021) eller drönare (Jalil et al., 2019). Även dessa strategier kräver ny hårdvara, som i sin tur kräver etablering av data-pipelines med tillhörande bearbetning och lagring av stora dataströmmar med låg latens. Detta leder till betydande kostnadsökningar. Inte heller dessa approacher är kompatibla för nedgrävda kablar.

Att kombinera analys av PQ-data med maskininlärning för proaktiva insikter i elnät är ännu ett relativt outforskat område. Ett exempel på ett pågående forskningsprojekt inom området är EarlyWarn, som leds av Sintef Energy (Sintef Energy, 2021). EarlyWarn är ännu på forskningsstadiet och det påpekas i deras senaste rapporter att modellen de utvecklar ännu inte är tillräckligt bra för direkt implementering i en operativ miljö. Vidare är tidshorisonten i EarlyWarn tämligen kort och träffsäkerheten låg. En för stor mängd falska positiva är också ett problem. Men projektet är likväl ett mycket intressant initiativ inom ett område där Eneryield kan vara med och bidra till fortsatt utveckling

3 Omfattning

Projektet baseras på forskning inom maskininlärning för PQ och elektriska system. I arbetet används data från PQ-mätare stationerade i Göteborg Energis elnät. Ingen ny hårdvara/mätutrustning behöver installeras. I stället handlar projektet om att maximera nyttan av den data som redan idag spelas in och sparas.

I projektet testas en maskininlärningsalgoritm på tillgänglig data. Den maskininlärningsbaserade lösningen möjliggör effektiv mönsterigenkänning och trendanalys, och iden är att kartlägga hur mindre avvikelser och störningar leder upp till mer allvarliga fel. Dvs, identifiera trender och mönster för begynnande fel i ett tidigt stadie. För att exemplifiera, lösningen möjliggör att kabelfel kan upptäckas innan allvarliga fel sker. På så vis kan underhåll och drift planeras optimalt.

Maskininlärningsalgoritmen kan förutsäga ifall en störning kommer att ske inom en viss framtida tidshorisont, vilken typ av störning det kommer att vara, hur allvarligt det kommer att bli samt dess ungefärliga position. Sannolikheten för att något kommer att ske presenteras i en sannolikhetsmatris med en flexibel predikteringshorisont. Således kan en tidig varning genereras, som meddelar driftcentralen att vidta nödvändiga åtgärder innan ett allvarligt fel inträffar. Vidare kan lösningen estimera plats och grundorsak för en specifik störning. Lösningen är helt unik och innebär en bättre överblick av elnätet och dess hälsotillstånd, som i förlängningen gör det möjligt att undvika avbrott, förlänga livslängden på hårdvara samt att nyttja energi mer effektivt.

Projektet är ämnat att verifiera och vidareutveckla Eneryields algoritmer med data från Göteborgs Energi samt att i samförstånd genomföra ett proof-of-concept för proaktiv analys i elnät.

3.1 SYFTE

Det saknas automatiserade digitala verktyg för att förstå hur fel uppstår samt hur de utvecklas över tid. Således inträffar idag störningar och fel som hade kunnat förhindras med en mer proaktiv metodik.

Projektet syftar till att genom en behovsstyrd och kundorienterad approach, belysa fördelarna med en maskininlärningsbaserad proaktiv strategi, vilket står i kontrast till dagens mer reaktiva sätt. Projektet fokuserar på följande utmaningar:

- Nuvarande approach till underhåll och övervakning är reaktiv, snarare än proaktiv
- Redan tillgänglig data används inte optimalt

3.2 MÅL

Projektets mål är:

- Framtagandet av ett *Proof-of-Concept* för maskininlärnningsbaserad prediktering av begynnande fel i elnät

- Utvärdera tillförlitlighet, robusthet och bredd på Eneryields lösning samt specificera krav på datainsamling och IT-säkerhet genom att besvara följande frågeställningar:
    - Kan Eneryields maskininlärningsmodell prediktera kabelfel?
    - Kan Eneryields maskininlärningsmodell prediktera andra typer av störningar?
    - Kan Eneryields maskininlärningsmodell estimera distansen till/plats för en störning?
    - Vilka krav ställs på data och datainsamling för att maximera värdet av Eneryields lösning?
    - Vilken nivå av konfidens kan tilldelas Eneryields lösning och vilka viktiga faktorer ligger bakom hög konfidens?
    - Vilka komplikationer finns utifrån ett IT-säkerhetsperspektiv, och hur bemöter man dessa?

4 Teori

I detta avsnitt introduceras metoder för maskininlärning, prestandaindikatorer samt elkvalitet (PQ).

4.1 METODER FÖR ARTIFICIELL INTELLIGENS OCH MASKININLÄRNING

Vi ger i det följande en sammanfattning av begrepp och metoder inom Artificiell Intelligens (AI) och Maskininlärning (ML) som är relevanta för projektet.

AI och ML. AI är ett datavetenskapligt område dedikerat till utveckling av den kategori metoder som besitter egenskaper liknande mänsklig intelligens, exempelvis inlärning och problemlösning. Inom paraplybegreppet AI kan ML definieras som den underkategori av metoder avsedda för inlärning, dvs att nyttja data för förbättrad prestanda på en uppsättning problem. Således kan metoder inom ML med fördel användas för att automatiskt hitta mönster i olika typer av data som är relevanta för specifika uppgifter.

Djupinlärning. Användning av Artificiella Neuronnät (ANN), mer specifikt Djupinlärning (DL), har på senare år ökat markant, både inom akademisk forskning, men även i industriella tillämpningar, tack vare dess förmåga att prestera mycket väl på många olika problem. ANN är en klass av ML-metoder, som karaktäriseras av komplexa arkitekturer av beräkningsstrukturer, bestående av ett antal basala beräkningsenheter som kallas *neuroner*, oftast strukturerade i flera lager. Ett ANN betraktas i regel som en DL-metod då arkitekturen innehåller tre eller fler sådana lager av neuroner. Utöver in- och ut-lagren i ett ANN (som utgörs av indatan, samt önskad utdata), så behandlar varje lager resultatet av föregående lager med en affin transformation/operator, betingad av en inlärd viktmatris, vilket sedan behandlas av en icke-linjär *aktiveringsfunktion*, resultatet av vilket ges vidare till efterföljande lager. Olika operatorer kan nyttjas i olika lager, beroende på topologin hos indatan; Exempelvis är (i) faltningslager lämpliga för bild-liknande data, tack vare dess förmåga till translationsinvariant mönsterigenkänning, och (ii) RNN-lager för data i tidssekvens, tack vare dess förmåga att aggregera tidigare indata i ett tillstånd, samt behandla in-sekvenser av variabel längd (se Tabell 1). Inom ramen av projektet utvecklas specialanpassade DL-arkitekturer som är optimerade för behandling av de aktuella datatyper som funnits tillgängliga, samt för den specifika uppgiften som adresseras, nämligen prediktering av kabelfel. Som exempel, eftersom sådan prediktering kräver kontinuerlig analys av tidsserier av variabel längd, lämpar sig RNN-baserade (Recurrent Neural Network) arkitekturer väl för en sådan tillämpning.

Namn	*Operator*
Perceptron	$out_n = \sigma\left(\sum_{i=0}^{\lvert in\rvert} w^n[i] \cdot in[i] + b\right)$
Faltning	$out_n^f = \sigma\left(\sum_{i=0}^{k} w^f[i] \cdot in[n+i] + b\right)$
Recurrent Neural Network (RNN)	$a^{(t)} = Wh^{(t-1)} + Ux^{(t)} + b$ $h^{(t)} = \sigma\left(a^{(t)}\right)$ $out^{(t)} = Vh^{(t)}$
ReLU	$\sigma(x) = \max(0, x)$
Sigmoid	$\sigma(x) = (1 + e^{-x})^{-1}$
Softmax	$\sigma_i(x) = \frac{e^{\text{out}_i}}{\sum_{j=1}^{K} e^{\text{out}_j}}$ for $i \in \{1, \ldots, K\}$

Tabell 1. Olika ANN-lager och aktiveringsfunktioner.

4.2 PRESTANDAINDIKATORER (KPI:ER)

Ett flertal storheter (KPI:er) nyttjas för estimering och evaluering av prestandan hos de olika metoder som utvecklas inom projektet, vilka vi sammanfattar nedan. För flera KPI:er är det relevant att bekanta sig med termerna *falska positiva/negativa* predikteringar (**FP, FN**), samt *sanna positiva/negativa* (**TP, TN**), sammanfattade i följande tabell:

		predikterat tillstånd	
		Positiva	Negativa
faktiskt tillstånd	Positiva	**TP**	**FN**
	Negativa	**FP**	**TN**

Tabell 2: Överblick över falska/sanna positiva/negativa predikteringar

Notera att **TP+FN** utgör samtliga positiva tillstånd, och **TN+FP** samtliga negativa.

MSE och korsentropi. Under träningsfasen av ett ANN tillämpas en optimeringsalgoritm för att minimera en *felfunktion*, som kvantifierar prestandan hos ett ANN i varje givet tillstånd av dess viktmatriser. Olika typer av problem kräver i regel olika felfunktioner, men *kvadratfel* (MSE) är ett lämpligt val i flera sammanhang:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

där Y, $\hat{Y}$ är vektorer av faktisk, respektive önskad, utdata, och n är antalet datapunkter. Vid högre kännedom om problemets och datans innebörd och topologi kan man med fördel välja mer specifika felfunktioner mer lämpliga för uppgiften. Exempelvis är *korsentropi* (CE), som kvantifierar entropiskillnaden mellan sannolikhetsfördelningar, lämplig för just klassificeringsproblem:

$$\mathrm{CE} = -\sum_{i=1}^{n} Y_i \cdot \log \hat{Y}_i$$

Träffsäkerhet (CR). Ett av de vanligare prestandamåtten inom ML är *träffsäkerhet*, vilket i många sammanhang kan ge en god indikation på metodens användbarhet:

$$\mathrm{CR} = \frac{\#\text{Korrekta prediktioner}}{\#\text{Prediktioner}} = \frac{TP + TN}{TP + FP + TN + FN}$$

Träffsäkerhet kan dock vara vilseledande för obalanserad data; Om en klass utgör 95% av all data kan 95% träffsäkerhet uppnås om metoden endast gissar på den dominerande klassen. Om de klasser som ska identifieras ej utgör balanserade andelar av datan är det därför nödvändigt att för en prestandaanalys komplettera med mått som tar hänsyn till klassernas storlek, exempelvis *känslighet* och *precision*.

Känslighet. För att mäta känsligheten hos en metod att (i) prediktera händelser av relevans, eller (ii) korrekt klassificera en specifik klass används *känslighet* med fördel:

$$\text{Känslighet} = \frac{\#\text{Korrekta positiva prediktioner}}{\#\text{Positiva}} = \frac{TP}{TP + FN}$$

Känslighet ger ett mått på hur många av de positiva fallen i datan som en modell korrekt klassificerade dem som sådana.

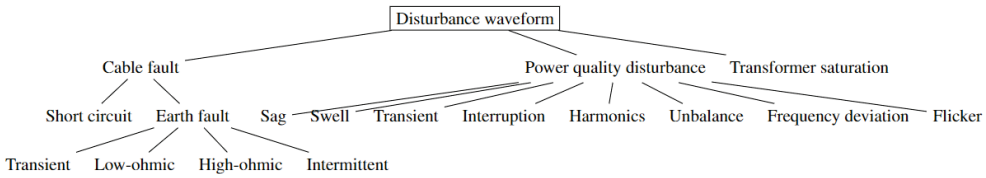
Precision. Det är relevant att en metod för prediktering presterar med hög *recall*, dvs. hittar många av de faktiskt relevanta händelserna. Notera dock att detta är möjligt att uppnå enkelt genom att utveckla en mycket känslig metod, med hög risk för falska positiva. För många relevanta problem är det dock extra viktigt att undvika just falska positiva, varför man med fördel även bör beakta metodens *precision*, som mäter andelen korrekt positiva predikteringar:

$$\text{Precision} = \frac{\#\text{Korrekta positiva prediktioner}}{\#\text{Positiva prediktioner}} = \frac{TP}{TP + FP}$$

Det är i allmänhet önskvärt att ha maximal prestanda i samtliga ovanstående storheter, men i praktiken krävs ofta en kompromiss mellan just känslighet och precision.

4.3 ELKVALITET (PQ)

Elkvalitet (PQ) kan definieras som kombinationen av mått på kvaliteten hos ström och spänning. I växelströmsystem motsvarar ideala ström- och spänningsvärden sinusoidala kurvor, med konstant frekvens och amplitud, samt att ström och spänning är i fas. Det är avvikelser från detta tillstånd som studeras inom PQ. Detta inkluderar *PQ-händelser* samt *PQ-variationer*, där det förstnämnda avser plötsliga effekter med en början och ett slut (exempelvis spänningsdippar, avbrott, etc.), och det sistnämnda mer tillfälliga (exempelvis transienter, övertoner). Inom projektet är även *kabelfel* relevanta, eftersom dessa är användbara indikationer som kan korrelera med allvarligare kabelfel och avbrott. En mer utförlig beskrivning finns att tillgå i (Balouji et al., 2020), och en överblick över flera av de mest relevanta störningarna visualiseras nedan:



Figur 1:

5 Om data och elnät

Här nedan ges en överblick för elnät och tillgänglig data.

5.1 INFORMATION OM DATA

Tre typer av data har använts i projektet:

PQ-data som samlades in från 61 PQ-mätare stationerade i Göteborg Energis elnät (MV/HV) under perioden 2008 - 2021. Datan registrerades i följande format:

- Störningsvågform som utlösts av en avvikelse som t.ex. spänningsdip;
- Löpande RMS av spänning och ström, som registrerades varje timme.

Avbrottsrapport (excel-ark) med start- och sluttider för avbrott (störningar) som inträffade i Göteborgs kommun mellan 2008 och 2021. Varje avbrott beskrivs med typ (exempelvis som planerad eller oplanerad), linje, fördelningsstation, felorsak m.m. Rapporterna innehåller information om totalt 29 756 avbrott.

Relä-data från fem mätande reläer stationerade i området som matas av specifik fördelningsstation. Varje inspelning innehåller fem vågformer: fyra ström-faser och en neutral spänning. Inspelningarna är en sekund långa. Totalt användes 11 filer med störnings-vågformer.

5.2 DATAINSAMLING

PQ-data inhämtas från mätar-leverantörens databas. Data från 2008 till 2021 fanns tillgänglig. Eneryield utredde all typ av information som fanns i databasen och identifierade de mest relevanta med avseende för detta projekt:

1) störnings-vågformer: inspelningar av vågformer som reflekterar avvikelser. Varje inspelning innehöll fyra vågformer (3-fas samt neutral) för spänning respektive ström och var 0.2 sekunder långa. Totalt extraherades 30 795 inspelningar;

2) löpande RMS: periodiska inspelningar samlade varje timme. Varje RMS innehåller fyra vågformer för spänning respektive ström över 0.2 sekunder.

I projektet användes data från 61 olika PQ-mätare utplacerade i MV- (10 kV) och HV- (130 kV) näten. I Fig. 1 nedan visas ett elnäts topologi för en fördelningsstation *KX*, med tre respektive transformatorstationer. Vid varje transformatorstation placeras en PQ-mätare, märkt som *PQ, PQ2 och PQ3*. I och med data från dessa PQ-mätare återfås en översikt över elnätets kvalitet i området som fördelningsstationen distribuerar el till.

Totalt samlar dessa 61 mätanordningar data från 18 fördelningsstationer, där majoriteten har en uppsättning med 3 PQ-mätare. I vissa fall kan upp till 6 PQ-mätare (på båda sidor av en transformatorstation) finnas.

Vidare i distributionssystemet visas linjer med reläer, märkt med *R1 … RX*. I avbrottsrapporten markeras alla avbrott utifrån specifik linje där avbrottet registrerades, samt motsvarande fördelningsstationsaddress. Denna information är viktig för att förstå var olika avbrott har inträffat, och därmed vilka PQ-mätare som är relevanta att använda för att förstå vad som förekom felet. På så vis kan en tidslinje byggas.

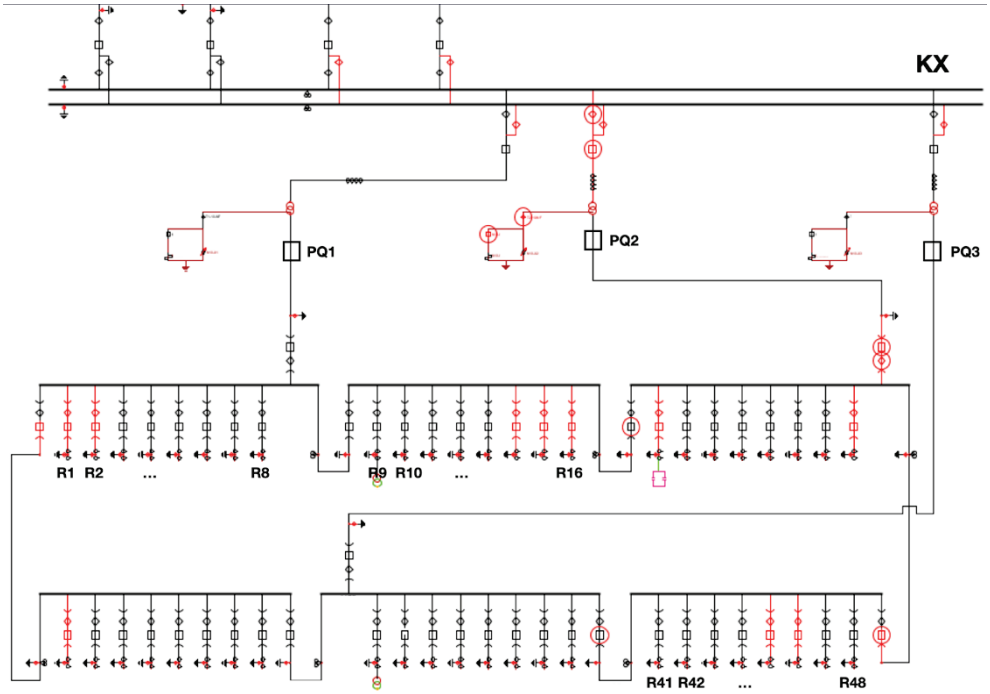


Fig 1. Fördelningsstation KX topologi.

5.3 TIDSMÄSSIG SYNKRONISERING AV DATA

För att utföra kvalitetskontroll av datan jämfördes tre olika datakällor: PQ-data, avbrottsrapport och inspelningar från relä. I ett idealt scenario förväntas det att PQ-mätare skall registrera varje dokumenterat avbrott som en störning. Alltså, för varje noterat avbrott i avbrottsrapporten och/eller relä-mätningar skall det i PQ-datan tidssynkront finnas motsvarande inspelning.

Först genomfördes en intern jämförelse mellan tider för inspelningar som gjorts av PQ-mätare: löpande RMS mot störningsvågformer. Förväntningen var att avbrott registrerade i löpande RMS skulle ha en motsvarande vågform i samma PQ-mätare. Det framkom dock att endast 46% av alla avbrott i löpande RMS hade en motsvarande störningsvågform. Detta innebär att mer än hälften, ***54% av all avbrott registrerade som RMS i PQ-mätare saknade motsvarande inspelning av störningsvågform.*** Detta trots att matchningen genomfördes med en timmes felmarginal, då datan samlas in på olika sätt: RMS registreras varje timme medan vågformer spelas in vid varje störning.

I nästa steg jämfördes tiderna för avbrott noterade i avbrottsrapporten med tiderna registrerade för inspelade störningsvågformer. Hänsyn togs till att

avbrottsrapporten ifylls manuellt, varför en marginal på en timme före/efter för att matcha avbrott med störningsvågformer användes även här. Jämförelsen gjordes oavhängt på PQ-mätarens placering, och en validerad matchning baserades på tid, oavsett PQ-mätarens område och fördelninsstationens adress som angivits i rapporten. Här påträffades en matchning på 24%, det vill säga att ***76% av avbrott registrerade i avbrottsrapporten saknade motsvarande inspelning av störningsvågform med en timmes marginal och oberoende av position.***

Slutligen jämfördes relä-inspelningar för 2021 med registrerad tid för störningsvågform. Där söktes en matchning inom 24 timmar. Det återfanns en inspelning i datan från PQ-mätare som motsvarade tiden registrerad i relä, om än med tre timmar tidsskillnad. Övriga ***10 av 11 registrerade avbrott i relä saknade motsvarande inspelning av störningsvågform inom ett dygn.***

Möjliga skäl för denna låga motsvarighet är:

1. Felaktig läsning av data. Detta eliminerades genom att noggrant kontrollera insamlad data: datan hämtades från databas på två olika sätt, med sökning i SQL programmeringsmiljö samt med sökning med hjälp av Pythons bibliotek för databashantering. De två gav samma resultat, av detta dras slutsatsen att läsningen av datan är korrekt.
2. Felaktig datalagring. Det är möjligt att störningar registrerades av PQ-systemet men inte sparas i databasen.
3. Felaktig placering av PQ-mätare. Om PQ-mätare inte är lämpligt placerade så kan störningar och avbrott missas.

Genomförd kvalitetskontroll indikerar att PQ-mätarna inte uppfattar alla fel i elnätet. Detta blir därmed en avgränsning för projektets omfattning och således en bestämmande faktor i urval av data för prediktion. Detta innebär dock inte nödvändigtvis att insamlad PQ-data inte kan användas för prediktering av begynnande kabelfel. För att utvärdera om tillhandahållen data och mätning är tillräcklig för prediktering av begynnande kabelfel tränas en maskininlärningsmetod på tillgänglig data varpå slutsatser dras baserade på metodens prestanda.

5.4 SAMMANFATTNING AV INSAMLAD DATA

Tre typer av data användes i detta projekt:

1) Data från PQ-mätare utplacerade vid transformatorstationer på både medel- och högspänning-sidan. Detta ger en översikt över elnätets 18 fördelningsstationer. Dataformat är binära inspelningar av spänning- och strömvärde i de tre faserna samt den neutrala. Mätningarna utgörs av vågformer för olika typer av störningar eller RMS-värden för varje timme. Data finns från 2008-2021.

2) Avbrottsrapport i excel-format, är den andra typen av data som fanns tillgänglig. Den tillhandahåller en mer detaljerad lokalisering av fel. Här anges exempelvis vilken linje i fördelningsstationens område som var felaktig. Här återfanns dessutom starttid, sluttid, plats och orsak för bekräftade avbrott som skedde i Göteborgs kommun 2008-2021.

3) Tredje datatypen är mätningar från relä i form av COMTRADE-filer. De är oplanerade avbrott som skedde under 2021: vågform av ström och neutral spänning för 11 störningar. Relädatan används som tillägg till avbrottsrapporten. Tillsammans anger de två källorna tiderna för de fel som ska predikteras. Således utgör de målvärde för maskininlärningsmodellen.

6 Analys och resultat

I detta avsnitt avhandlas genomförd analys med tillhörande metod samt dess resultat.

6.1 FÖRBEHANDLING AV DATA

Första steget av själva analysen är att förbehandla datan. Detta görs för att förbereda datan för maskininlärningsmodellens träning och validering.

Förbehandlingen genomförs i fyra steg:

1. **Datarensning** - Identifikation av avvikande mätdata och eliminering av brus som t.ex. dubbletter eller vågformsinspelningar som saknar värden.
2. **Datatransformation** - Omvandling av insamlad data från olika original-format - binära inspelningar från PQ-mätar-databas och COMTRADE filer från reläer - till enhetliga numeriska tabellformatet. Dessa normaliserades sedan för att minska effekt av absoluta värden. T.ex. normala värden för spännings-RMS från en PQ-mätare är 8 kV i 2008-2013 och 10 kV i 2014-2021. Utan normalisering så skulle dessa inte kunna jämföras på ett rimligt sätt.
3. **Datareducering** - Relevant data för analys identifierades, varpå relevanta datasetet skapades för klassificering och prediktering. T.ex. total kundavbrottstid från avbrottsrapporten är irrelevant för prediktering. Dessutom reducerades antal värden som representerar en störning genom att kalkylera olika skalära värden. Detta resulterade i transformation av ca. 20 000 värden i de råa vågformerna till ca. 100 koncisa sammanfattande värden.
4. **Dataintegrering** - Gjordes för att kunna använda två data-format tillsammans. Här synkroniserades vågforms-inspelningar och avbrottsinformation från rapporten på två sätt. Först skapades artificiella avbrott som vågformer för att expandera datasetet. Sedan berikades tiderna från avbrottsrapporten med tider för störningar och varpå en gemensam tidslinje skapades av avvikelser för varje fördelningsstations område.

I förbehandlingen gjordes datamängden mer precis och komplett. Felaktiga och saknade värden som är ett resultat av den mänskliga faktorn eller övriga mätfel eliminerades, samt vid behov fylldes de attribut som saknades i. Det andra målet med förbehandlingen är att förbättra konsistensen av data för att öka resultatens noggrannhet. Därmed jämnades datasetet ut och gjordes således lättare att tolka.

6.2 KLASSIFICERING: IDENTIFIERING AV HÄNDELSETYP OCH LOKALISERING AV STÖRNING/FEL

För att uppnå tillförlitlig prediktering är det viktigt att först identifiera relevant data, alltså att utröna vilka PQ-inspelningar som indikerar en utveckling av allvarliga fel. Därför genomfördes en data-driven automatiserad klassificering i två

steg samt analys för lokalisering av PQ-störningar. Analysen är delvis baserad på Eneryields metod som presenterades i företagets tidigare energiforsksprojekt som genomfördes under 2020 (Balouji et al., 2020).

Klassificering handlar om att kategorisera störningar, och genomförs utan manuell handpåläggning. Till exempel görs här distinktionen mellan kabelfel, avbrott, PQ-störning etc. Detta möjliggör att störningar som inte är relevanta för prediktering av begynnande fel kan filtreras bort.

En viktig aspekt av den PQ-data som använts är att en del av den utgörs av flera störningar registrerade av flera PQ-mätare med en minimal tidsskillnad (mindre än en sekund). Detta kan bero på:

- olika störningar som sker samtidigt i olika delar av distributionssystem
- felspridning, eller speglingseffekt: när en störning är så pass allvarlig att den påverkar spänning och ström i flera delar av ett område som därmed registreras av flera PQ-mätare.

I det senare fallet måste inspelningen som är inspelad där felet skedde skiljas från så kallade speglingseffekter. Speglingseffekterna bör identifieras som förbehandling till predikteringen.

Därför appliceras vår metod för att analysera de fel som registrerats i olika mätare men med en väldigt liten tidsskillnad (en sekund). Således klassificeras störningar som verkligt fel eller speglingseffekt. I projektet benämns proceduren *lokalisering*, mer specifikt identifieras det område där störningen sker. Som resultat tilldelas en PQ-störning två annoteringar (labels): händelsetyp och lokaliseringsbeteckning. Till exempel kan en PQ-störning kategoriseras som händelsetyp: kabelfel, med lokaliseringsbeteckning: speglingseffekt.

En övergripande process för ML-driven klassificering beskrivs i *Fig. 2*. Processens flöde är samma för varje vågformsinspelning från databasen. I exemplet laddas till schemat en omärkt vågform som är okänd för ML-modellen. Först sker klassificering beträffande händelsetyp, såsom exempelvis kabelfel. Om det finns flera störningar registrerade vid samma tidpunkt, utförs lokalisering. Analysen avgör alltså om störningen för specifik mätare är ett verkligt fel eller en speglingseffekt. Alltså avgörs om störningen skedde i området närmast en viss PQ-mätare, eller om det är en speglingseffekt som har färdats i nätet från längre bort.

I Fig. 2 överblickas processen som exemplifierar analysen för klassificering och lokalisering. Denna procedur genomfördes för samtliga tillhandahållna vågformer. Det förstås att lokaliseringen bara genomförs om det finns parallella störningsinspelningar. Finns det bara ett event vid en given tidpunkt antar följaktligen algoritmen att störningen är verklig.

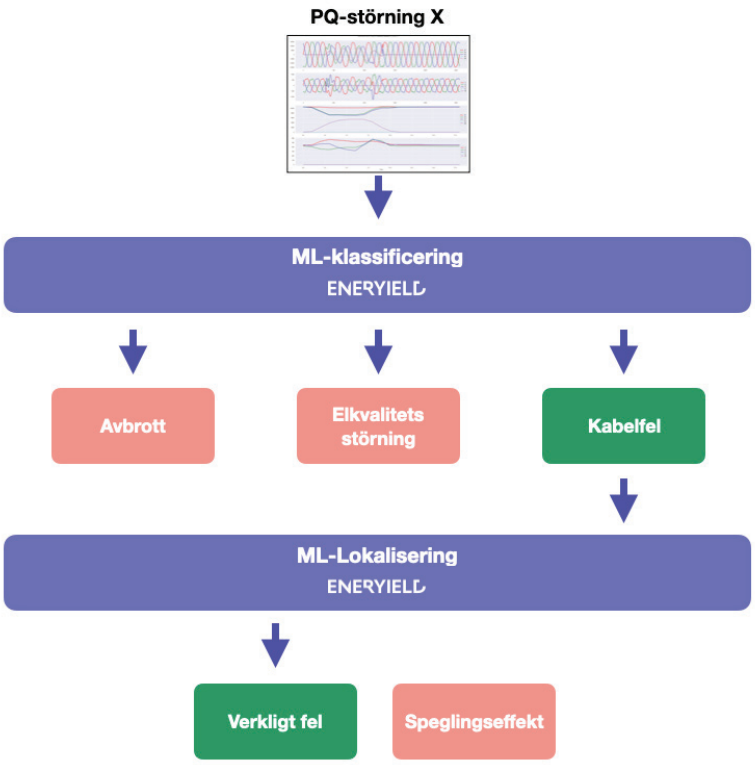


Fig 2. Överblick av metoden för störningsklassificering.

ML-arkitekturen är ett Faltningsnät (CNN) med tillhörande förbehandling av indata. I Fig. 3 visualiseras det föreslagna systemet för klassificering. Förbehandlade inspelningar från PQ-databasen matas genom separata kanaler till neuronätet. Bäst prestanda uppnåddes med en arkitektur bestående av fyra faltningslager.

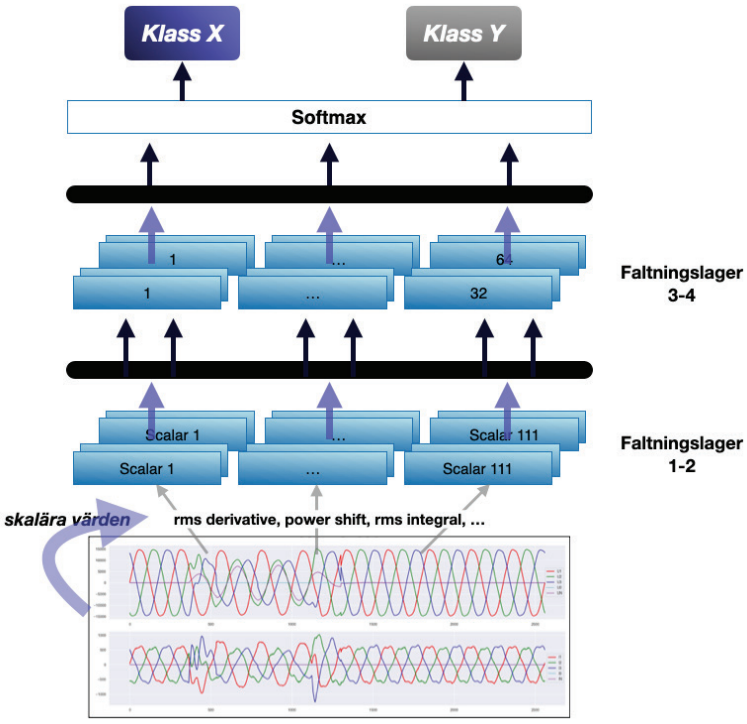


Fig 3. Upplägg för klassificering. Olika skalära värden matas in i CNN-arkitekturen som består av fyra faltningslager.

För att den utvecklade modellens resultat ska bli användbar och tillförlitlig behövs det en viss mängd datapunkter som har annoteringar av faktiska klasser.

I träningsfasen matas modellen med både vågformer för störning med dess respektive annoteringar, d.v.s. algoritmens förväntade utdata. Under detta träningssteg justeras parametrarna i modellen för att optimalt klassificera träningsdata baserat på den data som tillförs. Efter slutförd träningsfas utförs testfasen, där modellens prestanda observeras på den separata testdatan. I testfasen tillförs bara vågformer utan förväntade resultat, på så sätt kan modellens utdata och den faktiska annotering jämföras. Modellen klassificerar alla PQ-störningar från tillgänglig data först efter händelsetyp; som avbrott, kabelfel, etc. Efter detta tilldelas varje händelsetyp en lokaliseringsbeteckning: speglingseffekt eller verkligt fel.

I tabell 3 ges en överblick av den data som använts för att bygga modellen. Dessutom presenteras distributionen för olika händelsetyper och tillhörande lokaliseringsbeteckning. Detta ger en överblick av distribution för olika händelsetyper i vågformsdatan samt påvisar hur frekventa speglingseffekter är i det samlade datasetet.

Händelsetyp	Fördelning	Verkligt fel	Speglingseffekt
Kabelfel	90 %	30 %	70 %
Avbrott	6.1 %	81.6 %	18.4 %
Elkvalitet	3.9 %	77 %	23 %

Tabell 3. Fördelning av PQ-störningar händelsetyp och dess lokaliseringsbeteckning.

I och med genomförd klassificering var det möjligt att extrahera ytterligare information gällande störningarna registrerade av PQ-mätare. Framgångsrik klassificering genomfördes över två nivåer (typ och lokalisering). Klassificeringsresultaten möjliggör ett robust urval av tidiga tecken på begynnande allvarliga fel för prediktering. Detta ger även djupare förståelse av elnätets kvalitet genom att lokalisera känsliga områden i nätet.

6.3 PREDIKTERING AV BEGYNNANDE FEL

Målet med prediktering är att förutse begynnande fel innan de sker och på så vis förebygga strömavbrott och skadad utrustning. För detta behövs ett dataset som innehåller två typer av variabler: så kallade *målvärden* - värden som modellen ska prediktera, samt oberoende variabler - alla andra tillgängliga variabler som används för att estimera målvärden. I detta projekt är målet att prediktera huruvida ett strömavbrott kommer att ske snart, och detta blir således målvärdet. För att estimera detta analyseras datan från PQ-mätare - detta blir i det här fallet de oberoende variablerna. Därmed inkluderas ingen information om när framtida avbrott skall ske, utan endast historiska observationer från mätare i elnätet tills nu, samt tider för historiska avbrott. På så sätt reflekterar datan som används för prediktering en verklig situation, där framtida data är okänd.

Först gjordes vissa avgränsningar. I projektet fokuserades predikteringen till en femårsperiod (2017-2021) och en specifik fördelningsstation, KX. Avgränsningen gjordes baserat på tillgänglig datamängd, samt på att flera transformatorer i nätet ersattes 2017, vilket innebar skiften i datan före och efter bytet. Hädanefter används därmed som oberoende data inspelningar från tre PQ-mätare som placerats vid tre transformatorer i KX, nämligen PQ1, PQ2, PQ3 under samma tidsperiod, 2017-2021.

I den tillhandahållna avbrottsrapport kan det avläsas att 50.1% av alla tillgängliga 958 avbrottstider för valt fördelningsstationsområde är planerade, därmed är de inte aktuella att användas i analysen. Kvarvarande 478 oplanerade avbrott har olika felorsaker angedda i avbrottsrapporten. Dock måste även viss sållning göras bland de oplanerade avbrotten. Detta då vissa registrerade avbrott har felorsak av exempelvis typerna fallet träd, grävning, snö, sabotage. Dessa slumpartade händelser är svåra att förutsäga och sorteras därmed bort för att inte förvirra modellen. Istället riktas fokus mot de fel som har tillhörande orsak som indikerar att utvecklingen för felet kan ha registrerats tidigare i systemet. Det vill säga tillverknings-eller materialfel, säkringsbrott (utlöst säkring), överbelastning, felaktig montering/förläggning, bristande underhåll etc. Den här typen av fel utgör 22% av totala oplanerade avbrott. Det är tiderna för dessa avbrott som används

som målvärden i analysen. I *Fig. 4* återges en tidslinje för avbrott som skett under ett år sorterade efter avbrottstyp (planerade, oplanerade utom scope, oplanerade inom scope), markerade med olika färger. De långa röda linjerna visar avbrottsurvalet för prediktering.

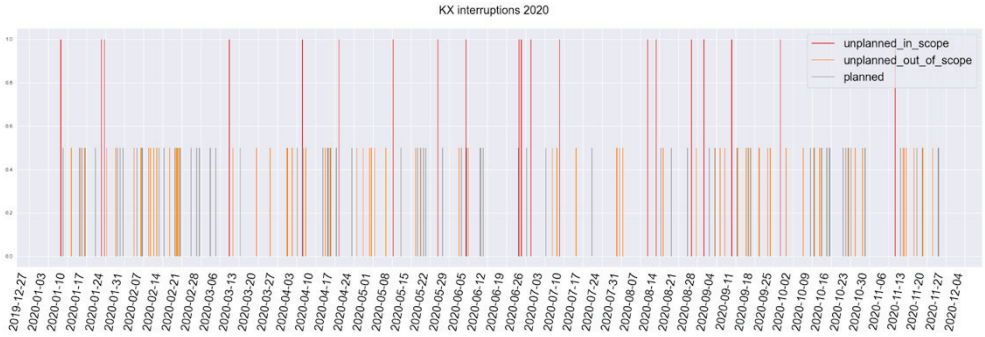


Fig 4. Tidslinje för avbrott under ett år. Tiderna för avbrott i fokus är märkta med röda långa linjer.

För predikteringen utvecklades en algoritm baserad på ett beslutsträd. Här är målet att exempelvis identifiera återkommande mönster i datan från PQ-mätaren på MV/HV-sidan av fördelningsstationen som korrelerar med allvarliga fel som sker på LV-sidan och som registrerats som avbrott i avbrottsrapporten. Ett exempel på identifierat mönster presenteras i *Fig. 5*.

Störningar i sekvensen som föranledde ett kabelfel har följande karakteristiska egenskaper:

- Klassificerade som kabelfel
- Tidsskillnad mellan störningar är mindre än en sekund, d.v.s. sekvensen är registrerad av flera PQ mätare
- Betydande avvikelser observeras i:
    i. Amplitud och mönster i signalen för neutral spänning (UN), samt
    ii. Derivatan av UN (förändringshastigheten).

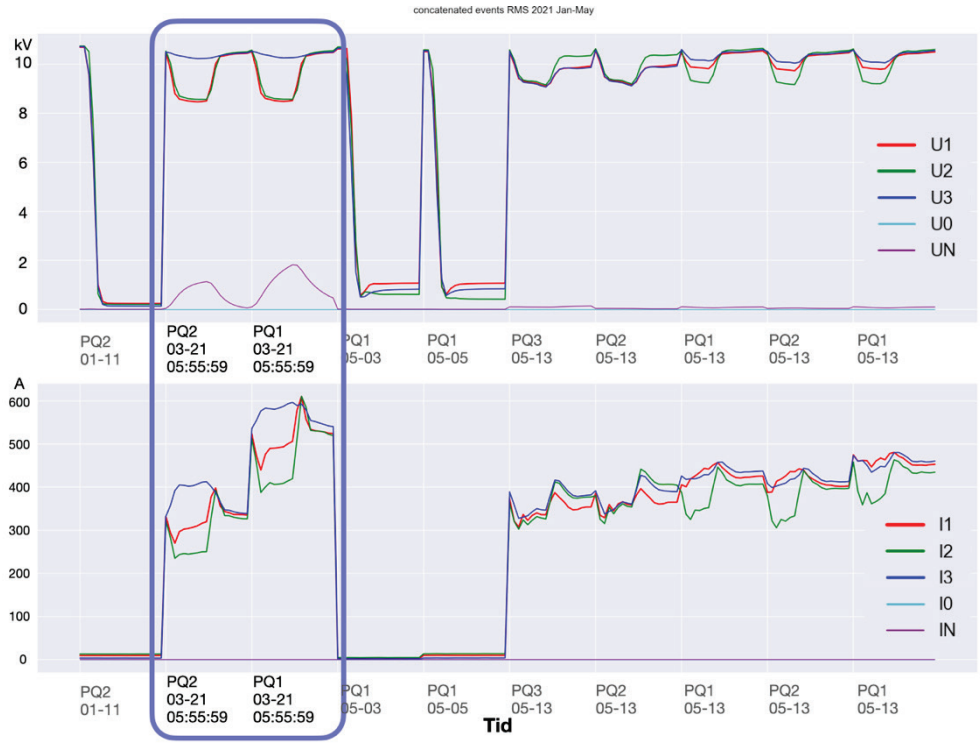


Fig 5. Ett exempel på identifierat mönster i PQ-datan som korrelerar med efterföljande allvarligt kabelfel.

Modellen matas med indatan iterativt genom en *sliding-window metod*. Alltså, en datamängd registrerad under ett visst "tidsfönster" av fast längd laddas och analyseras. Sedan förflyttas fönstret längs tidslinjen ett steg vidare och laddar datadelen som tillhör den nya tidsperioden. På så sätt emuleras en löpande dataövervakning. Varje dag görs, till exempel, predikteringen baserad på historisk data från de två föregående veckorna, det vill säga, med ett tidsfönster på fjorton dagar. Det vill säga om tidpunkten då en prediktering ska göras är den 15 juli, analyseras data från 1-14 juli. Nästa dag, den 16 juli förflyttas fönstret vidare och predikteringen uppdateras med ny information och estimerar således datan från 2-15 juli, osv.

Baserat på olika mönster som modellen identifierar i den historiska datan uppskattas sedan om ett fel kommer att inträffa inom en nära framtid. Det övergripande upplägget av predikteringsmekanismen finns att tillgå i *Fig. 6*. Tidsperioden för prediktering (prognosperioden) är fjorton dagar. Om algoritmen identifierar ett mönster som korrelerar starkt med allvarliga fel i den tillgängliga datan, triggar modellen en positiv förutsägelse, vilket således betyder att "inom den estimerade prognosperioden från och med nu" kommer det med hög risk att inträffa ett allvarligt kabelfel.

Modellen matas med indatan iterativt genom en *sliding-window metod*. Alltså, en datamängd registrerad under ett visst "tidsfönster" av fast längd laddas och analyseras. Sedan förflyttas fönstret längs tidslinjen ett steg vidare och laddar datadelen som tillhör den nya tidsperioden. På så sätt emuleras en löpande dataövervakning. Varje dag görs, till exempel, predikteringen baserad på historisk data från de två föregående veckorna, det vill säga, med ett tidsfönster på fjorton dagar. Det vill säga om tidpunkten då en prediktering ska göras är den 15 juli,

analyseras data från 1-14 juli. Nästa dag, den 16 juli förflyttas fönstret vidare och predikteringen uppdateras med ny information och estimerar således datan från 2-15 juli, osv.

Baserat på olika mönster som modellen identifierar i den historiska datan uppskattas sedan om ett fel kommer att inträffa inom en nära framtid. Det övergripande upplägget av predikteringsmekanismen finns att tillgå i *Fig. 6*. Tidsperioden för prediktering (prognosperioden) är fjorton dagar. Om algoritmen identifierar ett mönster som korrelerar starkt med allvarliga fel i den tillgängliga datan, triggar modellen en positiv förutsägelse, vilket således betyder att "inom den estimerade prognosperioden från och med nu" kommer det med hög risk att inträffa ett allvarligt kabelfel.

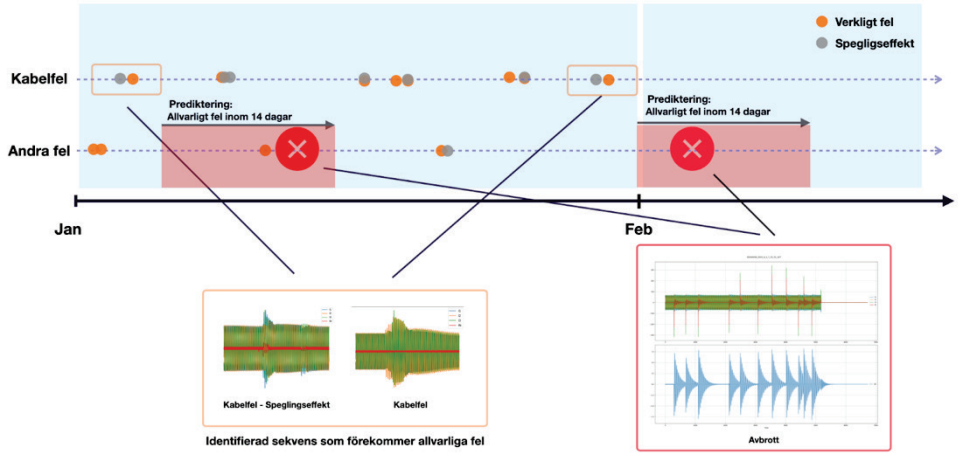


Fig 6. Överblick över metoden för prediktering av begynnande allvarligt kabelfel som ledde till avbrott.

Metodens predikteringar valideras sedan mot målvärden (tider för bekräftade avbrott från avbrottsrapporten). De primära prestandamåtten (KPI:er) som nyttjas här är precision och känslighet (se definition under Teori). Om det inom en tidsperiod skedde ett fel så markeras det som en positiv datapunkt, om inte - som negativ. Precision kan ses som ett mått på hur bra modellen är på att undvika falska positiva. Känslighet visar hur många av de avbrott som har skett som modellen lyckades förutspå. I *Fig. 7* presenteras modellens predikteringar samt målvärde på samma exempel över en ettårsperiod, vilket ger insikt i modellens prestanda. Här kan vi se att alla gånger som modellen gav en positiv prediktion, alltså förutspådde att ett fel är på väg att ske, så skedde det ett fel inom 14 dagar.

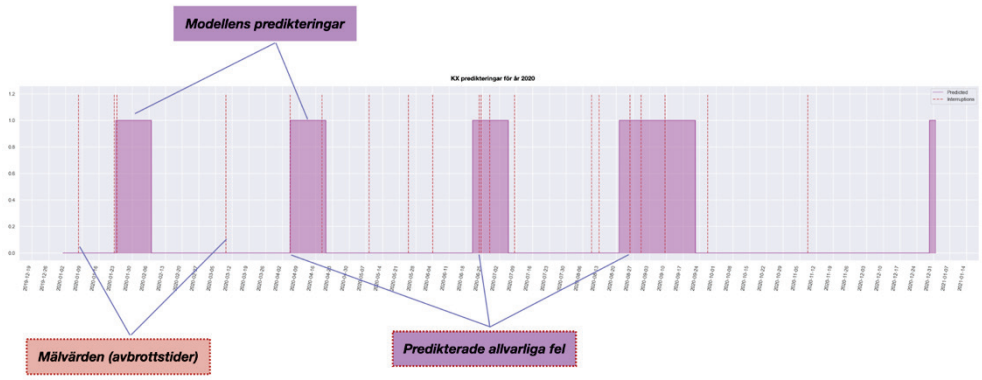


Fig 7. Felperioder identifierade av modellen markerade med lila fönster samt avbrottstider markerade med röda streckade linjer.

Modellens KPI för predikteringar med prognosperioden av två veckor som gjordes över perioden av 2017-2021 är:

- **Precision: 97%** (utan falska positiva resultat)
- **Känslighet: 29%** (andel fel som predikterats)

Lokaliseringen av störningar som genomfördes i klassificeringens andra steg användes även för prediktering av framtida fel. Det vill säga varningen säger inte bara att ett fel kommer att ske, utan även var i nätet det kommer att hända. Detta då störningar som observerats i den historiska datan tydliggör vilken del av elnätet som är felaktigt. På så vis kan platser ringas in i nät där utveckling av allvarliga fel pågår (exempelvis kabelfel). Analysen baseras på att PQ-mätare registrerar störningar i fördelningsstationsområdet på olika sätt beroende på i vilka linjer det sker. Med andra ord, värdet av spänning- och strömsignal i inspelningar som registrerats närmare platsen för störningen har vissa specifika egenskaper. Exempel på dessa är en högre derivata (förändringshastighet) av den neutrala spänningen (UN), karaktäristiska mönster i olika faser, tidigare tidsregistrering, etc.

I *Fig. 8* ses exemplet på lokalisering av predikterat framtida fel. Det lila området markerar predikterade felets läge: modellen estimerade att ett allvarligt fel skulle ske i någon av linjerna ett till åtta. I relädatan finns en inspelning av ett avbrott inom predikterad period i linje 7. Denna delen av analysen kan utvidgas med tillgång till mer relädata.

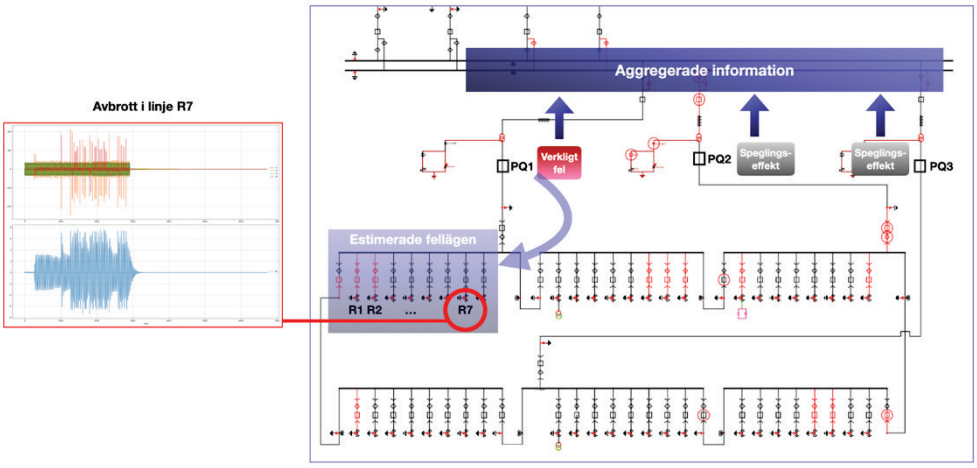


Fig 8. Exempel på framgångsrik prediktering av framtida position för fel.

7 IT-säkerhet

Metoder som utvecklats och studerats inom projektet, såsom *prediktering* av kabelfel, *analys* av PQ-störningar, mm., utgör stora möjligheter för en mer proaktiv hållning i övervakning och underhåll av elnätet. Detta kräver dock att de nya funktionerna görs tillgängliga för operatören, exempelvis genom att de kan opereras genom ett användargränssnitt (UI), som även används till att delge och visualisera resultatet av den analys som utförs. Detta ställer flera krav hos nätägare, operatörer, samt hos den aktör som tillhandahåller analys, gällande datainsamling, datatransmission, datalagring, analysinfrastruktur, m.fl., speciellt relaterade till *IT-säkerhet*.

7.1 DISKUSSION

Tre centrala begrepp inom IT-säkerhet som med fördel beaktas vid systemutformning är (i) tillgänglighet, vilket innebär graden till vilka systemfunktioner och tjänster är tillgängliga att nyttjas av användaren, (ii) sekretess, samt (iii) integritet, vilka garanterar att endast auktoriserad part har möjlighet att tillgå, respektive modifiera, den information som avses. I det följande diskuteras IT-säkerhetsaspekter, specifikt i relation till de metoder som utvecklas inom projektet, framför allt med fokus på punkterna (ii) och (iii), dvs. sekretess och integritet. Valet av fokus beror framför allt på att dessa aspekter anses vara de mest relevanta i sammanhanget, samt att just tillgänglighet är mycket beroende av hårdvaru- och topologiaspekter, vilka varierar kraftigt mellan olika nät, varför en sådan diskussion är lämplig att ha gentemot specifika nätägare och operatörer för att garantera relevansen.

Metoderna som utvecklas av Eneryield, som paketeras som en mjukvarumodul, exekveras i praktiken på någon beräkningsenhet i systemet, exempelvis på server/cloud-nivå hos nätoperatör, på intern server hos Eneryield genom en on-demand/API funktion som kan anropas av mjukvara hos nätoperatör, eller möjligtvis i mindre enheter närmare sensor, i.e. fog/edge-beräkning. Inom ramen för detta projektet, där syfte är att uppskatta användbarheten och prestanda hos Eneryields metoder, har samtliga beräkningar skett internt hos Eneryield, vilket inte kräver någon utförligare diskussion kring IT-säkerhet. Fokus ligger istället på integreringsmöjligheter, med dess potentiella risker, på sikt, med produktisering och lansering av densamma i åtanke. Detta kräver att infrastruktur för datakommunikation, dataanalys, och prediktering, tas i beaktning, speciellt från ett IT-säkerhetsperspektiv, där ökade risker relaterade till sekretess och integritet måste vägas mot möjligheter och nytta av moderna analys och predikteringsmetoder.

7.2 REKOMMENDATION PÅ SIKT

Rekommenderat upplägg. *Fig. 9* ger en överblick över det upplägg som rekommenderas för att på sikt möjliggöra en integrering av extern AI-baserad analys i nätoperatörers existerande infrastruktur. I den rekommenderade integreringen samlas mätdata in som tidigare till nätoperatörerns centrala moln-

lagring, men genomgår dessutom analys av lanserade AI-modeller, resultatet av vilka görs tillgängligt till operatör förslagsvis genom ett grafiskt användargränssnitt (GUI). Nätoperatören avgör vilken del av datan som kan delges till tillhandahållaren av AI-modeller, som sedan nyttjar denna data för att utveckla nya AI-modeller, som kontinuerligt gradvis kan förbättras allteftersom mer data samlas in från de nätägare som tar del i ett sådant samarbete. Nya AI-modeller bör valideras innan de lanseras skarpt, vilket görs genom (i) tester på insamlad data hos Eneryield, samt genom att skicka kandidater för AI-modellen ut till en eller flera nätoperatörer, där mer, potentiellt privat, data lagras. Hos nätoperatören kan sedan modellkandidater testas, resultatet av vilket (KPI:er) skickas tillbaka till Eneryield, som därigenom får en god uppfattning om hur modellerna presterar på flera uppsättningar data. Med detta upplägg behöver endast modellkandidater kommuniceras, inte mätdata, vilket minskar kommunikationsbelastning, samt IT-säkerhetsrisker. Vid lyckad validering lanseras sedan de uppdaterade AI-modellerna, genom att de skickas till nätägarens moln-server, där de ersätter existerande modeller.

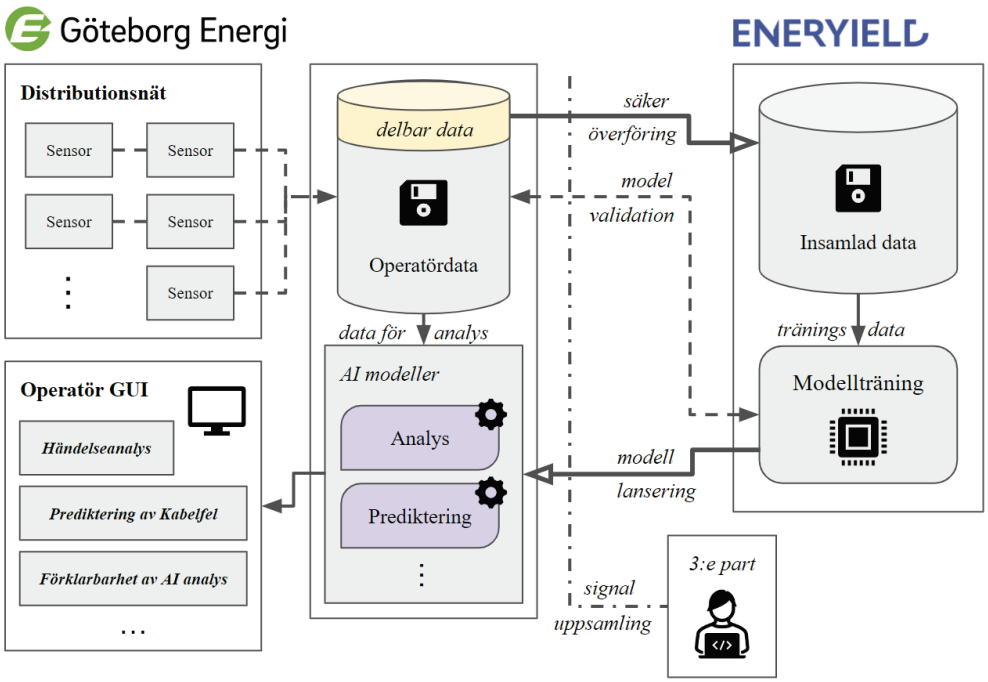


Fig 9. Överblick över rekommenderat upplägg för integrering av Eneryields analysmetoder

Analys. Här följer en kortfattad säkerhetsanalys, från perspektivet att en tredje part otillbörligen avlyssnar datakommunikation mellan nätägare och Eneryield, som illustrerat i *Fig. 9*. Kommunikation för insamling av mätdata ligger utanför ramen för diskussionen, då denna varierar beroende på den specifika nättopologin samt den utrustning som används av nätoperatören. Först och främst bör samtlig kommunikation över internet ske krypterat, med tidigare överenskomna nycklar, där krypteringsalgoritm lämpligen väljs baserat på vilken typ av data som avses, samt hur ofta överföring ska ske. Exempelvis för överföring av större mängder data vid enstaka tillfällen, så som ytterligare träningsdata, från '*delbar data*' hos

nätoperatör, till 'insamlad data' hos Eneryields interna server, anses en symmetrisk krypteringsmetod med hemlig nyckel tillräcklig. För regelbunden automatisk överföring av data, i form av mätdata eller nya AI-modeller, rekommenderas *asymmetrisk* kryptering, där krypteringsnyckeln är öppen, men dekrypteringsnyckeln endast är känd av mottagaren av datan, eftersom en symmetrisk nyckel löper större risk att spridas. Detta minskar drastiskt risken att tredje part får tillgång till den data som kommuniceras. Utöver detta är det rekommenderade upplägget i *Fig. 9* utformat för att minimera kommunikation av känslig data, och därmed även konsekvenserna om data mot förmodan ändå läcker till tredje part. Exempelvis är det av mindre nytta för tredje part att tillgå viktmatriser hos nya AI-modeller, eller KPI:er som skickas tillbaka till Eneryields server. Den svagaste länken är överföring av ny träningsdata, vilket är varför det rekommenderade upplägget endast kräver att detta sker vid vissa, antingen manuellt valda, eller periodiska, tillfällen, och inte nödvändigtvis kontinuerligt, vilket andra potentiella upplägg sannolikt krävt. Ett alternativt upplägg skulle exempelvis kunna vara en API-lösning, där insamlad data kontinuerligt skickas till Eneryields server, där analys sker, och resultatet skickas tillbaka till nätoperatör. Detta skulle kräva mer kommunikation av känslig data, vilket belastar kommunikationsinfrastrukturen, samt ökar antalet attackpunkter för tredje part. Detta alternativa upplägg är å andra sidan ett bra insteg mot integrering, då det i princip kan fungera som en emulering av ett fullintegrerat upplägg, utan att större modifiering behöver ske hos nätoperatör. Detta alternativ bör dock endast tillämpas på nät med icke-känslig data.

8 Reflektion

Här ges en reflektion kring några aspekter av projektet.

8.1 PQ-MÄTNINGAR OCH DATAINSAMLING

Genomförd analys påvisar att PQ-mätare inte registrerar alla avbrott och störningar som sker i nätet. Den jämförande analys som gjordes mellan projektets tre datakällor indikerade på inkonsekvens och lågt överensstämmande källorna emellan. Detta kan indikera att nuvarande system för PQ-mätning inte är optimal. Förslagsvis bör därför nuvarande övervakningssystem noggrant kontrolleras för att säkerställa att det fungerar som avsett.

Följande orsaker bör närmare kontrolleras:

- Tröskelvärden för hur PQ-mätare registrerar störningar
- Datakommunikation mellan mätning och databas
- Placering av mätare i nät - finns möjligen "blinda fläckar"

Om inkonsekvensen i insamlad data kan elimineras finns goda möjligheter till en mer exakt överblick av elnätet. Delvis i en så kallad nuläges-prognos, men framförallt ur ett proaktivt hänseende.

Att gå vidare med att säkerställa en optimal placering, samt eventuellt komplementera med ytterligare PQ-mätare är en sak. Mer intressant kan dock vara att uppdatera med moderna reläer kapabla till mer noggrann mätning samt effektiv datainsamling. Att använda störningar inspelade från reläer i analysen skulle möjliggöra:

- Mer precis lokalisering av störningar och avbrott (genom att utesluta fler linjer i predikteringen), och
- mer exakt estimering av tid till framtida fel

Sammanfattningsvis; att använda PQ-data tillsammans med relädata möjliggör predikteringar med högre precision samt mer noggrann identifiering av felaktig linje för begynnande fel.

8.2 UTMANINGAR - DATAINSAMLING OCH BEGRÄNSNINGAR

För att förenkla användandet av extern analys bör en plattform (data lake eller liknande) finnas tillgänglig där stora mängder sensordata kan lagras för att sedan effektivt inhämtas. I detta projekt och gentemot nuvarande system förelåg vissa utmaningar beträffande detta. Data från PQ-mätare insamlades genom att göra en kopia av leverantörens SQL-databas. Denna procedur för att tillgängliggöra data blir något omständig och dessutom uteblir alla typer av uppdateringar av dataset. I och med detta kan analysen inte göras kontinuerligt och i realtid.

En hållbar integrering av den här typen av analys skulle gynnas av standardiserade protokoll för datahantering. Om det är möjligt ur ett IT-

säkerhetsperspektiv, så kan datalagring i moln vara ett alternativ för enkelt utbyte. Vidare kan åtminstone en specialprogrammerad exportmodul för automatisk export av PQ-störningar vara av nytta vid offline-träning av modulen.

Omfattningen och tillförlitligheten för analysen kan också öka om inspelningen från PQ-mätare bättre reflekterar och därmed fångar upp de avbrott som faktiskt sker i nätet. Nu finns det perioder i den historiska datan där störningar helt saknas, men där avbrott enlig avbrottsrapporten har skett. Detta begränsar modellens möjlighet att i ett tidigt stadie uppfatta exempelvis begynnande kabelfel. Det kan vara så att delar av nätet ligger utanför PQ-mätningens "synfält", samt att vissa störningar är så små att de inte registreras. Att öka kvaliteten på inspelningar från PQ-mätare samt användandet av relädata komplementärt skulle troligtvis förbättra analysen och dess datadrivna predikteringar ytterligare.

9 Slutsatser

I det här projektet har Eneryield framgångsrikt genomfört ett *proof-of-concept* för proaktiv analys i elnät med gott resultat. Analys utfördes på tillgänglig data från Göteborgs Energi, med avseende på klassificering och lokalisering av störningar registrerade av PQ-mätare. Detta gav en mer djupgående förståelse för enskilda störningar och därmed elnätets generella hälsa. Med detta som grund utformades därmed algoritmer för maskininlärningsbaserad prediktering och lokalisering av begynnande fel. Både kabelfel och andra typer adresserades.

Modellen påvisade goda resultat, med en uppmätt precision på 97% och en känslighet på 29%. Rent praktiskt innebär det att nästan 1/3 av alla oplanerade avbrott kan predikteras, med ytterst få falska positiva. Vidare innebär modellen att processer som idag utförs manuellt kan automatiseras med angivna värdefulla insikter kring elnätets nuvarande och framtida status som följd.

Projektets resultat påvisar att det redan idag, inom ramen för existerande infrastruktur, är möjligt att genomföra en mer djupgående analys av PQ-data som kan tillgängliggöra proaktiva insikter. Men samtidigt belyser också projekten avsaknaden av standardiserade metoder för hantering av PQ-data. Det finns förbättringspotential för inspelning, insamling och export. Det förelåg också vissa brister i loggning av avbrott olika datakällor emellan. Kan detta åtgärdas, samt relä-data tillgängliggöras ännu smidigare för analys, så finns det goda möjligheter att resultaten kan bli ännu bättre.

10 Fortsatt arbete

I följande avsnitt resoneras det kring nödvändiga nästa steg för att ta genomfört arbete vidare. Projektet visade mycket lovande resultat, men för att modellen ska kunna användas i praktiken krävs ytterligare validering och utveckling.

Till att börja med så behöver algoritmen valideras på ytterligare data. Datan ska vara diversifierad och gärna från flera olika platser för att säkerställa prestandan. Det vill säga att för att maximera robustheten och generaliserbarheten hos algoritmen så måste den exponeras för så många olika typer av störningar som möjligt. Vidare kan mer data nyttjas för att kunna identifiera flera andra typer mönster som föranleder ett avbrott. Genom att identifiera samt träna på dessa kan precision och känsligheten öka ytterligare.

I projektet fanns endast begränsad tillgång till relädata. Genom arbetet blev det dock tydligt att relädata kan användas för att estimera distans och plats för störning mer noggrant. Som återsågs i resultat, kan viss avgränsning av nätet göras med endast PQ-data tillgänglig, men vill man mer noggrant precisera platsen för störningen krävs tillgång till relädata. Därmed föreslås som fortsatt arbete att i större skala inkludera relädata i analysen. Det skulle potentiellt räcka att inkludera RMS-värden för ström. Hur man tillgängliggör den här datan för att inkorporera i Eneryields analys är intressant att fokusera vidare på.

Efter detta blir sedan nästa steg att tillgängliggöra modellen för faktisk användning i realtid. Algoritmerna behöver förpackas på ett sådant sätt att de kan ta emot data kontinuerligt/periodvis, analysera densamma och därefter leverera tillbaka och visualisera ett resultat. För att möjliggöra detta krävs utveckling av en mjukvarumodul med tillhörande gränssnitt.

11 Referenslista

Balouji, E. & Salor, Ö. (2017). *Classification of power quality events using deep learning on event images.* 3rd International conference on pattern recognition and image analysis (IPRIA)

Balouji, E., Yu-Hua Gu, I., Bollen, M., Bagheri, A. & Nazari, A. (2018). *A LSTM-based Deep Learning Method with Application to Voltage Dip Classification*. 18th International Conference on Harmonics and Quality of Power (ICHQP).

Balouji, E., Yu-Hua Gu, I., Bollen, M. & Bagheri, A. (2018). *A robust transform-domain deep convolutional network for voltage dip classification*. IEEE Transactions on Power Delivery 33(6)

Balouji, E., Hammarström, T. & McKelvey, T. (2019) *Partial discharge classification in power*
electronics applications using machine learning. IEEE Global Conference on Signal and Information Processing (GlobalSIP), pages 1–5.

Balouji, E. Bäckström, K. McKelvey, T. & Salor, Ö. (2020). *Deep-learning-based harmonics and interharmonics predetection designed for compensating significantly time-varying EAF currents.* IEEE Transactions on Industry Applications, 56(3)

Balouji, E., Berggren E., Bäckström, K., Kärrman, I. & Rådemar, J. (2020). *Maskininlärning, elkvalitet och smarta elnät - Tillämpning av maskininlärning vid analys av elnät.* Energiforsk. RAPPORT 2020:691.

Balouji, E., Baghaee, H. R., Hajizadeh, A., Soltani, M., Lin, Z. & Savaghebi, M. (2022). *Locating high-impedance faults in DC microgrid clusters using support vector machines.* Applied Energy, 308:118338

Chen, K., Yue, Y. & Tang, Y. (2021) Research on temperature monitoring method of cable on 10 kv railway power transmission lines based on distributed temperature sensor. Energies, 14(12):3705.

Dashti, R. Daisy, M., Mirshekali, H. Shaker, H.R. & Aliabadi, M. H. (2021). *A survey of fault prediction and location methods in electrical energy distribution networks.* Measurement, 184:109947.

Energyprofessionals. (2022). Power Outages on the Rise in the US. Available at: https://energyprofessionals.com/power-outages-on-the-rise-in-the-us/ (2022-06-01)

Faisal, M.F., Mohamed, A. & Shareef, H. (2012) *Prediction of incipient faults in underground power cables utilizing s-transform and support vector regression*. International Journal on Electrical Engineering and Informatics, 4(2):186

Jalil, B., Leone, G. R., Martinelli, M., Moroni, D., Pascali, M. A. & Berton, A. (2019). *Fault detection in power equipment via an unmanned aerial system using multi modal data*. Sensors, 19(13):3014,

Sintef Energy. (2021). *Machine learning can predict faults in the Norwegian power system - But do we need more data sources?* Available at: https://blog.sintef.com/sintefenergy/machine-learning-can-predict-faults-and-disturbances-in-the-power-system/ (2022-06-01)

PREDIKTERING AV BEGYNNANDE FEL I ELEKTRISKA NÄT

I det här projektet belyser vi de möjligheter som kommer med maskininlärning och big data. Att prediktera fel, och få en tidig metod till underhåll och övervakning har ett stort värde för elnätsägare. Fokus i det här projektet är framtagandet av ett *Proof-of-Concept* för maskininlärningsbaserad prediktering av begynnande fel i elnät. I projektet används data från PQ-mätare i Göteborg Energis elnät. Projektet handlar om att maximera nyttan av den data som spelas in och sparas idag, ingen ny programvara behövs.

Projektets resultat visar att det redan idag, inom ramen för existerande infrastruktur, är möjligt att genomföra en mer djupgående analys av PQ-data som kan tillgängliggöra proaktiva insikter. Men samtidigt belyser också projekten avsaknaden av standardiserade metoder för hantering av PQ-data. Det finns förbättringspotential för inspelning, insamling och export.

Ett nytt steg i energiforskningen

Energiforsk är en forsknings- och kunskapsorganisation som samlar stora delar av svensk forskning och utveckling om energi. Målet är att öka effektivitet och nyttiggörande av resultat inför framtida utmaningar inom energiområdet. Vi verkar inom ett antal forskningsområden, och tar fram kunskap om resurseffektiv energi i ett helhetsperspektiv – från källan, via omvandling och överföring till användning av energin. www.energiforsk.se

Energiforsk