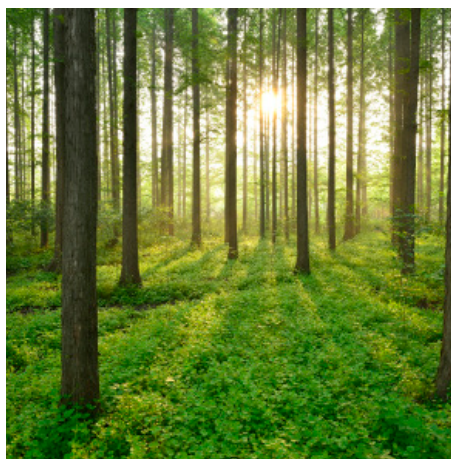
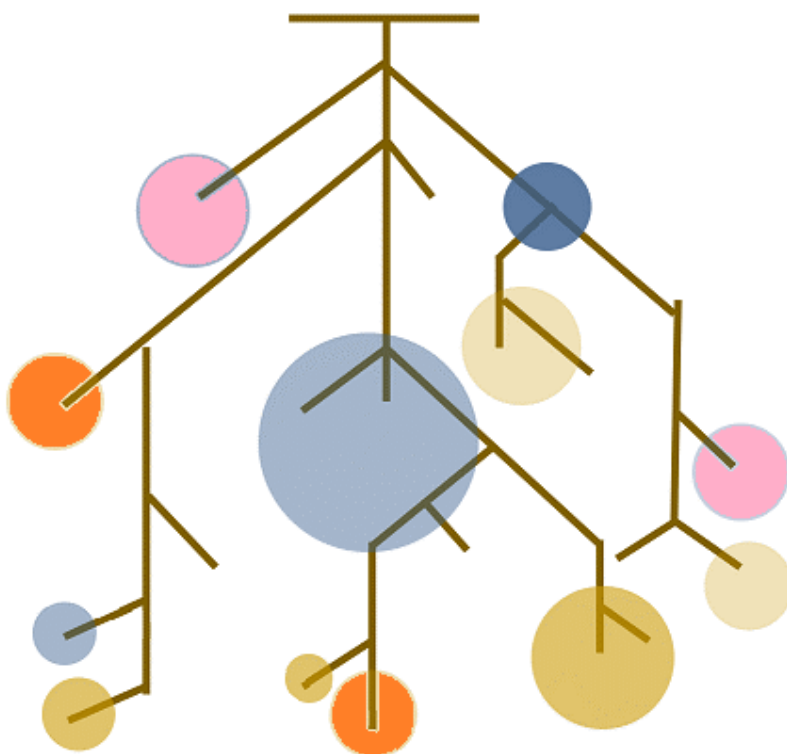


PLATTFORM FÖR PREDIKTIVT UNDERHÅLL

RAPPORT 2023:963



PLATTFORM FÖR PREDIKTIVT
UNDERHÅLL



Plattform för prediktivt underhåll

preDHiCt – Digital Platform for Predictive Maintenance
of District Heating Networks

OLOF MOGREN
OLLE PENTTINEN
MARTIN WILLBO
KRISTIN ÅKERLUND

ISBN 978-91-7673-953-.2 © Energiforsk juni 2023

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

Förord

I projektet Predhict – Plattform för prediktivt underhåll har ett proof-of-concept för prediktiv modellering utvecklats. Syftet var att undersöka om det finns en prediktiv signal i data och om det är möjligt att ge prediktioner om risker för nya skador i befintliga nät. Ett antal riskfaktorer som påverkar livslängden på ledningsnät identifierades och data från olika energibolag samlades in för att användas som indata till maskininlärningsmodellen. Modellen tränades därefter med en maskininläringsteknik som kallas för Random Forest.

Projektet har genomförts av Kristin Åkerlund på FVB, Olle Penttinen, Martin Willbo, Olof Mogren och Mattias Vesterlund på RISE Research Institute of Sweden. Julia Kuylenstierna och Katja Åström från Energiforsk har koordinerat projektet.

En referensgrupp bestående av Karolina Nygren (E.ON), Lennart Kramér (Göteborg Energi), Ronnie Sparrenholt och Anders Bylin (Karlstad Energi), Johan Barenfeld (Mölnadal Energi), Shahriar Badiei (Vattenfall), Niklas Johansson (Växjö Energi) och Magnus Ohlsson (Öresundskraft) har följt och kvalitetssäkrat projektet. Referensgruppen har också bidragit med data till modelleringen i projektet.

Projektet har finansierats av stiftelsen ÅForsk och av medverkande energibolag. Rapportförfattarna ansvarar för innehållet.

Sammanfattning

Det sägs att data är det nya guld. Här vänder vi på resonemanget och menar att vi kan spara "guld" genom att dra nytta av värdet i aggregerade dataset och på så sätt accelerera övergången till ett prediktivt underhåll av fjärrvärmenät.

Vår ursprungliga målsättning var att bygga en federerad prototypplattform för prediktivt underhåll av fjärrvärmenät. Initialt genomfördes därför en omvärldsanalys i vilken vi redogör för närliggande initiativ, belyser olikheter mellan dessa initiativ och vår egen ansats samt identifierar samarbetsmöjligheter. Sju energibolag har deltagit i en referensgrupp samt bidragit med data från deras NIS/GIS och underhållssystem. Utifrån information om tidigare skador har målet varit att ge en riskbedömning, vilken kan tolkas som en sannolikhet för skada, för varje del av bolagens befintliga ledningsnät. Projektgrupp och referensgrupp har tillsammans identifierat ett 40-tal möjliga riskfaktorer för att en skada ska uppstå på en fjärrvärmeledning. Tillgången på data har dock inneburit en begränsning, varför vi till slut fick fokusera på data från två bolag som hade möjlighet att leverera datamängder med tillräckligt många riskfaktorer för både skadade och oskadade ledningsobjekt. En maskininlärningsmodell baserad på Random Forests har tränats på denna data. Modellen ger rimliga riskestimat. Vi har visat att om data från två bolag kombineras och nyttjas för att träna en global modell presterar den bättre än en modell som tränats på ett enskilt bolags data. Vi har även simulerat en federerad inlärningsmiljö där grunddata inte blandas, men modellens inlärd parametrar ändå blir ett resultat av den sammanlagda datan. Detta ger en något sämre prediktionsförmåga i jämförelse med den globala modellen. De flesta ansatser uppvisar en F1-score över 0.7, maximalt 0.72, vilket får anses vara ett godkänt resultat baserat på de begränsade datamängderna.

Faktum är att en utökad lokal modell, där information om marktyp inkluderats med extra handpåläggning presterade något bättre än den globala modellen som saknar information om marktyp. Detta ger hopp om att en utökning av modellen med ytterligare riskfaktorer, skulle stärka den ytterligare. I första hand gäller detta relativ luftfuktighet som identifierats som en viktig faktor av referensgruppen.

När projektet summeras har vi landat i en prediktionsmodell med förmåga att estimerar risken för förekomsten av skador på ett ledningssegment. Det återstår dock flera steg till en fullödig beslutsstödsplattform för prediktivt underhåll. I slutet av rapporten stakar vi därför ut vägen för fortsatt arbete mot en sådan. Slutligen lyfter vi från projektet vikten av standardiserade former för datadelning i den här typen av samarbete. Datasäkerhet är och blir en viktig fråga framöver. Vi föreslår också branschorganisationen som en lämplig förvaltare av projektresultaten och det som eventuellt kommer därefter.

Nyckelord

Fjärrvärmenät; Random forests; riskfaktorer; federerad inläring; maskininläring; prediktivt underhåll; underhållshistorik; skattningar; estimat

Summary

It is said that data is the new gold. Here we reverse the reasoning and believe that we can save "gold" by taking advantage of the value in aggregated datasets and thus accelerate the transition to predictive maintenance of district heating networks.

Our original objective was to build a federated prototype platform for predictive maintenance of district heating networks. Initially, a market survey was therefore carried out in which we report on nearby initiatives, highlight differences between these initiatives and our own approach and identify opportunities for cooperation. Seven energy companies have participated in a reference group and contributed with data from their NIS/GIS and maintenance systems. Based on information about previous damages, the goal has been to estimate a risk index, a probability of damage, for the companies' existing district heating networks. The project group and reference group have together identified around 40 possible risk factors for damage to occur on a district heating pipe. However, the availability of data has been a limitation, which is why we finally had to focus on data from two companies that were able to deliver data sets with enough risk factors for both damaged and undamaged pipes. A machine learning model based on Random Forests has been trained on this data. The model provides reasonable risk estimates from a data science perspective. We have shown that if data from two companies is mixed and used to train a global model, it performs better than a model trained on a single company's data. We have also simulated a federated learning environment where the basic data is not mixed, but the model's learned parameters are still a result of the aggregated data. This gives a slightly worse prediction ability compared to the global model. Most approaches show an F1 score above 0.7, a maximum of 0.72, which can be considered an acceptable result based on the limited data sets.

In fact, an extended local model, where soil type information was included, performed slightly better than the global model without soil type information. This gives hope that an expansion of the model with additional risk factors would strengthen it further. One such factor is relative humidity, which has been identified as an important factor by the reference group.

To summarize, we have landed in a prediction model with the ability to estimate the risk of damage to a pipe segment. However, there are still steps to a full-fledged decision support platform for predictive maintenance. At the end of the report, we therefore outline the path for continued work towards such a goal. In summary, we highlight the importance of standardized forms of data sharing in this type of collaboration. Data security is and will be an important issue going forward. We also suggest the industry organization as a suitable manager of the project results and what may come after that.

Key words

District heat networks; Random forests; risk factors; federated learning; machine learning; predictive maintenance; maintenance history; estimates

1 Innehållsförteckning

1	Innehållsförteckning	6
2	Introduktion	8
2.1	Bakgrund	8
2.2	Syfte och mål	8
2.3	Aktörskonstellation	9
3	Omvärldsanalys	10
3.1	Genomförande	10
3.2	Resultat från omvärldsanalys	10
3.2.1	Smart energi	11
3.2.2	Gradyent	11
3.2.3	Danfoss Leanheat	12
3.2.4	Kamstrup	12
3.2.5	Cenosco	12
3.2.6	RörANN/Vass AI	13
3.2.7	RISEs och Göteborg Energis elnätsprojekt	13
3.2.8	NODA Intelligent Systems	14
3.2.9	Högskolan i Borås, DAD	14
3.2.10	CARE-W LTP och KANEW	15
3.2.11	Utilifeed	15
3.2.12	SAB, SAM och BSAM	15
3.2.13	Examensarbete hos Öresundskraft - Luftfuktighet	16
4	Önskemål och visioner	17
4.1	Kravspecifikation	17
4.1.1	Genomförande	17
4.1.2	Resultat från arbete med kravspecifikation	17
4.2	Gapanalys	17
4.2.1	Genomförande	17
4.2.2	Resultat från gapanalys	18
5	Koncept- och modellutveckling	19
5.1	Datainsamling	19
5.2	Genomförande	20
5.3	Resultat från modellutveckling	22
5.3.1	EON	23
5.3.2	Göteborg	26
6	Utveckling av prediktiv modell genom federerad inlärning	29
6.1	Genomförande	29
6.2	Resultat	29
6.2.1	Global modell	29
6.2.2	Federerad modell	34

7	Diskussion och slutsatser	40
8	Fortsatt arbete	43
9	Referenser	45
10	Bilagor	47
10.1	Korrelationsfigurer	47
10.2	Resultatöversikt	47
10.3	Källkod	47
10.4	Utdrag Predicerad risk ledningsobjekt	48

2 Introduktion

2.1 BAKGRUND

Fjärrvärmebranschen har en gemensam utmaning i att förvalta och hantera åldrande ledningsnät. Det är viktigt att optimera ledningsnätens livslängd och prioritera rätt underhållsinsatser för att uppnå hög försörjningstrygghet, energieffektivitet, resursanvändning, säkerhet och god ekonomi.

I Sverige började fjärrvärmenät byggas i slutet på 40-talet (Fjärrvärme - A real success story, 2009) och ökade i samband med oljekriserna och utbyggnaden av bostäder inom Miljonprogrammet. Idag finns cirka ca 62 000 km ledningsnät bestående av ledningar av olika material, konstruktion, storlek och ålder. Det totala värdet av distributionsnätet har uppskattats till 150 miljarder kronor (Andersson, o.a., 2015) varför korrekt förvaltning av näten är en viktig fråga även ur ett ekonomiskt perspektiv.

Idag finns olika metoder och tekniker för att upptäcka och identifiera skador i fjärrvärmeledningar så som fuktlarmsövervakning, termografering och snösmältning. Dessa metoder används för att identifiera läckage som redan uppstått, men används även till statusbedömning för att hitta skador som kan åtgärdas innan de leder till läckage på medierören.

Att utveckla metoder för effektiv statusbedömning har varit ett prioriterat område inom forskning- och utveckling, se till exempel (Sörensen & Penttinen, Utvärdering av icke-förstörande mätmetoder för lokalisering av skador på rörledningar , 2019), (Malm, Mokhlesi, Sernhed, & Yarahmadi, 2016) (Andersson, Molin, & Pletikos, 1999) (Sederholm, 2021) (Ohlsson, Smart Aktiv Box Smart fjärrvärmenät, 2022). Databaserade metoder för att lokalisera fel i undercentraler har presenterats av ett par forskningskonsortier (Wästberg, Hansson, & Edland, 2022) (Mbiydzennyuy G. , 2022) där det senare drogs slutsatsen att anomalidetektering i undercentraler hade underlättats av tillgång till nätinformation. Frågeställningarna angränsar till detta projekt och frågan är om dessa projekt liksom vårt hade gynnats av tillgång på både mät- och nätdata.

Ett antal kommersiella initiativ har etablerats de senaste åren. Affärsmodeller bygger till viss del på helhetskoncept, där kombinationen av bolagens hård- och mjukvara sägs ge högre precision i analysresultaten.

2.2 SYFTE OCH MÅL

Svensk ordbok förklarar ordet *prediktiv* med "som har att göra med (vetenskaplig) förutsägelse" och syftet i det här projektet har varit att utveckla en gemensam plattform för prediktivt underhåll. Tanken är att nätägare kan använda den som ett beslutsstöd för välavvägda riskklassningar och underhållsinsatser. Det är önskvärt att underhålla ett ledningssegment i rätt tid för att maximera den tekniska livslängden.

Målet med det här projektet har varit att demonstrera konceptet genom att ta fram en första version av plattformen för att visa dess funktion.

Inom projektet har fem delmål satts upp, dessa beskrivs nedan.

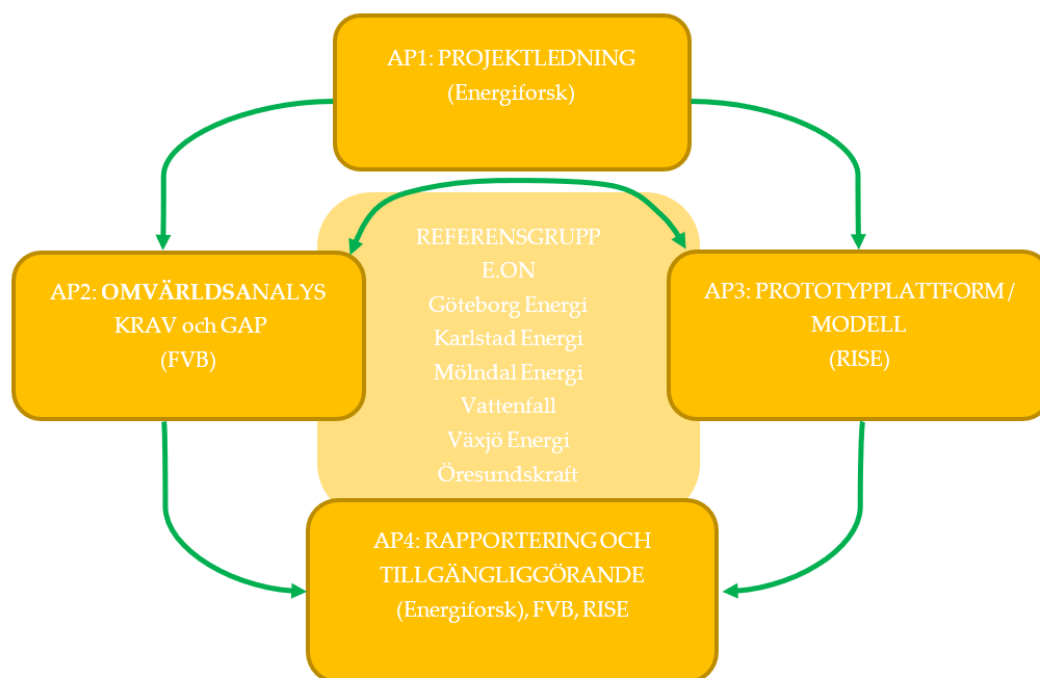
- Delmål 1: En sammanställning av liknande initiativ och samarbetsmöjligheter inom prediktivt underhåll av fjärrvärmenät
- Delmål 2: En kravspecifikation för plattformen och en gap-analys som visar vilken utveckling som behövs för att möta kraven
- Delmål 3: Konceptet demonstreras och implementeras på deltagande energibolags data
- Delmål 4: Resultatspridning bl.a. genom en projektrapport
- Delmål 5: En färdplan för uppskalning, utveckling och användning av plattformen

Målet inom ramen för projektet var att plattformen skulle nå en teknisk mognadsgrad 6 på skalan *Technology readiness level* (TRL), vilket innebär att en modell eller prototyp av systemet har demonstrerats under verklighetsliknande förhållanden.

2.3 AKTÖRSKONSTELLATION

Projektet har bedrivits i form av fyra arbetspaket (AP), där det övergripande AP1 innefattat projektledning. Här har projektet koordinerats av Energiforsk i samarbete med FVB och RISE. Inom AP2 för vilket FVB ansvarat har fokus legat på delmål 1 och 2, dvs omvärldsanalys, kravspecifikation och gap-analys.

I AP2 enades alla projektpartners också kring en lista med riskfaktorer som låg till grund för fortsatt arbete med prototypplattformen. AP3 har dedikerats åt delmål 3, att implementera en prototyp av ett beslutsstödsystem, eller modell som det faktiskt blev vilket vi beskriver senare i rapporten. För AP3 har RISE ansvarat. AP4 utgörs av rapportering och resultatspridning. Projektet har genomförts tillsammans med en referensgrupp bestående av sju energibolag. Dessa bolag har delat med sig av önskemål och synpunkter på en framtida beslutstödsplattform samt erfarenheter från skador och historiskt underhållsarbete.



Figur 1: Aktörer och arbetspaket. Gröna pilar indikerar interaktion mellan arbetspaket. Angivet bolag i parentes indikerar AP-ledare. Referensgruppens har involverats i AP2, AP3 och AP4.

Alla bolag i referensgruppen bidrog aktivt till AP2s genomförande, framför allt i arbetet med kravspecifikationen och riskfaktorer. Fem av sju bolag valde att förse projektet med data utifrån denna riskfaktorlista, men endast två bolag kunde erbjuda data av sådan typ och omfattning att det ansågs möjligt att använda i projektet. Bolagen har också bidragit i AP4, framför allt kring rapporteringen.

3 Omvärldsanalys

3.1 GENOMFÖRANDE

Det finns två primära syften med att kartlägga projekt och produkter med liknande mål som detta projekt. Dels att säkerställa att projektet inte skapar något som redan finns, dels att hitta potentiella samarbeten. Det sistnämnda då projektets långsiktiga strävan är att utveckla ett färdigt verktyg som kan användas av fjärrvärmebolag. Ingen av de aktörer som varit involverade i projektet är inriktade på de sista stegen i en produktutvecklingsprocess eller förvaltning.

Flera angränsande initiativ har identifierats. För att undvika begreppsförvirring skiljer vi på expertsystem och datadrivna system som bygger på maskininlärning. I ett expertsystem definierar en tillräckligt erfaren person, systemets beslutsregler genom att bestämma vilka faktorer och kombination av faktorer som påverkar tex ledningens status. Det är därför starkt beroende av expertens erfarenhet baserat på tillgänglig statistik eller tidigare forskningsresultat. Ett sådant angreppssätt har exempelvis använts i en reinvesteringsmodell för Göteborg Energis fjärrvärmenät (Åkerström, 2004).

Maskininlärning (ML) är ett datadrivet angreppssätt där statistik används för att lära vilka faktorer och kombination av faktorer som påverkar ledningens status och hitta beslutsreglerna. ML behöver en stor mängd data/statistik för att systemet ska bli så bra som möjligt. Precis som experten har lärt sig av teknikers och produkters styrkor och svagheter genom åren behöver också ML tränas på data som innefattar ett spann av goda och dåliga exempel (i vårt fall hela och trasiga ledningar). En ML-modell tränas för att generera en prediktion inom sitt träningsområde. Den kan därför inte säga något om sannolikheten för en skada för en ledningstyp i en viss typ av system eller med omgivningsfaktorer som är vitt skilda från dem den tränats på.

Det är viktigt att vara medveten om att digitala system (exempelvis de baserade på ML) kan ha fel trots att de kan "låta" säkra. Dessa digitala system kan aldrig utgöra "facit". De prediktioner, estimat eller resultat som databaserade ML-system genererar behöver liksom andra system alltid ifrågasättas och bedömas av experter.

Sammanställningen innefattar projekt och produkter som utförargruppen och deltagande energibolag har kännedom om genom sitt arbete inom detta område, inom och utanför projektet. Utöver det har kompletterande internetsökningar genomförts. Information om respektive initiativ har hämtats från forskningsrapporter, offentlig information via hemsidor och samtal.

Sammanställningen inkluderar projekt och produkter som fokuserar på digitala lösningar inom fjärrvärmedistribution eller andra branscher som förvaltar infrastruktur. Fältbaserade mätmetoder och visuella kontroller för förebyggande underhåll och statusbedömning av ledningsnät har exkluderats (tex termografering, fortlöpande tillsyn, fuktalarm). Däremot finns det förhoppningar om att data från statistik och mätningar med sådana metoder på sikt ska kunna inkluderas i de algoritmer som utvecklas i det här projektet.

3.2 RESULTAT FRÅN OMVÄRLDSANALYS

Nedan följer en beskrivning av de initiativ som identifierats och som på ett eller annat sätt fokuserar på liknande frågeställningar som preDHiCt.

3.2.1 Smart energi

Inledning

Smart Energi är en branschsamverkansorganisation bestående av 12 energibolag (Borås Energi och Miljö, E.ON, Göteborg Energi, Halmstads Energi och Miljö, Hässleholm Miljö, Kungälv energi, Mölndal Energi, Sandviken Energi, Sundsvall Energi, Tekniska verken i Linköping, Västra Mälardalens Energi & Miljö och Öresundskraft). Organisationen bildades för att gemensamt och i samarbete med akademi och forskning utveckla digitala system och lösningar för energibranschen.

Samarbetet är fortsatt aktivt och medlemsbolagen inom Smart energi håller på att bilda en ekonomisk förening.

Teknisk beskrivning

Inom Smart energi har forsknings- och utvecklingsprojekt genomförts, bl.a. projektet Branschsamarbete för avancerad analys av värmedistribution och uppvärmningsbehov, Data Science:BRAVA förkortat DS:BRAVA. Projektet finansierades av Smart energi, Energimyndigheten och Energiforsks och utfördes av Solita. Inom projektet vidareutvecklades analysverktyget K2. En Data Science-portal tillhandahölls för utvecklare, forskare och behovsägare. Metoder baserade på AI (artificiell intelligens), DL (djupinlärning) och ML (maskininlärning) användes för att detektera avvikelser och fel i kundens anläggning, tex styr-/reglerfel och mätarfel. Även algoritmer för att identifiera ledningsläckage har utvecklats. Algoritmer och modeller utvecklades med hjälp av deltagande bolags gemensamma data och ett ramverk för datadelning och benämning av data (taxonomi) togs fram inom Smart energi. Exempel på data som analyserats i projektet är kombinationer av flöde, fram- och returtemperatur från fjärrvärmekundernas värmemätare och utomhustemperatur. Inom Smart Energi undersöktes också inverkan på effektbehovet i fastigheter och hur det korrelerade med daggpunkt, vind etc.

3.2.2 Gradyent

Inledning

Gradyent är ett nederländskt bolag som erbjuder digitala lösningar för fjärrvärmebolag. En digital tvilling som återspeglar produktions- och distributionssystemet samt slutanvändare, har utvecklats. Den används för att optimera temperaturnivåerna i systemet, kontrollera trycknivåer, underlätta produktionsplanering, upptäcka fel och arbeta med förebyggande underhåll, kontrollera användardata samt för att designa och simulera framtida fall och stödja beslutsfattande. Geografisk information, väderdata och sensordata kombineras med fysikaliska modeller och artificiell intelligens.

På sin webbsida (Gradyent, 2023) uppger de att över 15 nät i Europa använder deras temperaturoptimering.

Teknisk beskrivning

För att möjliggöra prediktivt underhåll beskrivs hur den digitala tvillingen i realtid rapporterar om värmeförluster, stora tryckfall, för låga returtemperaturer och möjliga omdirigeringar för att lösa flaskhalsar. Gradyent erbjuder möjlighet att installera tryck- och temperatursensorer vid produktionsanläggningar och i nätet. Detta nyttjas med övriga data från kundcentraler och sensorer i ledningsnätet. Genom att tillhandahålla visualiseringar av systemets status, baserat på data från ovanstående sensorer, ska den digitala tvillingen bidra till effektiv planering av underhållsåtgärder.

Information om Gradyents produkter har hittats på företagets egen webbplats (Gradyent, 2023). Detaljerad teknisk information finns inte offentlig och det går inte att utläsa hur algoritmerna för prediktivt underhåll är uppbyggda.

3.2.3 Danfoss Leanheat

Introduktion

Danfoss Leanheat är mjukvaror och tjänster för att styra och optimera fjärrvärmeproduktion och distribution liksom byggnader och bostäder. Modulen Leanheat Network är ett modelleringsverktyg som utvecklats för att stödja planering, design och drift i fjärrvärmesystem med minskning av investerings- och driftkostnader som mål (Danfoss, 2023).

Beskrivning

Leanheat Network består av en termohydraulisk modell för att simulera nätet. Modellen är uppbyggd genom information om produktionsanläggningar, distributionssystem och dess komponenter. Modellen kan användas för att optimera utbyggnad och renowringar samt analysera effekten av dessa åtgärder. Vidare kan den användas för utveckling av beredningsplaner och utgöra en databas med kunskap om näten.

3.2.4 Kamstrup

Introduktion

Kamstrup har utvecklat en analysplattform (Kamstrup, 2023) som sägs fungera bäst tillsammans med Kamstrups egen värmemätare (Säljare Kamstrup, 2018), men att den även går att nyttja tillsammans med andra fabrikat. Mjukvaran är främst tänkt att användas för övervakning/anomalidetektering i fjärrvärmesystem.

Kamstrup är ett privatägt större aktieföretag med en bred kundbas. De räknar med att åtminstone 75% av omsättningen sker på marknader utanför Norden inom en 10-årsperiod.

Beskrivning

Heat Intelligence förenar fakta om nätet (rörlängder, dimensioner, isolering m.m.) med data från olika typer av mätare i systemet. Den uppdateras automatiskt med nya data. Identifierar värmeförluster och rundgångar, anomalier baserat på temperaturer och flöden. Heat Intelligence kombinerar en nätmodell med mätvärden och kan på så sätt identifiera värden som avviker från det modellerade förväntade beteendet. Deras analysresultat kan potentiellt nyttjas som träningsdata för vår typ av ML-algoritmer. I förlängningen skulle vår algoritm då kunna komplettera Heat Intelligence i nät där erforderligt data saknas.

3.2.5 Cenosco

Inledning

Cenosco tillhandahåller en analysplattform som kallas Integrity Management Solutions Suite (IMS). Målgruppen är företag inom processindustrin som kemi, olja, gas, papper och förnybar energi, specifikt med behov inom infrastruktur för högtryck. Inom plattformen finns teknik för att estimerar risk hos komponenter i olika typer av system som hanterar trycksatta medier.

Cenosco grundades år 2000 i Nederländerna.

Teknisk beskrivning

Ingen teknisk information finns tillgänglig från leverantören.

3.2.6 RörANN/Vass AI

Inledning

Stockholm Vatten och Avfall (SVOA) publicerade 2019 ett examensarbete (Rehn, 2019) där en modell för att prediktera status för olika delar av Stockholm stads vattenledningsnät utvecklats. Forsknings- och utvecklingsbolaget Sweden Water Research (SWR, 2023) som samägs av tre skånska VA-bolag koordinerade 2019-2021 ett uppföljningsprojekt med syfte att tillgängliggöra projektresultaten till fler VA-bolag. SWRs projekt (Sörensen & Nilsson, Ordning i RörANN, 2019) finansierades av Smart Built Environment. Parallellt med detta har systemmodellen VASS AI (Karlsson, 2022) utvecklats av branschorganisationen Svenskt Vatten, med samma syfte.

VASS AI finns tillgänglig för VA-bolag utan kostnad.

Teknisk beskrivning

Med fokus på VASS AI kan den sammanfattas som en omarbetad variant av SVOAs prediktionsmodell. Den tillgängliggörs kostnadsfritt för Sveriges VA-organisationer via Svenskt Vattens hemsida. Modellen är ett artificiellt neuralt nätverk (ANN) som predikterar ett konditionsvärde för ledningar som spänner från 0 till 1 där ett högt värde motsvarar "dålig" status. Det intressanta här är dels den slående likheten med projektets målsättning i förhållande till vår tillämpning, men det är också intressant att se hur Svensk Vatten som branschorganisation ser värdet i att förvalta och tillgängliggöra den här typen av verktyg. VASS AI kräver information om lednings-ID, byggår, dimension, längd, material, jordart, växtzon samt i vissa fall kommunkod. Det förväntas att användaren följer en standardiserad nomenklatur för exempelvis ledningsmaterial, jordart, växtzon och kommunkod.

3.2.7 RISEs och Göteborg Energis elnätsprojekt

Inledning

Göteborg Energi och RISE samverkade 2021–2022 i ett gemensamt forskningsprojekt rörande statusestimering av elnät under 10 kV (Göteborg Energi, 2020). Projektet finansierades av Göteborg Energis forskningsstiftelse. Medfinansiärer som listats i projektets ansökan är; Herrljunga Elektriska, Umeå Energi, Alingsås Energi, Mölndal Energi, Göteborg Energi Nät AB.

Syftet med projektet var att med maskininlärning och data om tidigare avbrott i elnätet bedöma vilka markkablar som borde prioriteras i elnätbolagets underhållsarbete. De resulterande maskininlärningsmodellerna skulle beräkna kablarnas hälsa vilket indirekt skulle ge en fingervisning om återstående livslängd. Detta för ett effektiviserat underhåll som minskar kostnader och avbrottstider för kunderna.

Teknisk beskrivning

Frågeställningen är väldigt lik vår, men förutsättningarna och tillgången på data skiljer sig åt vilket gör att också angreppssätten blir olika. RISE nyttjade en klassificeringsmodell som gick att träna relativt väl på den data de hade tillgång till i projektet. De gjorde även en ansats att titta på korrelationer mellan indata och uppkomna skador för att se hur olika faktorer påverkat skaderisken. Skador som uppkommit senare (ett 10-tal) hade fått en relativt hög riskklassning med den tränade modellen. Samarbetet med Göteborg Energi samt deras djupa domänkunskap var till stor hjälp enligt RISEs utvecklare Erik Weihs. Erik var i vår intervju ödmjuk i att det kan finnas andra modeller som presterar bättre för vår fjärrvärmertilämpning än just den han valt för 10 kV-nätet.

Estimeringen utgår från tidigare avbrott som orsakats av föråldrade och felaktiga kablar, tidigare skarvade kablar, ej gynnsamma markförhållanden och andra riskfaktorer. Projektet avgränsades till markkablar under 10 kV.

3.2.8 NODA Intelligent Systems

Inledning

NODA Intelligent Systems är ett företag som erbjuder AI-baserade lösningar för termiska system som fjärrvärme och fjärrkyla, värmepumpar och gas. För fastigheter erbjuds produkter för optimering av energianvändning mot ett önskvärt inomhusklimat. Både mängden energi och när i tiden energin används styrs för att sänka energikostnaderna. NODA erbjuder också produkter för energileverantörer. Dessa produkter bygger på dynamisk kontroll av tillgång och efterfrågan av energi genom laststyrning och optimering av framledningstemperaturen. (Noda, 2023)

NODA erbjuder färdiga produkter både till fastighetsbolag och fjärrvärmebolag och har partners och kunder i Europa, Nordamerika och Asien.

Teknisk beskrivning

I lösningarna ingår sensorer för temperatur och luftfuktighet för att möta inomhusklimat, väderprognoser och modeller av uppvärmningssystemet.

3.2.9 Högsolan i Borås, DAD

Inledning

Avdelningen Computer Science Lab på Högsolan i Borås drev 2018–2022 ett projekt kallat Data Analytics for Fault Detection in District Heating, förkortat DAD (Mbiydzenyuy G. , 2022). AB Bostäder i Borås, Borås Energi och Miljö, samt Noda Intelligent Systems, deltog också som partners i projektet. Tre målsättningar hade identifierats inom ramen för projektet; att skapa metoder för online-detektering av onormalt beteende på konsumentnivå, att identifiera suboptimerade undercentraler i fjärrvärmenät samt att klassificera upptäckta avvikelser.

Teknisk beskrivning

På det avslutande seminariet presenterades metoder för anomalidetektering i tidsserier. Där hade DL (djupinlärning) nyttjats för prediktioner och klassificering. Tydliga skillnader presenterades för olika undercentraler och det hade gått att dela in dem i kluster, men skillnader i karaktäristik mellan olika fastighetstyper gjorde det svårt att urskilja anomalier från datamängderna.

Projektet hade också ett arbetspaket som fokuserade på projektledning och forskningsmiljö. Under slutseminariet tog diskussionerna fart mellan närvarande forskare på slutseminariet. Det rörde konsensus i att tillgången på data ofta blev en begränsande faktor i forskningsprojekt, samt att det fanns en potential i att inkludera fler datakällor. DAD hade exempelvis saknat information om ledningsnätet och fastigheterna. Tillgång till detta hade förmodligen ökat modellens noggrannhet. Forskarkonsortiet publicerade också en nulägesbild för maskininlärning inom fjärrvärmeområdet. De visade att de flesta ML-relaterade artiklar fokuserar på prognostisering av energibehov. Avvikelser och läckagedetektering återfanns på andra plats, men dessa studier var relativt få i jämförelse (Mbiydzenyuy, o.a., 2021). En teknik de lyfte i samband med detta, som kan utgöra ett komplement till preDHiCt, kombinerar en hydraulisk modell med signaler från sensorer i distributionsnätet för att identifiera avvikelser (Xue, o.a., 2020).

3.2.10 CARE-W LTP och KANEW

Inledning

CARE-W är ett EU-finansierat projekt som syftar till att utveckla metoder och mjukvara för att möjliggöra effektiv förvaltning av vattenledningar. Den slutliga rapporten publicerades 2005 och information om projektet finns på projektwebben (Sintef, 2023). Elva parter från åtta olika länder utförde projektet tillsammans med 13 slutanvändare i Europa. Projektets mål har varit att skapa ett ramverk för beslutsfattande för underhåll av vattenledningsnät, såsom VA- och fjärrvärmeledningar. Slutprodukten är ett beslutsstödsystem, *Decision Support System (DSS)*, för att göra rätt underhållsåtgärd på rätt ledning vid rätt tidpunkt, dvs innan en ett fel inträffat. De huvudsakliga leveranserna från projektet är en prototyp (KANEW) och en handbok.

Resultaten från forskningsprojektet har tagits vidare och kommersialiserats under namnet KANEW 3S (3S Consult GmbH, 2023)

Teknisk beskrivning

KANEW 3S är ett expertsystem för planering av underhållsåtgärder på vattenledningsnät. Programvaran uppskattar det nuvarande och framtida tillståndet för vattenledningar, d.v.s. prestandaindikatorer, predikteringar av fel i ledningsnätet och beräkning av pålitlighet för vattenförsörjning. Vidare ingår rutiner för att uppskatta långsiktiga investeringsbehov samt urval och rangordning av underhållsprojekt. Dessa verktyg är integrerade i CARE-W-prototypen.

Indata till modellen är GIS, information om ledningsnäten, SCADA, ekonomi, hydraulisk modell och viss annan information. Ut från modellen fås en prioriteringslista bland objekt som behöver underhållas och strategiska planer och kostnader för framtida underhåll.

3.2.11 Utilifeed

Inledning

Utilifeed är ett privatägt aktieföretag som grundades 2016. De har utvecklat en plattform för analys och optimering av fjärrvärme som de erbjuder fjärrvärmebolag. Plattformen har moduler för produktion, distribution, försäljning och marknad samt mätdata. Till exempel inkluderas lastprognoser och en digital tvilling för produktionsoptimering, dimensionering och övervakning av distributionssystemet. Bolaget har deltagit i ett antal forskningsprojekt, där fokus bl.a. legat på digitaliserad nätövervakning vilket gör att de befinner sig i gränslandet mot vad vi avser göra inom preDHiCt (Kensby, 2022).

Ett flertal energibolag använder plattformen, främst svenska.

Teknisk beskrivning

Utilifeed har utvecklat en tjänst där AI/ML används för att spegla verkligheten och analysera fjärrvärmesystem. Resultatet från analyserna ska möjliggöra en optimerad dimensionering samt identifiera avvikelser och fel hos kund. Den molnbaserade plattformen ska beräkna kundernas dimensionerande effektbehov, möjligheten till nätlagring och lastprognoser. Den ska även genom övervakning möjliggöra upptäckt av mätfel och läckage.

3.2.12 SAB, SAM och BSAM

Inledning

Ett mindre forskningskluster med energibolaget Öresundskraft och teknikutvecklaren Arne Jensen AB som de centrala aktörerna har under senare år arbetat för att utveckla metoder som upptäcker den primära skadan i ett tidigt skede. Om dessa fel åtgärdas vid upptäckt kan ledningarnas livslängd maximeras eftersom följdskador som bl.a. leder till medierörsläckage minimeras. I klustret har även

konsultbolag inom fjärrvärme, energi och företagsorganisation såsom Enelex (numera upplöst) och Indepro samt Sweheat and Cooling ingått.

Teknisk beskrivning

Projektet Smart Aktiv Box – SAB - Smart fjärrvärmenät (Ohlsson, 2022), finansierades av Energimyndigheten 2018–2021. Här skedde mycket av grundutvecklingen till det expertsystem Öresundskraft förfogar över idag. Detta system bygger delvis på de sensorboxar som utvecklats av Arne Jensen AB som mäter stålrensens tjocklek samt ett antal miljöparametrar (fukt, temperatur etc.). Tolkning av mätvärden sker i Power BI.

Smart Asset Management, SAM, ett Vinnovafinansierat UDI2-projekt, pågick under 2018 – 2021 som ett samverkansprojekt mellan energibolag, företag, universitet och högskolor. Projektet består av flera delar men i det arbetspaket som ligger närmast preDHiCt har arbetet med ovan nämnda SAB-system fortsatt. Det finns också en koppling till DAD (avsnitt 3.2.9 ovan) då deras nulägesanalys för ML i relation till fjärrvärme delfinansierats av SAM (DIVA Digitala Vetenskapliga Arkivet, 2021).

Slutligen ingår också arbete med SAB-systemet i Interreg South Baltic samarbetet Baltic Smart Asset Management, BSAM (Hogland, 2023). Här utgör Öresundskrafts implementering av systemet (Ohlsson, 2021) ett av två tekniska pilotfall som redovisats i projektet. Syftet är att visa på tekniska möjligheter att säkra en hög leverans kvalitet och hållbar resursanvändning genom att utveckla ett beslutsstödsystem som möjliggör statusbedömning vilket underlättar planering av underhållsarbeten och reinvesteringar. Det tjänar sannolikt som inspiration för andra fjärrvärmeleverantörer i regionen och ökar samtidigt möjligheterna för framtida internationalisering av svensk teknik.

3.2.13 Examensarbete hos Öresundskraft - Luftfuktighet

Inledning

På uppdrag av Öresundskraft genomförde Olof Olsson, Lunds universitet, ett examensarbete (Olsson, 2022). Arbetet slutredovisades 2022-06-01. Syftet med arbetet var att utveckla programmet Power BI till att bli ett beslutsstödsystem för fjärrvärmedistribution genom att använda data från tidigare nämnda sensorboxar (SAB) för att göra bättre statusbedömningar av fjärrvärmenätet.

Teknisk beskrivning

I projektet används ledningstyp, luftfuktighet och temperatur i kammare, utomhustemperatur samt nederbörd för att programmera Power BI till att identifiera samband mellan luftfuktighet och påverkan i olika typer av fjärrvärmeledningar. Resultaten visade att förändringshastigheten i kammarens luftfuktighet kunde användas för att avgöra om ett inläckage skett via betäckning eller genom kammare/kulvert. Olika kammare- och ledningstyper påverkade korrelationen mellan kammarens luftfuktighet och den relativa luftfuktigheten utomhus. Detta kunde kopplas till delsystemets ventilationsförmåga. Detta möjliggör att i ett tidigt skede utföra rätt underhållsåtgärder på rätt plats.

I arbetet har olika nyckeltal som kan användas för att jämföra kammare eller kammartyper identifierats, samt vilken koppling dessa nyckeltal har till förekomsten av fuktbetingad skada. Den här typen av nyckeltal skulle potentiellt kunna ingå som ytterligare indata i den typ av modell som tas fram inom preDHiCt.

I arbetet utvecklades ett expertsystem där flera intressanta samband identifierades, ett förslag på fortsatt arbete från examensarbetet var att utveckla analysmodellen med hjälp av maskininlärning.

4 Önskemål och visioner

4.1 KRAVSPECIFIKATION

4.1.1 Genomförande

Kraven på plattformen fastställdes genom de önskemål om funktioner för plattformen som togs fram och prioriterades under tre workshops om prediktivt underhåll av fjärrvärme inom Energiforsks värmekluster (WS1 maj 2019, WS2 oktober 2019 och WS3 februari 2020). Diskussionerna fördes med en slutlig beslutsstödsplattform i åtanke vilket återspeglas i summeringen av kravspecifikationen nedan. Projektgruppen och referensgruppen utvecklade och prioriterade gemensamt krav för en färdig plattform och identifierade tillsammans rimliga målsättningar för projektet. Resultaten redovisas i sammanfattad form nedan.

4.1.2 Resultat från arbete med kravspecifikation

Plattformen ska användas som ett beslutsstöd vid planering av underhållsåtgärder i fjärrvärmeledningsnät. För att bolagen ska kunna agera och planera sitt underhållsarbete med stöd av denna plattform behöver några ska-krav vara uppfyllda. Referensgruppen definierade också ett antal bör-krav, vilket sågs som rimliga förväntningar på den slutliga plattformen och då avser vi den plattform bolagen avser att använda, inte nödvändigtvis inkluderat i den prototyp som utvecklats inom ramen för preDHiCt. Sist framfördes också ett par önskemål om utökad funktionalitet.

Referensgruppens ska-krav innefattade att plattformens algoritmer skulle prediktera sannolikheten för en skada på ledningssegment. De förväntade sig också en estimering av återstående livslängd hos ledningssegmenten. Modellen skulle utöver att ingå i en distribuerad bolagsgemensam infrastruktur också kunna användas lokalt utan risk för oavsiktlig datadelning. Plattformen skulle kunna användas som ett fristående verktyg och den skulle också återge information om ledningsnäten (historiska åtgärder, etc.).

Det fanns förväntningar på att plattformen borde kunna visualisera modellresultat. Dessa resultat borde också härledas, för att höja tilltron och förståelsen för estimaten. I förlängningen bör plattformen kunna kvantifiera eller redovisa konsekvenser av incidenter och det borde finnas möjlighet att addera datalager till befintlig modell lokalt om det råder lokala avvikelser kring tillgängliga datalager bland användarna.

Referensgruppen framförde också önskemål om att ett estimat av ekonomiska konsekvenser av åtgärder skulle inkluderas. De önskade en kontinuerlig övervakning av ledningsnäten med tillförsel av nya data och uppdatering av tränade algoritmer.

Generellt har vi också behövt förhålla oss till icke-funktionella aspekter som kompatibilitet, användarvänlighet, uppdateringsmöjligheter, skalbarhet och säkerhet.

4.2 GAPANALYS

Syftet med gapanalysen var att identifiera nödvändiga åtgärder för att skapa en plattform för beslutsstöd i enlighet med referensgruppens kravspecifikation. Den belyser också skillnaden mellan det läge som rådde vid projektets start samt det läge vi nu befinner oss i vid projektslut.

4.2.1 Genomförande

Projektgruppen sammanställde en gapanalys utifrån de krav som identifierades i workshops och förberedande samtal som hållits med bolagen inför projektstart. Gapanalysen bröts ned på

arbetspaketsnivå och vi identifierade sedan nödvändiga åtgärder för att överbrygga gapen. Vissa åtgärder prioriterades, inte enbart utifrån referensgruppens kravspecifikation. Även tekniska och administrativa förutsättningar har påverkat projektets prioriteringar. GAPet för varje arbetspaket redovisas i avsnittet nedan.

4.2.2 Resultat från gapanalys

AP1: Projektets övergripande målsättning har varit att etablera en POC (proof of concept) för en unik samarbetsplattform i syfte att erbjuda ett beslutsstöd för prediktivt underhåll av fjärrvärmenät. Tre nycklar för denna POC var dels att förstå bolagens behov, krav och förväntningar på en plattform, dels att utreda hur bolagens data skulle kunna hanteras i projektet och dels hitta former för hur ytterligare datakällor skulle kunna nyttjas. Bolagens behov och krav togs fram och sammanställdes i en kravspecifikation i arbetet som beskrivs i avsnitt 4.1 ovan. I inledningen av projektet hölls även ett antal möten för att enas kring former för datadelning. Inledningsvis enades också ingående parter i ett gemensamt partneravtal, en process som tog längre tid än beräknat.

AP2: Ett mål var att genomföra en omvärldsanalys av aktuella initiativ inom prediktivt underhåll för fjärrvärmenät. Detta för att identifiera lämpliga aktörer för samarbete, vidareutveckling eller förvaltning av POCen samt att säkerställa att projektet inte skapar något som redan finns. Omvärldsanalysen är genomförd och sammanställd i avsnitt 3 ovan. Även POCens indata, dvs de riskfaktorer som påverkar ledningarnas livslängd, behövde identifieras. Arbetet genomfördes i samarbete med referensgruppen och sammanställdes i en lista, se avsnitt 5.1 nedan.

AP3: När det kom till utvecklingen av själva POCen syntes ett antal utmaningar i utgångsläget. En förutsättning för ett lyckosamt genomförande var tillgången på data. De utmaningar projektgruppen identifierade handlade om datadelning och hanteringen av bolagens data på ett tillförlitligt sätt. Under avtalsprocessen framkom att bolagens organisations och IT-strukturer skiljer sig åt, att bolagen sinsemellan hade olika inställning till att dela data, att de har olika inställning till säkerhetsarbetet och dataskydd, att de värderar känsligheten i samma egenskaper i data olika samt att de har tillgång till olika data. Projektteamet har arbetat för att definiera skillnaderna och hitta ett sätt att möta bolagen i detta. Bolagen har försökt involvera personer med rätt mandat och kompetenser, avtalet har reglerat eventuella skillnader och bolagen har tillåtits vara med under olika premisser. De bolag som varit mest öppna när det kommit till datadelning och bidragit med mest relevant data har prioriterats. När det kommer till att kvalitetssäkra plattformen mot kravspecifikationen är projektet inte i hamn i alla delar. Vi är ännu inte i ett läge där vi har en federerad distribuerad samverkansplattform. Mycket till projekttiden har gått åt till att definiera former för datadelning samt även att modifiera RISEs IT-infrastruktur för att uppfylla de säkerhetskrav som specificerats i projektavtalet. Projektteamet har i stället fokuserat på att bygga en prediktionsmodell med så hög noggrannhet som möjligt, samt att testa den i en artificiell federerad inläringsmiljö. Vår bedömning är därför att vi landat i TRL 4 istället för 6, som var vår ursprungliga målsättning.

AP4. 1 Resultatspridningen ser vi från projektgruppen som nödvändig för att tillgängliggöra forskningsresultaten. Målsättningen var att publicera material i FVBs nyhetsbrev (genomfört), samt på respektive organisations hemsidor. I dagsläget finns information om preDHiCt publicerad på Energiforsks websida (Kuylenstierna, 2022). Slutrapporten bör rimligtvis också publiceras via dessa kanaler.

AP4. 2. För att lyfta projektets resultat från TRL 6 till TRL9 behövde en färdplan för fortsatt utveckling och förvaltning tas fram. En sådan återfinns i avsnitt 8 Fortsatt arbete.

5 Koncept- och modellutveckling

5.1 DATAINSAMLING

Modellens indata, dvs de riskfaktorer som påverkar ledningarnas livslängd, har identifierats i möten med bolagen i referensgruppen och i diskussioner inom projektgruppen. Dessa riskfaktorer är bl.a. information om befintliga ledningar, från fortlöpande tillsyn, skadehistorik och omgivande faktorer. När dessa faktorer faststälts, gjordes en inventering av vilka av dessa data som fanns tillgängligt hos bolagen och vilka faktorer som kan hämtas från externa källor. Därefter ombads bolagen att göra exporter från sina NIS-system av tillgängliga data.

Det finns vissa förutsättningar för att data ska vara användbar i maskininlärning. För att kunna koppla ihop data för de olika faktorerna i modellen måste det finnas unika nycklar (ID, koordinater el liknande) som gör detta möjligt. Deltagande bolag måste även ha samma typ av data samt att denna data benämns på samma sätt. För att göra en prediktion av ledningars skaderisk måste utöver ledningsbeskrivningarna även ledningar med och utan historiska skador finnas med. För att lösa problemet med data som har olika benämningar togs en översättningsnyckel fram som skapar gemensamma benämningar.

Inför modelleringen valde vi att hämta data från de bolag som hade flest samstämmiga riskfaktorer i tillräcklig mängd dokumenterade, EON och Göteborg Energi. Utifrån det 40-tal genomgångna riskfaktorerna som referensgruppen identifierat som intressanta återstår endast de ledningsbeskrivande faktorerna i Tabell 1. I tabellen kan också modellens förväntade värden på dessa inputs ses.

Tabell 1. Översikt av ledningsbeskrivande faktorer som modellen använder sig av som indata.

Risikfaktor / Indata	Typ	Tillåtna värden/Kommentar
Vattenmängd, m ³	Kontinuerlig	[0, inf). 0 anses vara ett saknat värde.
Rördimension, mm	Kontinuerlig	[0, inf). 0 anses vara ett saknat värde.
Rörplacering	Kategorisk	marklagd, inomhus, ovanlagd, nan.
Installationsår	Kontinuerlig	[0, inf). 0 anses vara ett saknat värde.
Manteldimension, mm	Kontinuerlig	[0, inf). 0 anses vara ett saknat värde.
Dränering	Kategorisk	Ja, Nej, nan
Rörisolering	Kategorisk	mineralull, polyuretan, cellbetong, nan.
System	Kategorisk	fast, flex, hålrör, nan.
Mantel	Kategorisk	pe-flex, ac, peh, betong, stål, nan.
Medierör	Kategorisk	stål, koppar, nan.

Data som beskrivs ovan har använts i en modellering där faktorer använts som indata till en maskininlärningsmodell vilken tränats för att estimerar vilka rörobject (som beskrivs av faktorerna ovan) som råkat ut för en skada. Under träningen används de historiska skadorna som facit för modellen, och när modellen ska användas på rör som ännu inte råkat ut för skador, så kommer modellen genom statistiska egenskaper att ge ett estimat av sannolikheten för att just detta rör ska råka ut för en skada den närmaste tiden. En maskininlärningsmodell är en struktur som kan lära sig från egenskaper i en datamängd. Under inläringen så ställs parametrar in till värden som gör att modellen gör de prediktioner som är rimliga i förhållande till träningsdata. Av största vikt är att denna modell kan generalisera, det vill säga att den fungerar även för data som inte fanns med i

träningmängden. För att testa detta så har en separat valideringsmängd, för att följa hur väl modellen fungerar på data som inte fanns med under träningen, använts under utvecklingen.

5.2 GENOMFÖRANDE

Vi har använt en maskininlärningsteknik som kallas Random Forests.

Random Forests fungerar genom att bygga upp en stor mängd beslutsträd på slumpmässiga delmängder av träningsdata. Ett beslutsträd är en modell som delar in data i mindre och mindre delar genom att ställa frågor om variabler i data, för att till slut nå en slutsats eller en prediktion.

Genom att bygga flera beslutsträd och kombinera resultaten av dessa kan Random Forests göra mer robusta förutsägelser. Algoritmen använder slumpmässighet för att välja delmängder av data och variabler att använda i varje träd, vilket minskar risken för överträning (att modellen anpassas för mycket till träningsdata och därför generaliserar dåligt till data den inte sett under träning) och gör den mer flexibel för att hantera olika typer av data.

Efter att alla träd har byggts beräknas medelvärdet av deras utfall för att göra en slutgiltig förutsägelse. Detta ger en predicerad sannolikhet eller 'risk' för att ett ledningsobjekt är skadat. Ett värde nära ett indikerar att sannolikheten för skada är stor. Omvänt gäller att ett värde nära noll indikerar låg sannolikhet för skada.

I detta projekt är data tabulär. Det innebär att den består av rader och kolumner där varje kolumn representerar en egenskap eller faktor, i detta fall en riskfaktor (exempelvis ledningens yttermantel eller ålder) och varje rad representerar en observation (i detta fall ett segment i fjärrvärmenätet).

Random Forests är väl lämpade för att arbeta med tabulär data, på grund av följande anledningar:

1. De kan hantera olika typer av variabler: Random Forests kan hantera både kontinuerliga och kategoriska variabler vilket gör den lämplig för att arbeta med tabulär data som vanligtvis innehåller en mix av dessa.
2. De kan hantera hög dimensionalitet: Tabulär data kan ha en stor mängd faktorer vilket kan göra det svårt att hitta vilka som är relevanta eller att undvika överträning. Random Forests använder slumpmässighet för att välja en delmängd av funktioner för varje träd vilket minskar risken för överträning.
3. De är robusta mot outliers: Tabulär data kan innehålla outliers eller odefinierade värden. Random Forests är robusta mot dessa då det byggs flera träd på slumpmässiga delmängder av data vilket minskar risken för att en enskild observation eller mätning ska påverka resultatet för mycket.
4. De ger enkla förklaringar: Slutresultatet för en Random Forests algoritm kan tolkas som ett medelvärde eller majoritetsröstning från flera beslutsgränser vilket gör det lätt att förstå hur en prediktion har gjorts. Dessutom kan insikt i vilka variabler som har störst betydelse för prediktionen fås med hjälp av algoritmen.

Sammanfattningsvis, Random Forests är en robust och flexibel algoritm som är väl lämpad för att hantera tabulär data genom att kunna hantera olika typer av variabler, hantera hög dimensionalitet, vara robust mot outliers och ge enkla förklaringar på resultatet.

Det finns andra maskininlärningstekniker, exempelvis support vector machines (SVMs) och djupa neuronät, som också går att tillämpa för tabulär data. Djupa neuronät har visat fungera väldigt väl på högdimensionella data med komplexa mönster, men kommer till sin rätt främst när det en mycket stor mängd träningsdata finns tillgänglig, eller när data är rå (exempelvis data som kommer direkt från en sensor), och har behov av representationsinlärning. I detta projekt är dock de existerande

faktorererna redan goda representationer för de objekt som ska analyseras och därmed är Random Forests i stället ett rimligt val.

Som nämns ovan kan Random Forests hantera både kontinuerliga och kategoriska variabler. Dock måste kategoriska variabler först transformeras till ett format som maskininlärningsmodeller kan processa. Ett exempel på en kategorisk variabel i detta projekt är till exempel 'rörisolering'. I detta projekt använder vi *one-hot encoding* för att transformera kategoriska variabler. Det är en teknik som används inom maskininläring och dataanalys för att representera kategoriska variabler som binära vektorer. Längden på den binära vektorn är lika med det totala antalet unika kategorier i variabeln. Varje position i vektorn motsvarar en specifik kategori, och om en datapunkt tillhör den kategorin sätts motsvarande position till 1, medan alla andra positioner sätts till 0. De kontinuerliga variablerna normaliseras eller processas ej på något annat vis.

Nedan kommer vi prata om data i termer av klasser. Vårt träningsdata består av två klasser, skadade och oskadade ledningar. Även om modellen tränas på kategoriskt data kan den ge ett statistiskt estimat på sannolikheten för en skada i ett kontinuerligt intervall. I data för detta projekt, oberoende av vilket bolag som lämnat den, finns det en grov klassojämlikhet. Det finns betydligt fler observationer i data av oskadade ledningsobjekt än skadade ledningsobjekt. Vid träning av maskininlärningsmodeller kan detta bidra till minskad modellprestanda och behöver vanligtvis åtgärdas. Att ha lika många exempel av varje klass i träningen av en klassifikationsmodell kan bidra till att undvika skevhet eller partiskhet i modellens prediktioner. Om en viss klass har mycket färre exempel än de andra klasserna i träningsdata kan modellen ha svårt att lära sig att skilja den klassen från de andra klasserna och kan göra felaktiga prediktioner. Genom att ha lika många exempel av varje klass i träningsdata kan modellen lära sig att identifiera varje klass på ett balanserat sätt och undvika partiskhet. Detta kan bidra till att förbättra modellens prestanda på nya data och göra den mer tillförlitlig. En metod för att få lika många exempel av varje klass i träningsdata är *downsampling*. När *downsampling* används för att hantera obalanserade datamängder med olika antal exempel i olika klasser, tas exempel från den överrepresenterade klassen bort så att antalet exempel i varje klass blir jämnt fördelat eller mer balanserat. Till exempel, om en klass har 1000 exempel och en annan klass har bara 100 exempel, kan 900 exempel slumpmässigt tas bort från den överrepresenterade klassen för att skapa en mer balanserad datamängd. För detta projekt innebär detta att slumpmässigt välja ut samma antal oskadade ledningsobjekt från varje enskilt dataset som det finns skadade ledningsobjekt.

Det är viktigt att notera att *downsampling* kan leda till förlust av information, eftersom vissa exempel tas bort från datamängden. Dessutom resulterar två applikationer av *downsampling* på dataseten inte i samma träningsmängd för modellen, detta för att de oskadade ledningsobjekten sannolikt inte kommer vara desamma.

Den här påverkan kan undersökas genom att göra *downsampling* flera gånger och titta på medelvärdet av resultaten. Det är vanligt att dela upp data i träningsmängd och valideringsmängd när maskininlärningsmodeller tränas och evalueras. Genom att dela upp data i en träningsmängd och en valideringsmängd undviks överanpassning av modellen på träningsdata. En modell som har överanpassats på träningsdata har memorerat data och kan därför inte generalisera till nya data. Genom att använda valideringsdata mäts modellens prestanda på oberoende data och därmed undviks överanpassning. Detta gör det möjligt att få en mer realistisk uppfattning om modellens förmåga att generalisera till nya data.

För att kort sammanfatta förfarandet så *downsamplas* först datasetet så att det innehåller lika många exempel av de båda klasserna, oskadade och skadade ledningsobjekt. Denna mängd delas sedan slumpmässigt in i träningsdata och valideringsdata. Modellen tränas sedan på träningsdata och evalueras på valideringsdata. Genom att göra detta upprepade gånger och titta på medelvärde får vi en uppfattning om hur väl modellen kan förväntas prestera på oseddosedda data och hur stor påverkan *downsampling* och andra slumpmässiga moment har på denna prestanda. I denna rapport

presenteras statistik utifrån 20 sådana slumpmässiga nedsamlingar/delningar. Detta ger ett signifikant medelvärde över modellens prestanda.

Vid evaluering av maskininlärningsmodeller användas metriker vars syfte är att representera modellens prestanda med hänsyn till aspekter som anses vara av vikt för problemet. Vanliga metriker är *accuracy* och *F1-score* för klassificeringsuppgifter. *Accuracy* är en vanlig metrik som mäter hur många av de observationer som finns i tränings- eller valideringsdata som modellen har rätt på. Det vill säga hur många av de oskadade eller skadade ledningsobjekten som modellen har klassificerat på det viset. Detta ger ofta en bra fingervisning om hur väl modellen fungerar men säger inget om hur väl modellen klassificerar var klass för sig. *F1-score* till skillnad från *accuracy* ger en balanserad bedömning av både *precision* och *recall* (återkallande) för modellen. *Precision* är andelen korrekt förutsagda skador jämfört med det totala antalet förutsagda skador av modellen, medan *recall* är andelen korrekt förutsagda skador jämfört med det totala antalet faktiska skador. *F1-score* är ett harmoniskt medelvärde mellan *precision* och *recall*, vilket ger en balanserad bedömning av modellens prestanda för båda måtten.

I resultatsektionerna i denna rapport visar vi både modellens *accuracy* och *F1-score* på valideringsmängden. Detta görs i form av figurer där figurerna beskriver fördelningen över dessa metriker, som fåtts genom att upprepa experimentet som beskrivs som ovan. Detta ger ett medelvärde på dessa metriker som är uttryckt i en orange linje inuti boxarna. Utöver dessa visar vi även en fördelning över riskuppskattningar som modellen har gjort. Denna risk ligger i intervallet $[0,1]$ där en riskuppskattning på 1 motsvarar väldigt hög sannolikhet att ledningsobjektet är att betrakta som skadat. Denna fördelning visualiseras i en figur där oskadade ledningar färgkodas med blått och skadade ledningar med rött. Överlappning visas som lila, dvs de fall där det finns oskadade och skadade ledningar med samma predicerade risk. Det bör alltså vara extra intressant att titta närmare på de blå eller "lila" oskadade ledningar som fått ett högt riskvärde.

Random Forest modeller har ett par inställningar som påverkar de resultat som fås. Dessa är exempelvis hur många beslutsträd som ska ingå i modellen. Dessa har i detta projekt bestämts via en metod för hyperparametersökning som kallas grid search. Detta innebär att olika värden prövas för varje hyperparameter (till exempel antal beslutsträd) för att se vad som funkar bäst. Vad resultatet av denna sökning blev för respektive modell går att utläsa i resultatsektionerna. De modeller som är framtagna i detta projekt kräver försumbar beräkningskraft både för träning och prediktion. Modeller med olika hyperparametrar är fortfarande jämförbara eftersom de bedöms utifrån deras prestanda på ett eller flera utvärderingskriterier, i detta fall *F1-score* på valideringsmängd. Genom att jämföra modellernas prestanda på dessa metriker kan bedömning av vilken modell som fungerar bäst för det specifika problemet göras, i detta fall bolags specifika datamängder. Med andra ord betyder detta att välja till exempel det antal träd för den modellen som tränas på bolag A:s data som leder till bäst *F1-score* på valideringsmängden för bolag A. En modell med dåligt valda hyperparametrar kan få sämre prestanda än en modell med väl valda hyperparametrar. Därmed är det inte garanterat att en modell tränad på bolags A data ger bäst resultat om den innehöll det antal träd som gav bäst resultat för en modell tränad på bolags B data.

Korrelationen mellan riskuppskattning och variationen i enskilda faktorer har också visualiserats. Vi har valt att inkludera en figur per modell som visar detta för faktorn vattenmängd. Den är intressant att reflektera kring, då det är en faktor modellen lägger mycket vikt på vilket visas i resultatdelen. Resterande figurer fås via länk som anges i Bilaga 10.1.

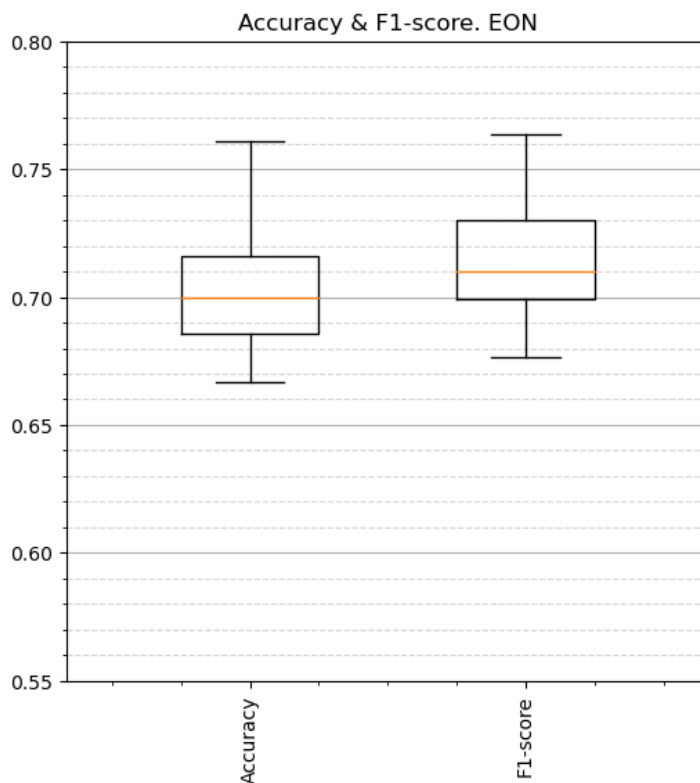
5.3 RESULTAT FRÅN MODELLUTVECKLING

I denna sektion presenterar vi resultat där modellen är tränad och evaluerad på data från enbart ett bolag i taget. Detta ger en baslinje, det vill säga vi kan då utvärdera vilken förändring i prestanda modellen får vid federerad träning. Alltså när modellen ser data från mer än ett bolag under träning.

Resultat från federerad träning ses sedan i Avsnitt 6.2 nedan. En översikt av resultaten för alla modeller återfinns i Bilaga 10.2. Källkod och tränade modeller fås via länk som anges i Bilaga 10.3.

5.3.1 EON

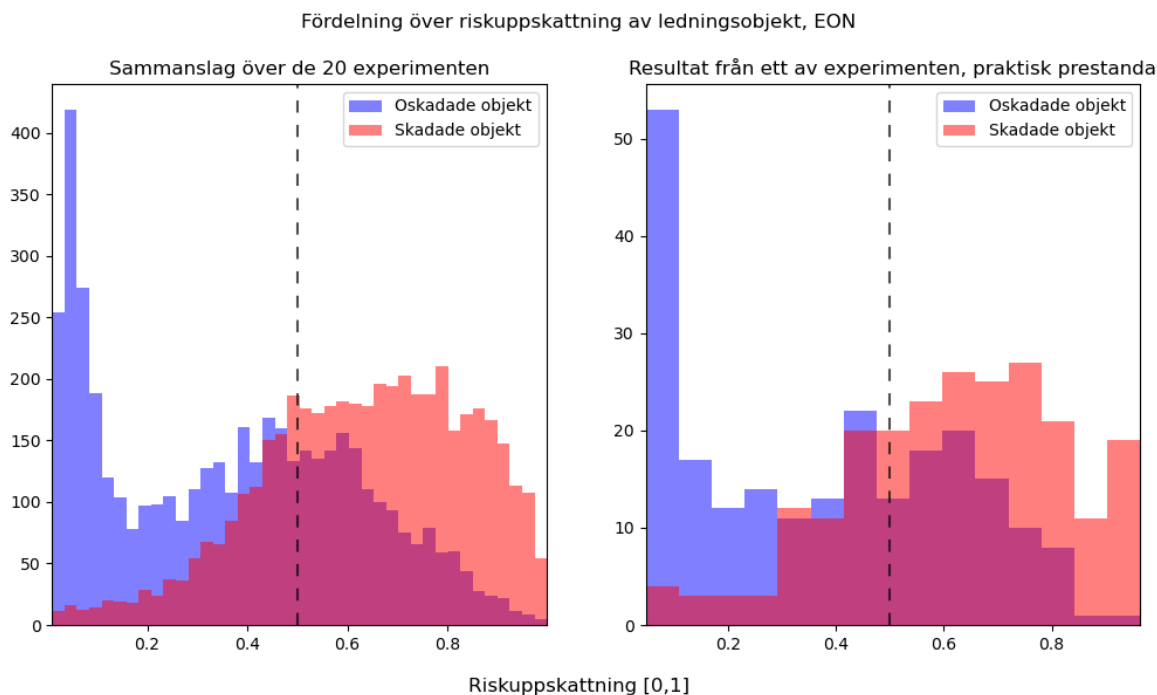
RandomForest modellen som tränats och evaluerats på EONs data består av totalt 150 beslutsträd där noder kan delas vid minst 10 datapunkter vid nod samt där varje löv måste innehålla minst fyra datapunkter. I Figur 2 ses modellens accuracy och F1-score på valideringsmängderna från de 20 experiment som gjorts. Det vill säga i vilken utsträckning modellen predicerar rätt när det kommer till om en ledning är skadad eller inte.



Figur 2: Accuracy och F1-score för modell tränad och validerad på data från EON. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.

Modellen tränad och evaluerad på data från EON visar en relativt hög prestanda och att modellen har kapacitet till att lösa problemet, om än inte fullständigt.

I Figur 3 ser vi fördelningen över vilken risk som modellen tillskrivit de ledningar som ingått i valideringsmängderna.



Figur 3: Till vänster visas en sammanslagning över de 20 experimenten. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Överlapp visas i lila. Data från EON.

Till vänster i Figur 3 ser vi en sammanslagning av de 20 experimenten och till höger riskfördelningen för enbart ett experiment, alltså enbart för en modell på sin egna valideringsmängd. Fördelningen över risk tyder på att modellen har lärt sig problemet väl. Fördelningen över risk ser lovande ut på valideringsmängden när det kommer till att kunna använda sig av modellen för att uppskatta vilken av de ledningsobjekt som bör bevakas men som ännu inte har några rapporterade skador. Till exempel de oskadade objekten med en risk större än ca 0.8–0.9. Se bilaga 10.4 för ett utdrag av oskadade ledningsobjekt med hög predicerad risk.

I Tabell 2 ser vi vilken vikt modellen har lagt på de olika faktorerna för att separera träningsmängden så väl som möjligt. Detta innebär vilken faktor som ges störst vikt när det kommer till att dela träningsmängden i oskadade och skadade objekt med så hög precision som möjligt.

Tabell 2: Normaliserad vikt modellen lagt på de olika faktorerna för att separera träningsmängden i oskadade och skadade objekt. Högre tal betyder att mer tyngd har lagts för separering.

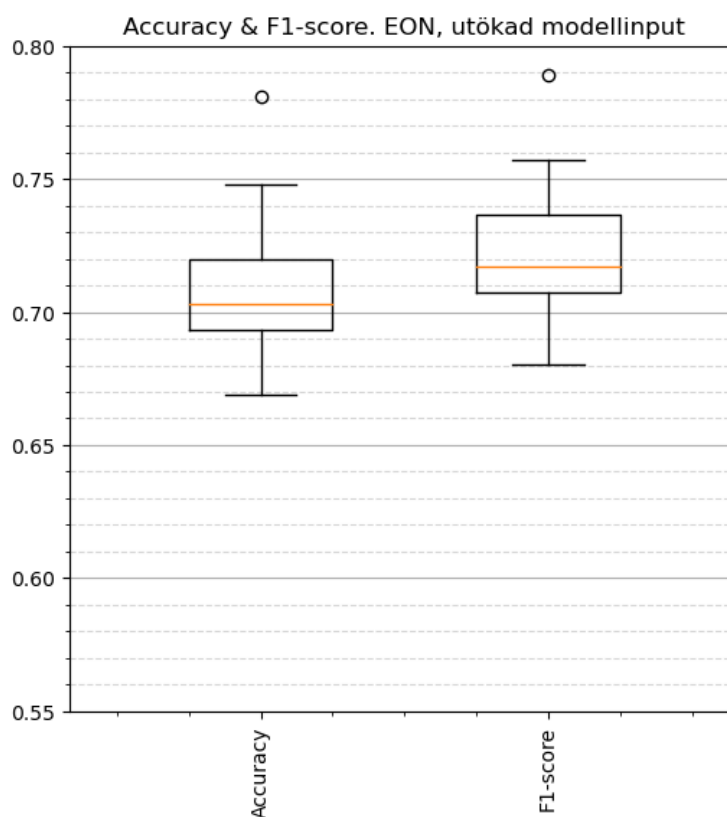
Risikfaktor	Vikt	Risikfaktor	Vikt
Vattenmängd	0.283	Dränering	0
Rördimension	0.109	Rörisolering	0.016
Rörplacering	0.076	System	0.085
Installationsår	0.152	Mantel	0.121
Manteldimension	0.072	Medierör	0.085

Det här resultatet kan tolkas som att ju större den vikten är desto mer beskrivande är just denna input för att göra en korrekt prediktion. Notera dock vid tolkning av detta att det inte bara är en input som har föranlett modellen till att göra en prediktion, snarare ett samspel mellan modellens alla inputs. För EONs data har vi haft möjlighet att utvärdera hur modellprestanda påverkas om ytterligare riskfaktorer inkluderas under träning och evaluering av modellen då denna datamängd innehåller koordinater över ledningsobjekten. Detta gör att data går att kombinera med kartlager från Norrköpings kommun. Främst är det information om marktyp där ledningar ligger, till exempel industriområde, låg/hög bebyggelse etc. Dessutom var kommunen saltar vägar så att avstånd från ledning och/eller betäckningar till närmaste saltade väg kan inkluderas. I Tabell 3 ses medelvärdet av F1-score vid enskild och fullständig inkludering av dessa två ytterligare inputs.

Tabell 3: Modellresultat vid inkludering av ytterligare riskfaktorer i modellen under träning och evaluering. I cellerna visas medelvärde av F1-score över de olika experimenten med 95% konfidensintervall inom parentes.

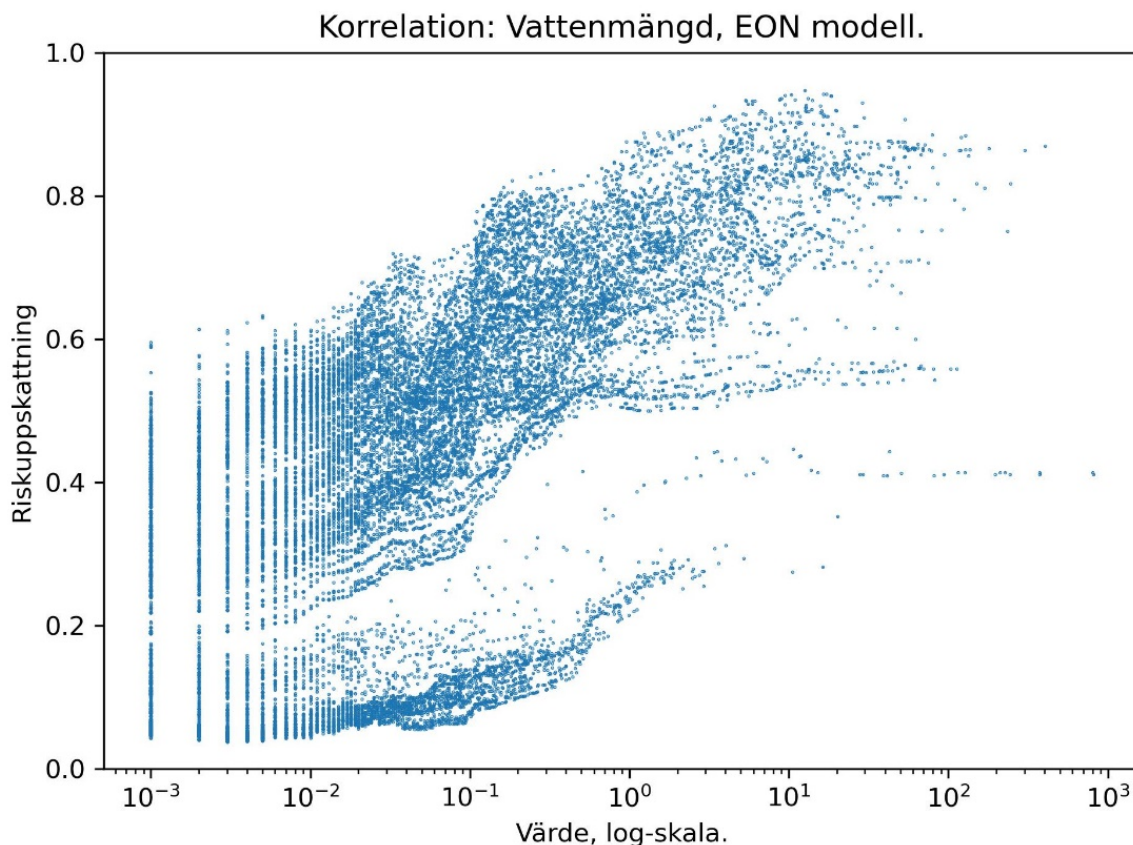
Marktyp	Avstånd saltad väg	Medelvärde F1-score
Exkluderad	Exkluderad	0.714 (± 0.0104)
Inkluderad	Exkluderad	0.722 (± 0.012)
Exkluderad	Inkluderad	0.714 (± 0.0111)
Inkluderad	Inkluderad	0.721 (± 0.012)

Baserat på dessa iakttagelser ger inkludering av marktyp högre medelvärde för F1-score, medan avstånd till saltad väg inte tycks påverka modellens F1-score nämnvärt. Accuracy och F1-score för den fullständiga inkluderingen visualiseras också i Figur 4.



Figur 4: Accuracy och F1-score för modell tränad och validerad på data från EON, både marktyp och avstånd till saltad väg har inkluderats. Resultat från 20 experiment. Visar varians från downsamling och tränings- och valideringsdelning av data.

Faktum är att inkludering av marktyp leder till ett F1-score som motsvarar den globalt tränade modellen vi redovisar i avsnitt 6.2.1. Tabell 2 visar att vattenmängden hade viktats högst av de ursprungliga faktorerna. Faktorn vattenmängd motsvarar mediets volym och är beräknad utifrån ledningssektionens innerdiameter och längd. Riskskattningen ökar med vattenmängden, se Figur 5 nedan.



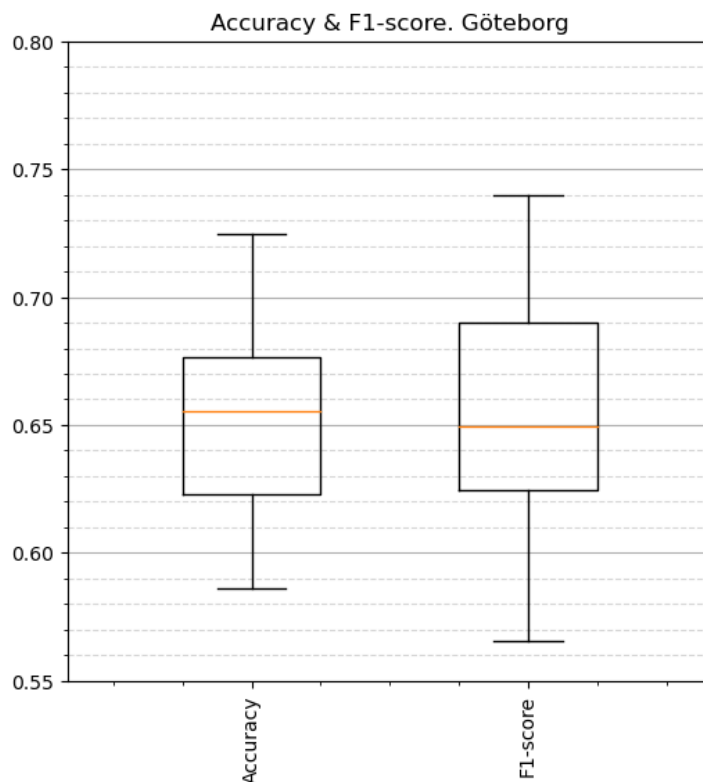
Figur 5. Estimerad risk för skada kontra ledningssegmentets beräknade vattenmängd (volym) för modell som tränats och utvärderats på EONs data.

Ur ett modelleringsmässigt perspektiv har vattenmängden tillskrivits en hög vikt och den blir därför nödvändig för att modellen ska göra bra prediktioner på valideringsdata. Modellen är tränad på att identifiera korrelationer mellan de faktorer som ingått i träningen och skador, inte att identifiera kausala samband mellan alla verkliga riskfaktorer och risken för skada. Därmed finns en risk att övertolka ett sådant här modellresultat. Riskfaktorn vattenmängd kan mycket väl ha viktats högt då sannolikheten för skada är högre för en längre ledning. Vi har sett att inkludering av marktyp förbättrar modellens prediktionsförmåga. Därför rekommenderar vi användare att om möjligt inkludera flera riskfaktorer, tex omgivningsförhållanden. Användare bör nyttja sin domänkunskap både i val av faktorer och vid tolkning av resultat. Som tidigare nämnts utgör ML-modeller inte facit utan snarare ett stöd för erfarna beslutsfattare.

5.3.2 Göteborg

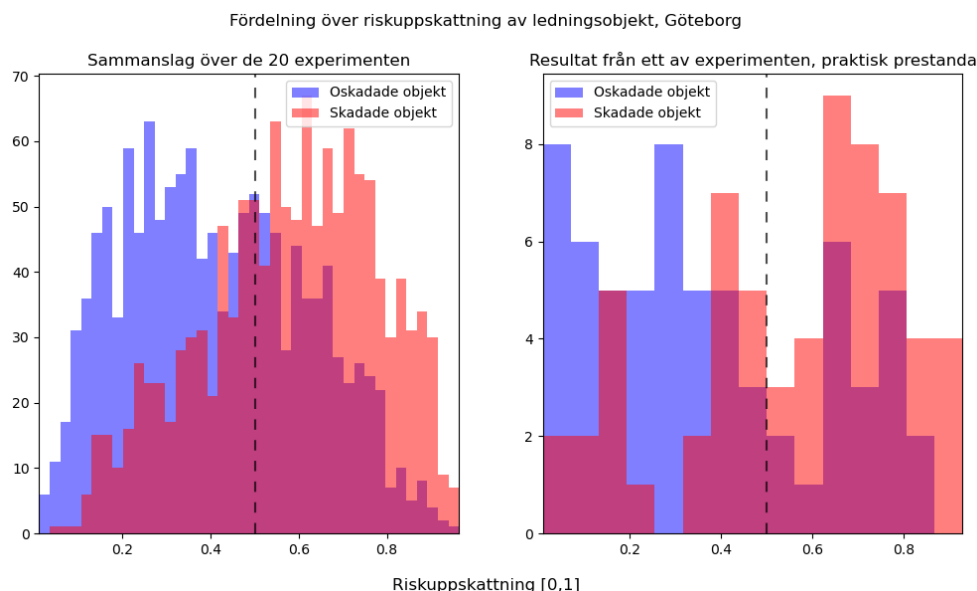
RandomForest modellen tränad och evaluerad på Göteborgs data består av totalt 50 beslutsträd där noder kan delas vid minst åtta datapunkter vid nod samt där varje löv måste innehålla minst en datapunkt.

Denna modell visar något svagare resultat än från EON. Det kan bl.a. bero på det färre antal skador som modellen har haft tillgänglig för inlärning, se Figur 6.



Figur 6. Accuracy och F1-score för modell tränad och validerad på data från Göteborg. Resultat från 20 experiment. Visar varians från downsampling och tränings och valideringsdelning av data.

Liksom för EON har vi också undersökt riskfördelningen i valideringsmängderna, se Figur 7.



Figur 7: Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimentet. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Data från Göteborg.

Till vänster i Figur 7 ser vi en sammanslagning över de 20 experimenten och till höger riskfördelningen för enbart ett utav dem. Alltså enbart för en modell på sin egen valideringsmängd. Som för EON finns det potentiellt möjlighet att använda modellen som stöd för prediktivt underhåll.

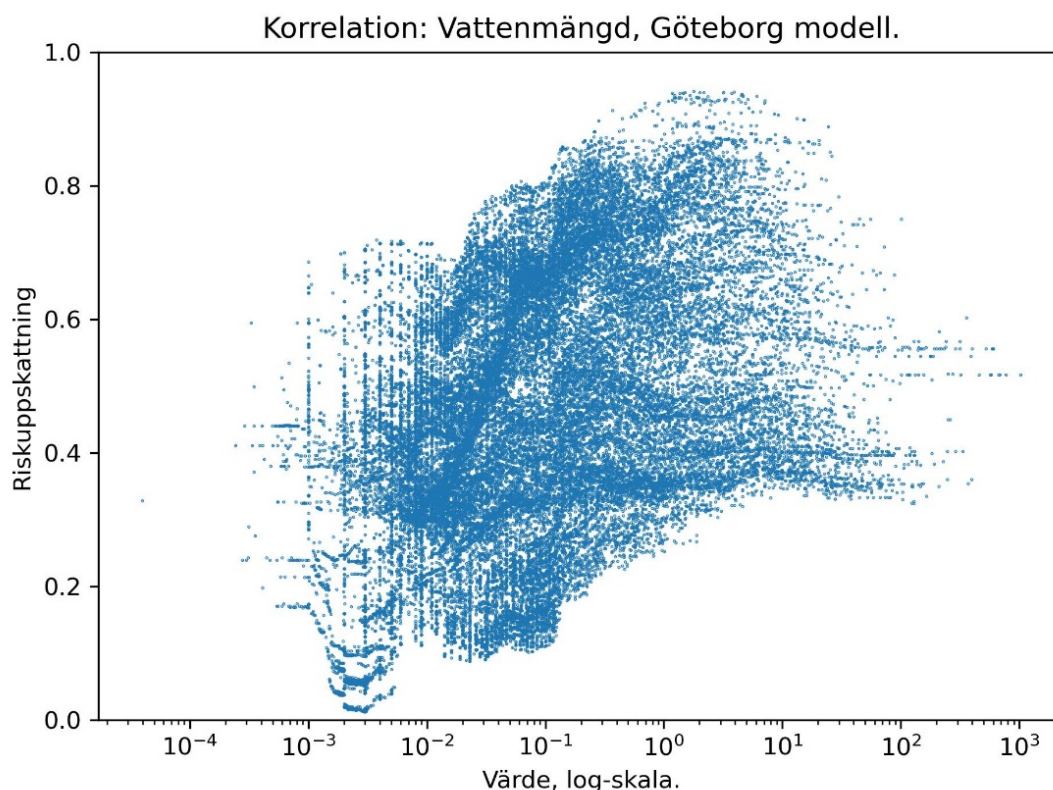
En märkbar skillnad mellan bolagen är att modellen ger ett flertal hittills oskadade ledningar en riskskattning < 0.1 i EONs fall medan vi inte ser samma mönster för oskadade ledningar hos Göteborg Energi. Här är det relativt sett färre ledningar som givits en riskskattning < 0.1. Detta kan troligen härledas till hur bolagens nät är konfigurerade, dvs att det existerar systematiska skillnader mellan bolagen i de faktorer som modellen nyttjar.

Tabell 4 visar vilken vikt modellen har lagt på de olika faktorerna för att separera träningsmängden så väl som möjligt.

Tabell 4: Normaliserad vikt modellen lagt på de olika faktorerna för att separera träningsmängden i oskadade och skadade objekt. Högre tal betyder att mer vikt har lagts för separering.

Risikfaktor	Vikt	Risikfaktor	Vikt
Vattenmängd	0.269	Dränering	0.079
Rördimension	0.146	Rörisolering	0.031
Rörplacering	0.028	System	0.042
Installationsår	0.233	Mantel	0.043
Manteldimension	0.095	Medierör	0.034

I Figur 8 visas hur faktorn vattenmängd korrelerar med riskuppskattning för modellen tränad på Göteborg Energis data.



Figur 8. Estimerad risk för skada kontra ledningssegmentets beräknade vattenmängd (volym) för modell som tränats på Göteborg Energis data.

Liksom för EON ser vi att en större vattenmängd har en tendens till att få högre riskuppskattning av modellen än mindre vattenmängder.

6 Utveckling av prediktiv modell genom federerad inlärning

6.1 GENOMFÖRANDE

Federerad inlärning (Federated Learning) är en metod för maskininlärning där modellen tränas på data som distribueras över flera enheter utan att data centraliseras på en plats. Detta innebär att en modell kan lära sig från mer än en datakälla utan att data behöver delas. Här tillämpas en variant av federerad inlärning genom att ta data från två olika fjärrvärmebolag.

En RandomForest-modell användes för att göra en riskestimering av fjärrvärmenätet baserat på en uppsättning av faktorer som påverkar risken för driftstörningar och skador (se Tabell 1). Modellen tränades separat på data från varje fjärrvärmebolag, och sedan kombinerades modellerna genom att ett antal av beslutsträden från varje bolag valdes ut slumpmässigt. De utvalda träden kombinerades på en central server till en gemensam modell. I övrigt är utförandet identiskt med det som beskrivits i 5.25.2.

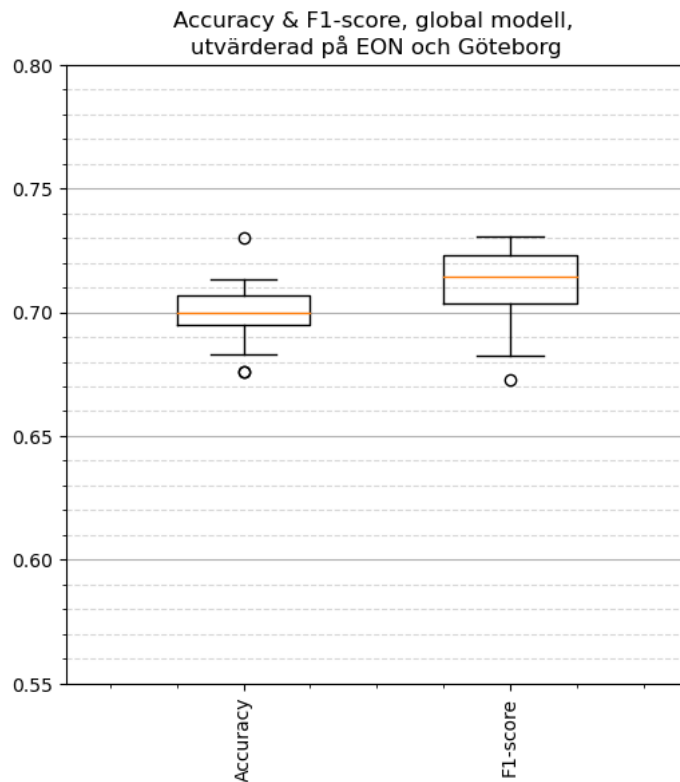
Utöver den federerade modellen som håller data från olika källor separat tränades även en *global* modell där all data först samlas centralt. Detta bildar ytterligare en baslinje som det går att jämföra den federerade modellen emot. I praktiken kan dock tränandet och evalueringen av en sådan global modell vara ogenomförbart på grund av datadelningsdirektiv eller liknande, något som också kringgås genom federerad inlärning.

6.2 RESULTAT

Här visar vi först resultat från den globala modellen som har haft tillgång till all data centralt och sedan resultat för den federerade modellen vilket motsvarar en situation där data aldrig lämnar bolagen. Det som visas motsvarar det som återfinns i 5.35.3.

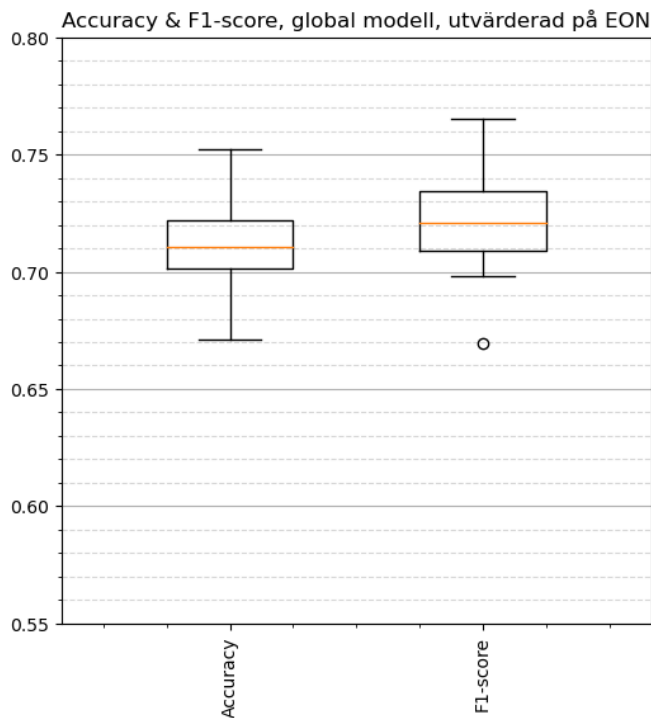
6.2.1 Global modell

RandomForest modellen tränad på data från EON och Göteborg på det globala viset består av totalt 100 beslutsträd där noder kan delas vid minst 10 datapunkter vid nod samt där varje löv måste innehålla minst en datapunkt. I Figur 9, Figur 10 och Figur 11 visas accuracy och F1-score för den globala modellen. Skillnaden mellan dessa är vilken data som modellen har evaluerats på. Initialt evalueras modellen på en sammanslagen valideringsmängd med både data från EON och Göteborg, se Figur 9.

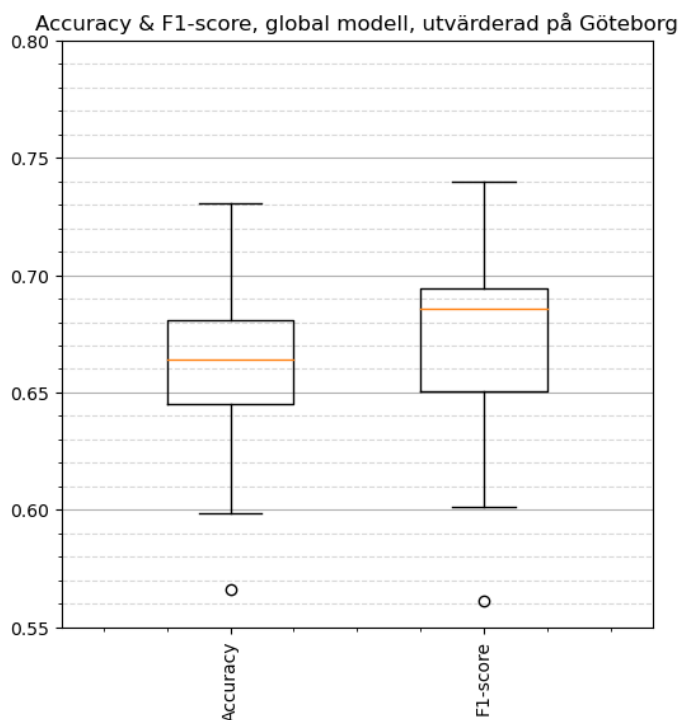


Figur 9: Accuracy och F1-score för global modell tränad och validerad på data från EON och Göteborg. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.

Redan här ser vi att den globala modellens spridning minskar jämfört med tidigare lokala modeller. Den utökade datamängden verkar alltså gynnsamt med avseende på detta. Vi evaluerar också modellen med data från respektive EON och Göteborg, se Figur 10 och Figur 11. Detta möjliggör en jämförelse mellan den globala modellen och modeller enbart tränade på enskilda bolags datamängder.



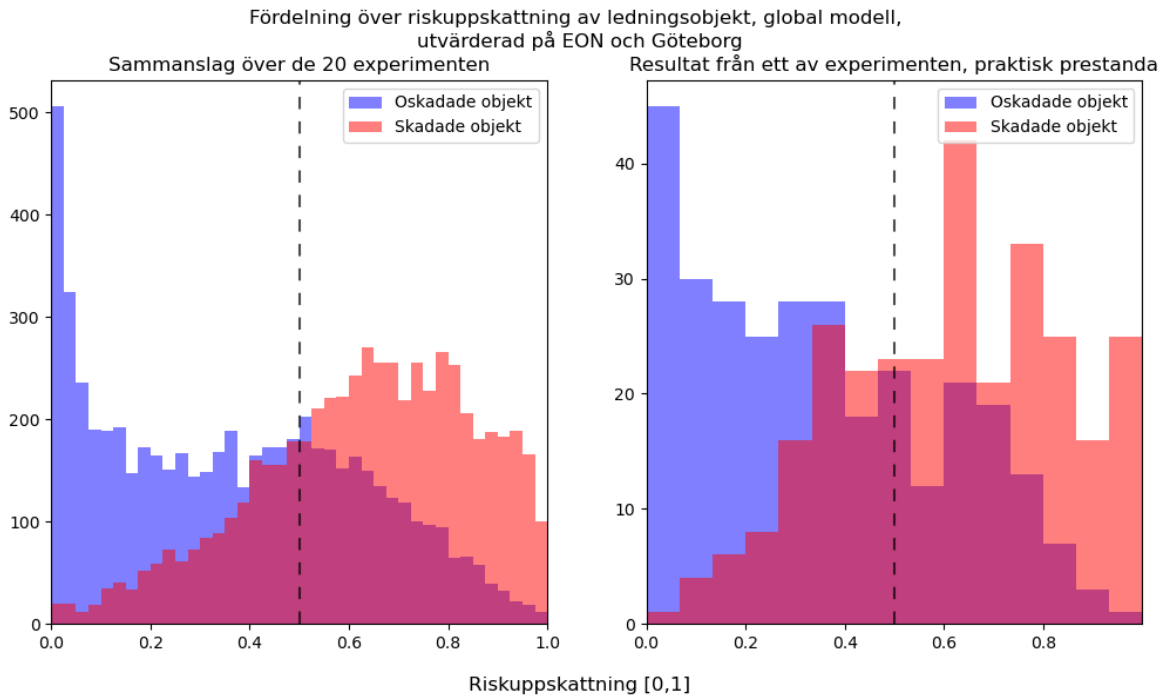
Figur 10. Accuracy och F1-score för global modell tränad på data från EON och Göteborg. Evaluerad enbart på data från EON. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.



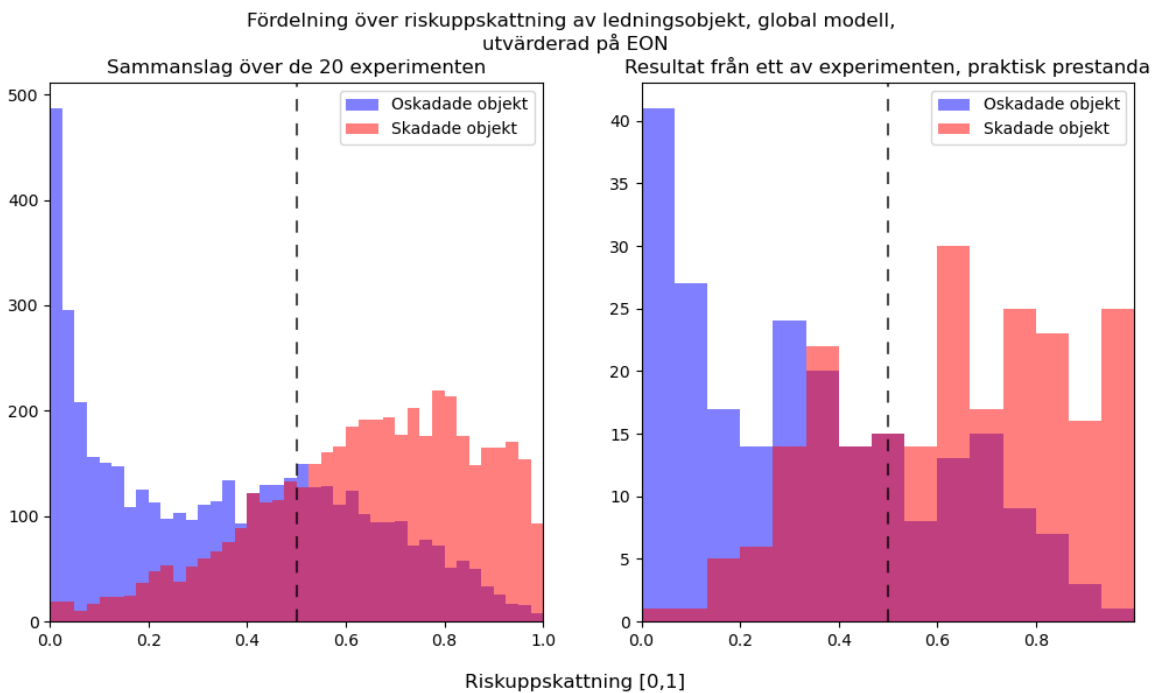
Figur 11: Accuracy och F1-score för global modell tränad på data från EON och Göteborg. Evaluerad enbart på data från Göteborg. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.

För båda bolagen ser vi en förbättring på valideringsmängder för den globala modellen. Störst ökning ser vi när den globala modellen används för evaluering av Göteborgs data jämfört med modellen som enbart är tränad på Göteborgs data. Detta indikerar att data från EON generaliserar åtminstone till viss del till data från Göteborg. Att den globala modellen får högre F1-score på valideringsmängden från Göteborg än den lokala modellen tränad på Göteborgs data betyder att det finns information i EONs data som den globala modellen kan lära sig av när det kommer till att predicera Göteborgs data med större träffsäkerhet.

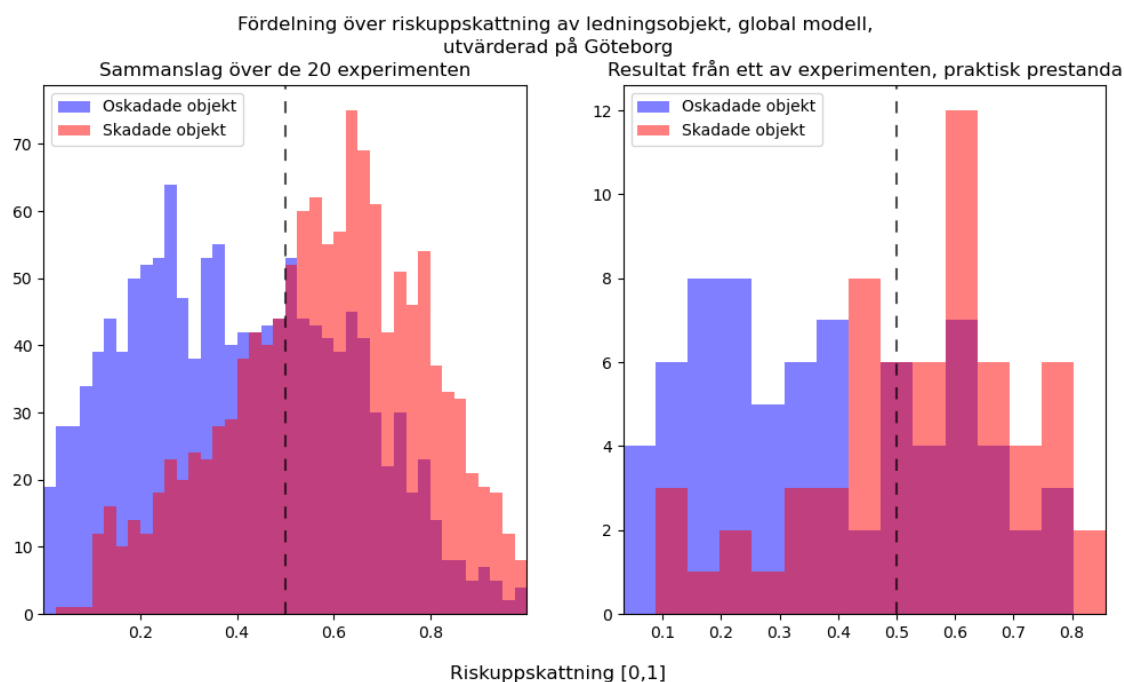
I Figur 12, Figur 13 och Figur 14 ser vi samma typ av fördelningar över risk som modellen tillskrivit de ledningar som ingått i valideringsmängderna som i avsnitt 5.3 som för accuracy och F1-score resultaten visas den globala modellens riskfördelning på en sammanslagning av valideringsmängderna från EON och Göteborg, på EONs valideringsmängd och på Göteborgs valideringsmängd.



Figur 12: Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimenten. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Global modell tränad och utvärderad på data från EON och Göteborg.



Figur 13. Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimenten. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Global modell tränad på data från EON och Göteborg. Valideringsdata från EON.



Figur 14. Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimentet. Till höger visas fördelningen från ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Global modell tränad på data från EON och Göteborg. Valideringsdata från Göteborg.

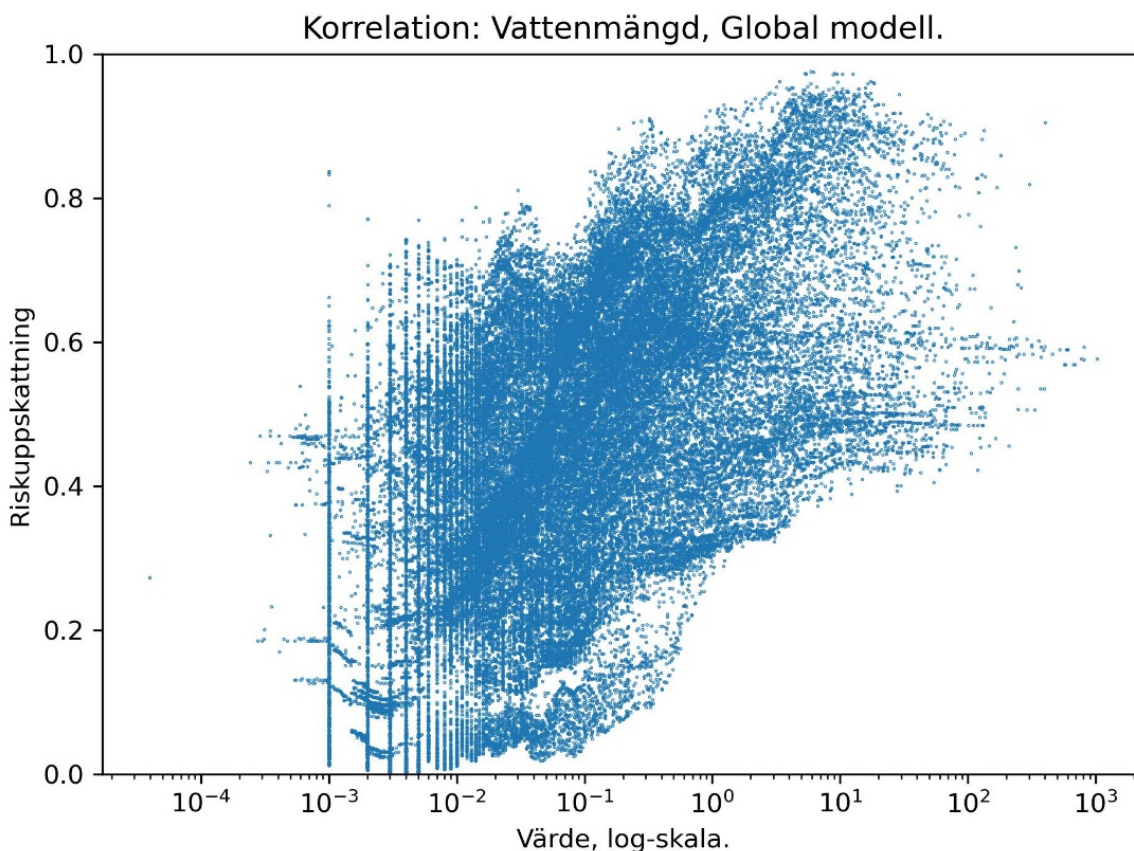
Här återfinns återigen de karakteristiska distributionerna för respektive bolag. I

Tabell 5 ser vi vilken vikt den globala modellen har lagt på de olika faktorerna för att separera träningsmängden så väl som möjligt.

Tabell 5: Normaliserad vikt modellen lagt på de olika faktorerna för att separera träningsmängden i oskadade och skadade objekt. Högre tal betyder att mer vikt har lagts för separering.

Risikfaktor	Vikt	Risikfaktor	Vikt
Vattenmängd	0.297	Dränering	0.03
Rördimension	0.126	Rörisolering	0.02
Rörplacering	0.054	System	0.059
Installationsår	0.2	Mantel	0.068
Manteldimension	0.079	Medierör	0.071

Liksom tidigare modeller tycks modellen vikta riskfaktorerna på ett snarlikt sätt. Återigen ges vattenmängden störst vikt. I Figur 15 visas hur faktorn vattenmängd korrelerar med riskuppskattning för den globala modellen.

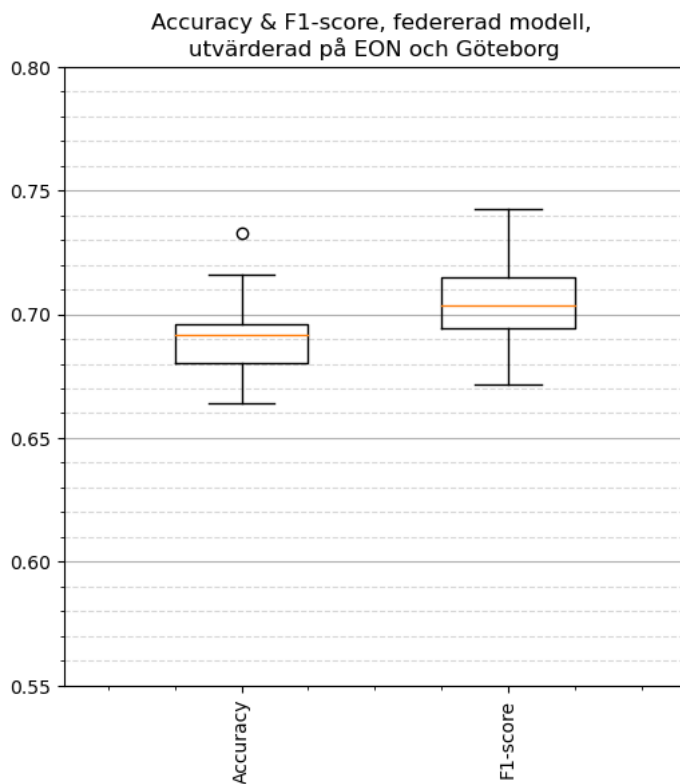


Figur 15. Estimerad risk för skada kontra ledningssegmentets beräknade vattenmängd (volym) för den globala modellen. Global modell tränad och utvärderad på data från EON och Göteborg.

För den globala modellen syns samma trend som för de lokalt tränade modellerna. Alltså att en större vattenmängd leder till en av modellen högre riskuppskattning. Korrelationsdiagram för ytterligare riskfaktorer har tillgängliggjorts elektroniskt. Vi hänvisar till dessa i Bilaga 10.1.

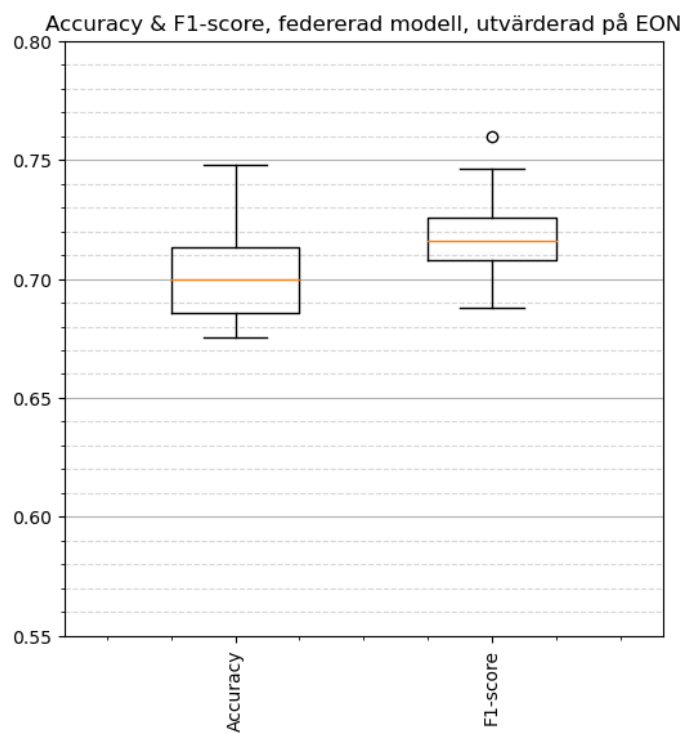
6.2.2 Federerad modell

RandomForest modellen tränad federerat på data från EON och Göteborg består av totalt 250 beslutsträd, där varje bolag bidrar med 125 beslutsträd till den slutgiltiga modellen, där noder kan delas vid minst två datapunkter vid nod samt där varje löv måste innehålla minst en datapunkt. I Figur 16, Figur 17 och Figur 18 visas accuracy och F1-score för den federerade modellen. Skillnaden mellan dessa är vilken data som modellen har evaluerats på. Initialt evalueras modellen på en sammanslagen valideringsmängd med både data från EON och Göteborg, se Figur 16.

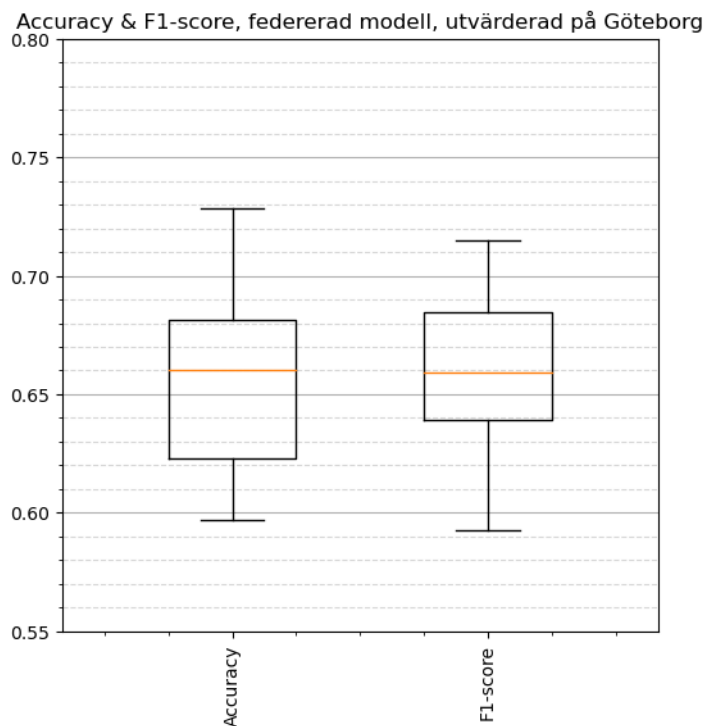


Figur 16: Accuracy och F1-score för federerad modell tränad och validerad på data från EON och Göteborg. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.

I Figur 17 och Figur 18 evalueras modellen på enbart data från respektive EON och Göteborg



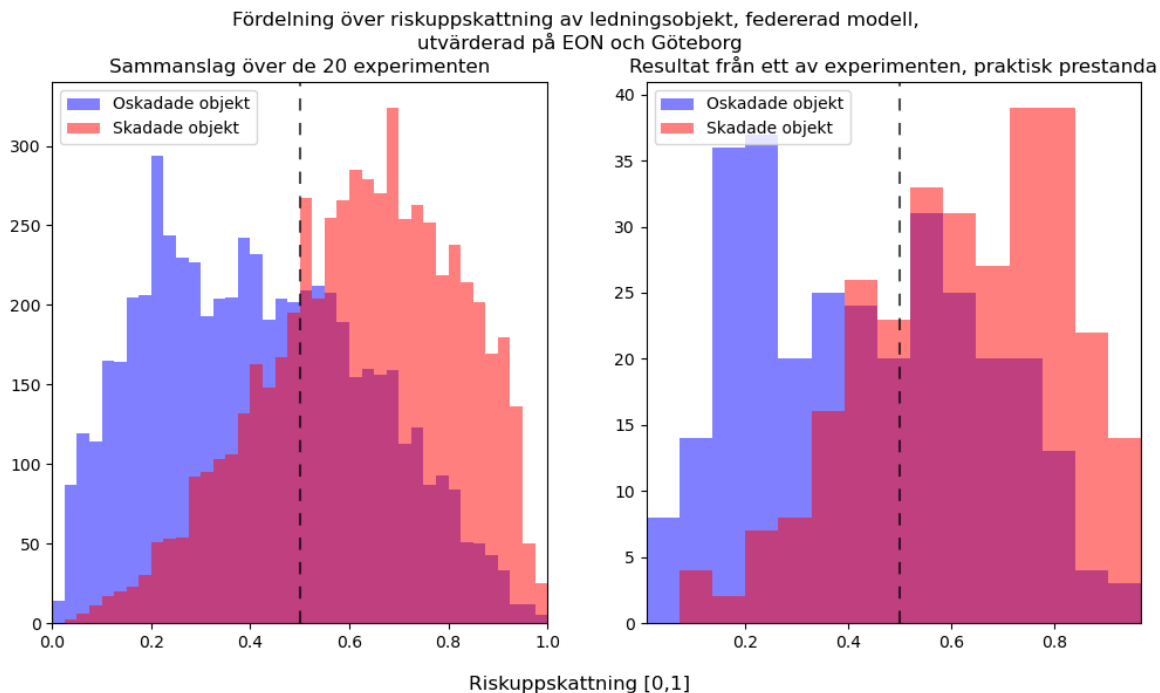
Figur 17: Accuracy och F1-score för federerad modell tränad på data från EON och Göteborg. Evaluerad enbart på data från EON. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.



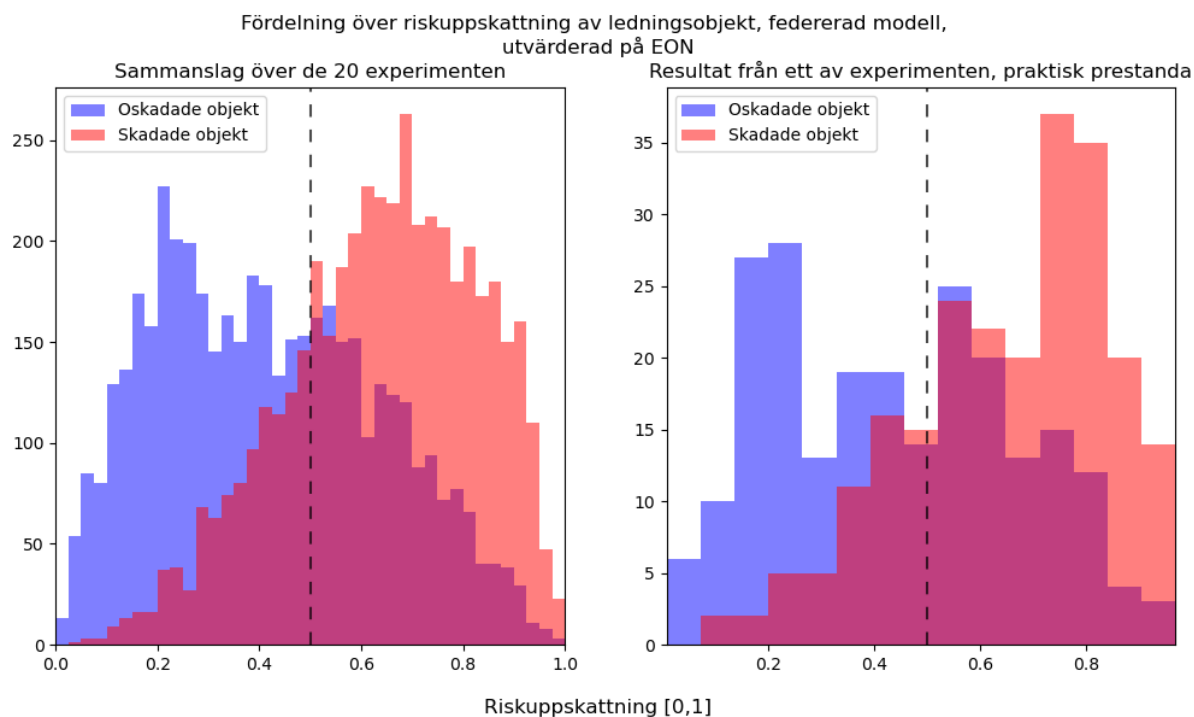
Figur 18: Accuracy och F1-score för federerad modell tränad på data från EON och Göteborg. Evaluerad enbart på data från Göteborg. Resultat från 20 experiment. Visar varians från downsampling och tränings- och valideringsdelning av data.

Resultaten för den federerade modellen är svagare än för den globala modellen. Det är vanligt vid federerad inlärning att det sällan blir lika bra som att ha en global modell med centrerat data. Resultatet för den federerade modellen är relativt oförändrat vid evaluering på EONs valideringsmängd, men för evaluering på Göteborgs valideringsmängd syns ett tapp i F1-score på ett par procentenheter jämfört med den globala modellen. Dock är resultatet för den federerade modellen evaluerad på Göteborgs valideringsmängd fortfarande starkare jämfört med den modell som enbart tränats lokalt på Göteborgs data. Detta visar på en vinst vid federerad inlärning.

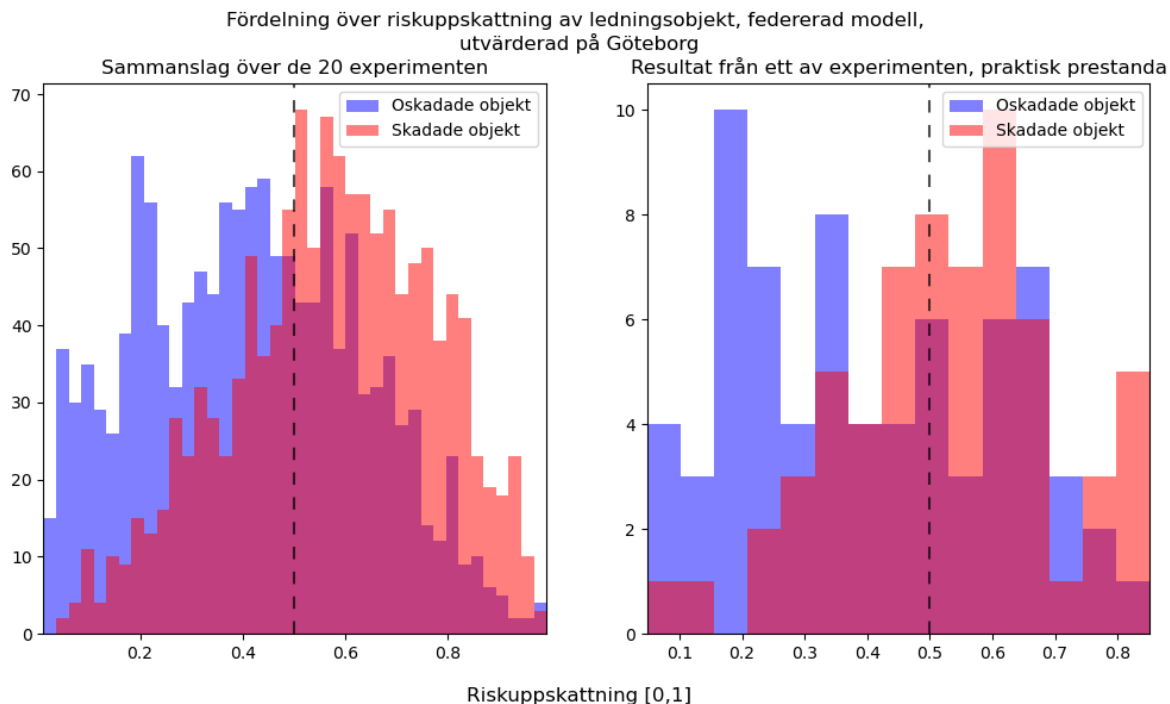
Som för accuracy och F1-score resultaten visas den federerade modellens riskfördelning på en sammanslagning av valideringsmängderna från EON och Göteborg, Figur 19, på EONs valideringsmängd, Figur 20, och på Göteborgs valideringsmängd, Figur 21.



Figur 19: Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimentet. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Federerad modell tränad på data från EON och Göteborg.



Figur 20: Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimentet. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Federerad modell tränad på data från EON och Göteborg. Valideringsdata från EON.



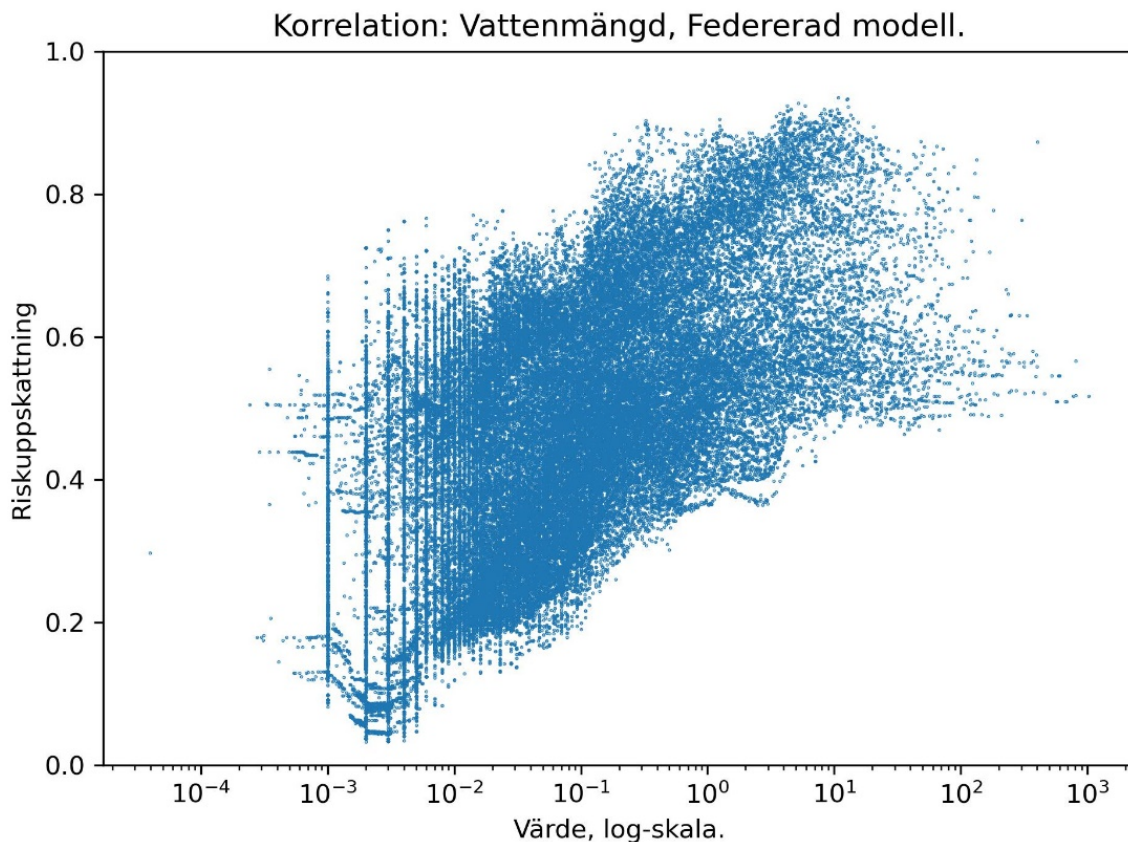
Figur 21: Fördelning över riskuppskattning av ledningsobjekt i valideringsdata. Till vänster visas en sammanslagning över de 20 experimentet. Till höger visas fördelningen över ett experiment. Ledningsobjekt märkta som oskadade visas i blått, ledningsobjekt märkta som skadade visas i rött. Notera att överlapp visas i lila. Federerad modell tränad på data från EON och Göteborg. Valideringsdata från Göteborg.

I Tabell 6 ser vi vilken vikt den globala modellen har lagt på de olika faktorerna för att separera träningsmängden så väl som möjligt.

Tabell 6: Normaliserad vikt modellen lagt på de olika faktorerna för att separera träningsmängden i oskadade och skadade objekt. Högre tal betyder att mer vikt har lagts för separering.

Risikfaktor	Vikt	Risikfaktor	Vikt
Vattenmängd	0.337	Dränering	0.03
Rördimension	0.132	Rörisolering	0.02
Rörplacering	0.034	System	0.035
Installationsår	0.251	Mantel	0.045
Manteldimension	0.083	Medierör	0.031

I Figur 22 visas hur faktorn vattenmängd korrelerar med riskuppskattning för den federerade modellen.



Figur 22. Estimerad risk för skada kontra ledningssegmentets beräknade vattenmängd (volym) för den federerade modellen. Federerad modell tränad och utvärderad på data från EON och Göteborg.

För den federerade modellen syns samma trend som för de lokalt tränade modellerna. Alltså att en större vattenmängd leder till en av modellen högre riskuppskattning. Den globala och federerade modellen har inte identisk vikt på faktorn vattenmängd, den federerade modellen lägger något större vikt på denna faktor. Som ett potentiellt resultat av detta så har den federerade modellen en mindre spridning i riskuppskattning per vattenmängdsvärde. Med andra ord så förklarar värdet på faktorn vattenmängd mer av modellens riskuppskattning för den federerade modellen jämfört med den globala modellen.

7 Diskussion och slutsatser

Allmänt

Vägen från ax till limpa har ofta några kurvor, så även i detta fall. Från det att ansökan skrevs till att projektet tog fart hade världsordningen ändrats. Det var krig i Europa. Detta har påverkat oss alla och därmed också projektets genomförande. Stor vikt fick därför läggas runt säkerhetsfrågor, framförallt dataskydd och formuleringar av lämpliga projektavtal. Det har till viss del medfört förseningar i projektet. Vi fick även ett något lägre bidrag än äskat, vilket gjort det svårare att uppnå ansökans projekt mål till fullo. Vi hade som mål att nå TRL6 men landar enligt vår bedömning på TRL4. Då vi av nämnda anledningar blev försenade gentemot ursprungsplan har fokus framförallt legat på ett intensivt modelleringsarbete och vi har känt oss tvungna att fokusera på de bolagen från referensgruppen med bäst förutsättningar att förse oss med strukturerade data. Kommunikationen med referensgruppen hade under vissa tider kunna vara bättre, men vi är som projektgrupp tacksamma för visad förståelse.

Projektgruppens sammansättning med representanter från Energiforsk, FVB och RISE har utgjort en dynamisk och tät mix av personer med domänkunnskap och expertis inom datavetenskap. Den här typen av konsortium har möjlighet att överbrygga de kompetensgap som trots allt existerar. Detta gäller i båda riktningar. Det tar tid att lära sig varandras domäner oavsett om vi pratar fjärrvärme eller maskininlärning. Detta har vi också märkt i dialog med deltagarna i referensgruppen. Ett exempel är att vi under projektets gång insåg vi att behövde tydliggöra skillnaden mellan expert- och rena maskininlärningsbaserade system. Viktiga aspekter vid val av metod och system för prediktivt underhåll är om behovet eller önskemålet är ett expertsystem eller ett ML-system samt vilka personella resurser som finns tillgängliga. Detta är relevant både vid utveckling och användande. Utöver resursfrågan är säkerhetsaspekten en viktig fråga om data ska delas.

Vi tar med oss några konkreta slutsatser från projektet och väljer att redovisa dem i den ordning de behandlats i rapporten.

Omvärldsanalys

Kartläggningen av initiativ visar att det finns ett flertal kommersiella aktörer eller scale-up (NODA, Utilifeed, Kamstrup, Gradyent, Danfoss) som erbjuder digitala lösningar för fjärrvärmeföretag. Dessa lösningar fokuserar i regel på att optimera och / eller övervaka system och att hitta anomalier eller avvikelser för att identifiera läckage, rundgångar etc. Men vi har inte hittat någon kommersiell aktör som kombinerar statistik, riskfaktorer och dynamiska data för riskklassning av ett ledningssegment. Dock erbjuder till exempel Gradyent och NODA tjänster som dom relaterar till *prediktivt underhåll*. Dessa tjänster bygger på stället på att identifiera avvikelser i driftdata. Något vi återkommit till i samtal med andra personer ur forskarkollegiet är tillgången eller bristen på data för att bygga träffsäkra ML-algoritmer. Det finns en efterfrågan från behovsägare och en önskan från forskarkollegiet att analysera kombinationen av mät- och nätdata. Detta gäller oavsett om intresset är avvikelser i nät eller kundcentraler.

Kartläggningen har också identifierat ett antal relevanta forsknings- och utvecklingsprojekt inom fjärrvärmeområdet (SAB, K2 och DAD), men även inom elnät och vattenledningsnät. Resultaten från dessa projekt har inte direkt använts i utvecklingen av våra algoritmer, däremot har de tjänat som inspiration. Vi har tydliggjort skillnaden mellan expertsystem och system baserade på maskininlärning. Det finns strategier för att kombinera expertbaserade system med ML-baserade system, men detta har lämnats för framtida arbete. Exempel på detta kan var lösningar där ett expertsystem bidrar med träningsdata för ett inlärningsbaserat system, eller ett ML-baserat system som med hjälp av osäkerhetskvantifiering avgör när ett expertsystem behöver tas in. Det går även att låta ett expertsystem påverka inlärningen genom en modifierad inlärningssignal under träningen.

Lösningar som dessa har använts inom fysikmodellering, där datadrivna (ML-baserade) system ofta visar upp en bra generaliseringsförmåga samt ofta en beräkningsekonomisk fördel men där expertsystem kan bidra med kunskap om fysikaliska egenskaper som om de inte beaktas kan leda till sämre prestanda i vissa fall.

I kartläggningen söktes även lämpliga aktörer för samarbete, vidareutveckling eller förvaltning av plattformen.

Svårigheten att samarbeta med konsortier där flera aktörer ingår handlar om vem som ska vara den framtida förvaltaren av projektet. Det är även svårt att se hur aktiva de olika initiativen är idag. I de initiativ som genomförts i projektform eller examensarbete finns ingen samarbetspartner eller framtida förvaltare. Däremot kan det vara möjligt att de metoder som utvecklats kan integreras med, användas i eller kompletteras av andra plattformar.

En förutsättning för ett framtida samarbete med kommersiella aktörer skulle vara att hitta samverkansformer och affärsmodeller som gynnar alla aktörer. Vår spaning blir att flexibla startups med högt engagemang och ett agilt förhållningssätt till en förändrad produktportfölj potentiellt skulle vara lättare att hitta samarbeten med.

Föreningen inom Smart Energi bedöms vara en potentiell samarbetspartner och möjlig förvaltare av resultaten då föreningen styrs av energibolag och fokuserar på branschgemensam digital innovation och datadelning. På samma sätt som Svenskt vatten sprider mjukvaran VASS AI är en branschförening som Energiföretagen Sverige en potentiell framtida distributör av plattformen preDHiCt.

Dessa slutsatser, tillsammans med insikten om att den produkt projektgruppen strävat efter att utveckla inte återfinns i redan existerande produkter, tar vi med oss från denna del av arbetet.

Datainsamling

Modellens indata, dvs de riskfaktorer som påverkar ledningarnas livslängd, har identifierats i möten med bolagen i referensgruppen och i diskussioner inom projektgruppen. Därefter inventerades vilka av dessa data som fanns tillgängligt hos bolagen och vilka faktorer som kan hämtas från externa källor

En förutsättning för att data ska vara användbar i maskininlärning är att de olika faktorerna i modellen måste innehålla unika nycklar (ID, koordinater el liknande) som gör det möjligt att kunna koppla ihop data. En annan förutsättning är att de deltagande bolagen måste ha samma typ av data samt att denna data benämns på samma sätt. Vi valde därför att använda data från de bolag som hade flest samstämmiga riskfaktorer i tillräcklig mängd dokumenterade. För att skapa gemensamma benämningar tog vi fram en översättningsnyckel för denna data.

Vi har trimmat modellen så långt som möjligt utifrån de data vi haft tillgång till, främst ledningsbeskrivande faktorer. De omgivningsfaktorer som vi intuitivt förväntar oss få en större vikt vid riskestimering är förmodligen nödvändiga för att göra en bättre prediktering. Det indikerar vårt försök med inkludering av marktyp i EONs modell. Om denna information finns i bolagens NIS/GIS-system finns de nödvändiga kopplingarna för att data ska kunna användas på ett enkelt sätt.

Koncept och modellutveckling

I detta projekt har vi tränat Random Forest-modeller för att göra riskestimering av fjärrvärmenät. Totalt har vi tränat och evaluerat fyra modeller. Två av dessa har tränats och evaluerats på separata datamängder, en för EON och en för Göteborg. Resterande två har tränats och evaluerats med tillgång till båda bolagens data. En genom att data lokaliserats på samma plats, detta har kallats den globala modellen; En där vi nyttjat tekniker som motsvarar att data aldrig lämnar respektive bolag, även känt som federerad inlärning.

Vi har för varje modell utfört 20 experiment för att uppskatta den varians som uppstår vid träning av modellen. Resultaten redovisas som medelvärde och spridning över dessa 20 experiment. Från detta ses att de lokalt tränade modellerna, de med enbart tillgång till ett bolags data, lär sig att separera oskadade och skadade ledningsobjekt i tränings- och valideringsdata väl. Utöver redovisad accuracy och F1-score för dessa modeller visas också fördelningar över riskuppskattning som modellen gör över objekt i valideringsmängden. Detta indikerar att modellen kan användas i ett förebyggande syfte då få objekt märkta som oskadade i valideringsdata får en predicerad riskuppskattning på över 0.8–0.9.

Dessa lokalt tränade modeller tjänar som måttstock när vi senare presenterar resultat över de modeller som sett mer än ett bolags data under träning. Vi visar att en modell som har haft tillgång till båda bolagens data 'på en och samma plats' presterar avsevärt mycket bättre än de lokalt tränade modellerna. Framför allt vid utvärdering på Göteborgs valideringsmängd. Detta visar att en modell kan dra nytta av data från mer än ett bolag vid träning och potentiella möjligheter till vidare förbättring över enbart lokal träning. En trolig förklaring till detta är att kombinationen av data från olika källor ökar mångfalden av exempel som modellen tränas på. Detta kan leda till att modellen blir mer allmängiltig och därmed bättre på att generalisera till nya data. Den federerade modellen visar också viss förbättring över de lokalt tränade modellerna om än mindre än den globala modellen.

Federerade modeller presterar oftast sämre än modeller som får se data centrerat. Detta till viss del på grund av fundamentala matematiska hinder. Fördelen med federerad inlärning är att sekretess kan upprätthållas hos varje bolag och inga känsliga data behöver delas. I detta projekt har vi valt att konstruera den federerade modellen genom att skapa en ensemble av beslutsträd lokalt tränade från respektive bolag. Det är inte givet att denna typ av federering är den som ger bäst resultat för detta problem och att det finns möjlighet till vidareutveckling.

Genom att använda Random Forests för att göra en riskestimering av ledningsobjekt kan vi också identifiera vilka faktorer som är mest relevanta för att bedöma skaderisken. Dessa har presenterats för respektive modell. Notera att samspelet av faktorer är det som i slutändan leder till en prediktion och inte en summering av enskilda faktorer bidrag. Men denna typ av presenterad vikt kan ändå ge en insikt i vilken faktor det är som väger tyngst när modellen har lärt sig att separera oskadade från skadade ledningsobjekt i träningsdatan.

Vi visar även att inkluderingen av ytterligare riskfaktorer kan förbättra modellens prestanda. Denna undersökning var begränsad till data från EON av praktiska skäl.

Sammanfattningsvis har vi i detta projekt visat att det går att modellera skadestatistik från fjärrvärmebolag med Random Forest modeller. Även i databegränsade situationer, både med antalet faktorer och exempel i åtanke. Vi har visat att en modell tränad på mer än ett bolags data sedan går att använda av enskilda bolag med bättre prediktioner än om dessa modeller enbart hade tränats lokalt.

8 Fortsatt arbete

Som förberedande workshops och ansökan återspeglar har vi inom preDHiCt arbetat för en prototypplattform på TRL-nivå 6. Utvärderar vi projektets måluppfyllelse kan vi av redovisade skäl säga att vi inte riktigt nått hela vägen. Vi befinner oss snarare idag på TRL4. Vi har applicerat utvecklat och tillämpat maskininlärningsmodeller på en begränsad mängd riskfaktorer från ett urval av deltagande bolag. Vi har också lokalt satt upp en skyddad miljö för federerad maskininläring, för att simulera en tänkt framtida infrastruktur för en vidareutvecklad plattform. Allt modelleringsarbete har skett på RISEs servrar. Vi identifierar fyra nyckelaktiviteter för att lyfta projektresultaten till TRL9.

Standardiserat format för datadelning

Det här har vi inte explicit avhandlat i rapporten tidigare. Däremot har vi under arbetets gång identifierat ett behov av ett standardiserat format för datadelning. Detta har vi i projektgruppen själva erfarit. Det har också lyfts av flertalet deltagare i referensgruppen. I nuläget använder bolagen ett flertal NIS och GIS-mjukvaror. Hur objekt benämns i dessa följer inte en gemensam standard vilket får till följd att aggregering av data från flera bolag blir mycket komplicerad. Vi efterlyser därför en branschgemensam standard som åtminstone kan efterlevas vid export eller upprättande av API:er från nämnda system. Exempel är Energiföretagens skadenyckel där ett branschgemensamt format för skador finns. Arbete pågår även för att utveckla en importfunktion för import av data från de olika GIS/NIS-leverantörerna (Energiföretagen Sverige, 2023) samt tidigare EU-initiativ riktat mot smarta städer och smart infrastruktur (FIWARE, 2023) (ETSI, 2023). Det bör också nämnas att liknande röster hörts tidigare, exempelvis från det svenska initiativet Smart Energi, där arbete gjorts för en gemensam fjärrvärmesaxonomi. På Energiföretagens hemsida finns rapport FVF 2000:02 Ledningsdokumentation. Syftet är att den ska utgöra branschrekommendation och hjälpmedel vid dokumentation. Den är i behov av uppdatering, men visar att liknande ambition finns hos Energiföretagen.

Utveckling av modellen

Vi har presenterat en modelleringsapproach där vi identifierat ett 40-tal riskfaktorer och inhämtat de som funnits tillgängligt hos bolagen. Ett urval har sedan gjorts baserat på expertkunskaper (förväntad prediktiv signal) samt för att en så stor andel som möjligt av de utvalda egenskaperna ska finnas tillgänglig bland de olika bolagens data. Modellen kan göra prediktioner baserat på indata som saknar vissa värden, men ju fler saknade värden desto större osäkerhet kommer riskestimeringen att ha. Vi har påvisat att inkludering av externa faktorer som markförhållanden stärker modellens prediktionsförmåga. I dialog med referensgruppens medlemmar har vi en gemensam önskan om att få med riskfaktorer som indikerar fukt i systemet. Denna data har inte gått att vikta mot sannolikheten för skada i detta projekt, men det skulle sannolikt förbättra modellen ytterligare. Erfarenhet har visat att det är en faktor med stor betydelse för medierörens livslängd.

Implementera bolagsövergripande plattform

RISE har utvecklat en modell för riskestimering, men har inte för avsikt att förvalta en framtida lösning. I nuläget befinner vi oss i ett modelleringsstadium i en skyddad "laboratoriemiljö". Nuvarande lösning kan potentiellt lyftas till en bolagsgemensam federerad plattform om datatillgången anses tillräcklig. Det var så den ursprungliga visionen såg ut. En koordinerande part behöver i så fall sätta upp en centraliserad tjänst som administrerar det federerade nätverket. En klientlösning installeras sedan som en del av respektive bolags IT-infrastruktur. Systemet tränas sedan vilket resulterar i en gemensam prediktionsmodell, trots att bolagens data inte lämnar dess ägare.

Ägandeskap/förvaltning

Oavsett implementering krävs en central aktör med en affärsmodell som stödjer den här formen av förvaltande. I nuläget kanske det är enklare att tillgängliggöra en färdigtränad modell som en gemensam resurs via ett API eller extern webbtjänst. Det kan då vara lättare att hitta en lämplig förvaltare. Detta har vi sett och lyft exempel på från VA-sektorn. I omvärldsanalysen identifierade vi branschorganisationen som en potentiell aktör som borde ha tillräckliga incitament för att förvalta en liknande plattform.

9 Referenser

- 3S Consult GmbH. (2023). *Kanew 3S*. Hämtat från <https://www.3sconsult.de/en/kanew-3s/>
- Andersson, H., Bjurling, G., Fransson, A., Jakobsson, H., Johansson, P., Larsson, L., . . . Åkerlund, K. (2015). *Underhållshandboken för fjärrvärmedistribution*. Svensk Fjärrvärme.
- Andersson, S., Molin, J., & Pletikos, C. (1999). *Underlag för riskbedömning och val av strategi för underhåll och förnyelse av fjärrvärmeledningar*, *Fjärrvärmeföreningen* 1999:41, *FoU* 1997:12, ISSN 1402-5191. Fjärrvärmeföreningen.
- Danfoss. (2023). *Leanheat® Network*. Hämtat från <https://www.danfoss.com/en/products/dhs/software-solutions/danfoss-leanheat-software-suite-services/leanheat-network/#tab-overview>
- DIVA Digitala Vetenskapliga Arkivet. (2021). Hämtat från Opportunities for Machine Learning in District Heating: http://www.diva-portal.org/smash/record.jsf?dswid=4084&pid=diva2%3A1576333&c=3&searchType=SIMPLE&language=sv&query=dad+district+heating&af=%5B%5D&aq=%5B%5B%5D%5D&aq2=%5B%5B%5D%5D&aqe=%5B%5D&noOfRows=50&sortOrder=author_sort_asc&sortOrder2=title_sort_asc
- Energiföretagen Sverige. (2023). *Skadenyckeln*. Hämtat från <https://www.energiforetagen.se/fragor-vi-driver/listsida/fjarrvarmefragor/statistikinsamling/>
- ETSI. (2023). *ETSI GS CIM 009 V1.7.1 (2023-06)*. Hämtat från ETSI: https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.07.01_60/gs_CIM009v010701p.pdf
- FIWARE. (2023). Hämtat från FIWARE About us: <https://www.fiware.org/about-us/>
- (2009). *Fjärrvärme - A real success story*. Energiföretagen Sverige. Hämtat från https://www.energiforetagen.se/globalassets/energiforetagen/om-oss/fjarrvarmens-historia/fjarrvarme_story.pdf
- Gradyent. (2023). *Our real-time Digital Twin Platform puts you in control with six solutions that matter*. Hämtat från <https://www.gradyent.ai/solutions/solutions-overview>
- Göteborg Energi. (2020). *20-09: Prediktiv tillståndsbedömning av kablar i elnät baserat på dataanalys och maskininlärning*. Hämtat från <https://www.goteborgenergi.se/om-oss/vad-vi-gor/forskning-utveckling/forskningsprojekt/20-09-prediktiv-tillstandsbedomning-av-kablar-i-elnat-baserat-pa-dataanalys-och-maskininlarning>
- Hogland, W. (2023). *Projekt: Smart tillgångsförvaltning för fjärrvärmesystem i Östersjöregionen (BSAM)*. Hämtat från Linnéuniversitetet: <https://lnu.se/en/research/research-projects/project-baltic-smart-asset-management-bsam/>
- Kamstrup. (2023). *Heat Intelligence*. Hämtat från <https://www.kamstrup.com/se-se/varmelosningar/analys-av-varme/heat-intelligence>
- Karlsson, E. (2022). *AI Vattenledningsnät*. Hämtat från <https://www.svensktvatten.se/vattentjanster/rornat-och-klimat/ai-vattenledningsnat/>
- Kensby, J. (2022). *Digitaliserad nätövervakning och scenariobaserad nätdimensionering med hjälp av AI och ML*. Energimyndigheten. Hämtat från <https://termoinnovation.se/wp-content/uploads/2022/07/51528-1-optimera-investeringar-i-energisystem-med-ai-docx.pdf>
- Kuylentierna, J. (2022). *Plattform för prediktivt underhåll, Predhict*. Hämtat från Energiforsk: <https://energiforsk.se/program/plattform-for-prediktivt-underhall-predhict/>
- Malm, A., Mokhlesi, S., Sernhed, K., & Yarahmadi, N. (2016). *Kartläggning av materialfördelning och skadestatistik av vatten- och fjärrvärmeledningar i Sverige som underlag till rekommendationer för datainsamling*. SP Hållbar Samhällsbyggnad.
- Mbiydzennyuy, G. (2022). *DAD*. Hämtat från <https://www.hb.se/en/research/research-portal/projects/data-analytics-for-fault-detection-in-district-heating-dad/>
- Mbiydzennyuy, G., Nowaczyk, S., Knutsson, H., Vanhoudt, D., Brage, J., & Calikus, E. (2021). *Opportunities for Machine Learning in District Heating*. Applied Sciences. doi:<https://doi.org/10.3390/app11136112>
- Noda. (2023). Hämtat från <https://noda.se/> april 2023
- Ohlsson, M. (2021). *SAB Smart Aktiv Box in Helsingborg*. Helsingborg: Öresundskraft.
- Ohlsson, M. (2022). *Smart Aktiv Box Smart fjärrvärmenät*. Energimyndigheten. Hämtat från <https://www.energimyndigheten.se/forskning-och->

- innovation/projektdatabas/sokresultat/GetDocument/?id=20899ac8-6332-485b-888e-15c111381f2d&documentName=Slutrappport_2.pdf.pdf
- Olsson, O. (2022). *Undersökning av luftfuktighetens förändring över tid och dess påverkande faktorer i fjärrvärmekulvert*. Lund: Lunds Tekniska Högskola. Hämtat från <https://lup.lub.lu.se/luur/download?func=downloadFile&recordOid=9087991&fileOid=9088011>
- Rehn, D. (2019). *En AI-modell för vattenledningsnätet*. Hämtat från Sweden Water Research: https://www.swedenwaterresearch.se/wp-content/uploads/2019/05/Rapport_AI_2019-ny.pdf
- Sederholm, B. (2021). *Fältförsök för statusbedömning av fjärrvärmeledningar*. Energiforsk.
- Sintef. (2023). *CARE-W Computer Aided REhabilitation of Water Networks*. Hämtat från <https://www.sintef.no/projectweb/care-w/>
- SWR. (2023). *Sweden Water Research*. Hämtat från <https://www.swedenwaterresearch.se/om-sweden-water-research/>
- (2018). Säljare Kamstrup. (O. Penttinen, Intervjuare)
- Sörensen, J., & Nilsson, E. (2019). *Ordning i RörANN*. Hämtat från Ordning i RörANN – en smart ANN-modell för minskat läckage på dricksvattennätet: <https://www.smartbuilt.se/media/talnx22k/rapport-sme-r%C3%B6rann.pdf>
- Sörensen, J., & Penttinen, O. (2019). *Utvärdering av icke-förstörande mätmetoder för lokalisering av skador på rörledningar*. Lund: Sweden Water Research.
- Vatten, S. (2020). Hämtat från AI-prediktioner i VASS: <https://docplayer.se/202401609-Ai-prediktioner-i-vass.html>
- Wästberg, F., Hansson, M., & Edland, R. (2022). *Branschsamarbete för avancerad analys av värmedistribution och uppvärmningsbehov*. Energiforsk. Hämtat från <http://smartenergi.org/datasciencebrava/>
- Xue, P., Jiang, Y., Zhou, Z., Chen, X., Fang, X., & Liu, J. (2020). *Machine learning-based leakage fault detection for district heating networks*. Energy and Buildings. doi:<https://doi.org/10.1016/j.enbuild.2020.110161>
- Åkerström, Å. (2004). *Reinvesteringsmodell för befintligt fjärrvärmenät*. Institutionen för Kemisk Teknologi Kemisk Apparatteknik LUNDS TEKNISKA HÖGSKOLA.

10 Bilagor

10.1 KORRELATIONSFIGURER

Korrelationsfigurer för faktorer som ingått i träningen återfinns på följande länk:
["https://git.ri.se/martin.willbo/predhict/-/tree/main/corr"](https://git.ri.se/martin.willbo/predhict/-/tree/main/corr).

10.2 RESULTATÖVERSIKT

Tabell 7: Översikt av resultat. På raderna är de olika modellerna som har tränats och evaluerats i arbetet. I kolumnerna är de olika valideringsmängderna. I cellerna visas medelvärde av F1-score över de olika experimenten med 95% konfidensintervall inom parentes.

Modell / Valideringsmängd	EON valideringsmängd	Göteborg valideringsmängd	Sammanslagen valideringsmängd
EON lokal	0.714 (± 0.0104)	x	x
EON lokal utökad	0.721 (± 0.012)	x	x
Göteborg lokal	x	0.655 (± 0.0218)	x
Global modell	0.721 (± 0.01)	0.67 (± 0.0204)	0.71 (± 0.0078)
Federerad modell	0.718 (± 0.0083)	0.652 (± 0.0235)	0.705 (± 0.0079)

10.3 KÄLLKOD

Källkod och tränade modeller förvaras av RISE och nås via länken:
["https://git.ri.se/martin.willbo/predhict."](https://git.ri.se/martin.willbo/predhict)

10.4 UTDRAG PREDICERAD RISK LEDNINGSOBJEKT

Tabell 8: Exempel på oskadade objekt klassade som högriskledning (riskobjekt) från EONs datamängd. Prediktioner är gjorda med modellen tränad på data från EON. Ett x i cellerna motsvarar ett saknat värde i datamängden för den faktorn.

water_content	pipe_dimension	pipe_location	installation_year	jacket_dimension	drainage	pipe_isolation	system	jacket	mediapipe	score
1.607	125.0	x	1972.0	442.0	x	mineralull	hålrör	ac	stål	0.942
0.017	80.0	x	x	334.0	x	mineralull	hålrör	ac	stål	0.942
0.101	80.0	x	x	334.0	x	mineralull	hålrör	ac	stål	0.941
0.345	100.0	x	1973.0	388.0	x	mineralull	hålrör	ac	stål	0.936
0.013	50.0	x	x	280.0	x	mineralull	hålrör	ac	stål	0.935
0.062	100.0	x	x	225.0	x	polyuretan	fast	peh	stål	0.935

Tabell 9: Exempel på riskobjekt från Göteborgs datamängd. Prediktioner är gjorda med modellen tränad på data från Göteborg.

water_content	pipe_dimension	pipe_location	installation_year	jacket_dimension	drainage	pipe_isolation	system	jacket	mediapipe	score
x	25.0	inomhus	1999.0	x	x	mineralull	x	x	koppar	0.942
1.139	125.0	marklagd	1977.0	225.0	x	polyuretan	fast	peh	stål	0.942
0.034	40.0	marklagd	1958.0	x	x	mineralull	hålrör	ac	stål	0.941
4.150	125.0	marklagd	1982.0	250.0	x	polyuretan	fast	peh	stål	0.936
0.037	65.0	inomhus	1997.0	x	Ja	mineralull	x	x	stål	0.935
0.117	50.0	inomhus	1975.0	x	x	mineralull	x	x	stål	0.935

Tabell 10: Exempel på riskobjekt från EONs datamängd. Prediktioner är gjorda med den globala modellen.

water_content	pipe_dimension	pipe_location	installation_year	jacket_dimension	drainage	pipe_isolation	system	jacket	mediapipe	score
0.039	80.0	x	x	334.0	x	mineralull	hålrör	ac	stål	0.975
0.185	100.0	x	1972.0	388.0	x	mineralull	hålrör	ac	stål	0.973
0.101	80.0	x	x	334.0	x	mineralull	hålrör	ac	stål	0.973
0.017	80.0	x	x	334.0	x	mineralull	hålrör	ac	stål	0.972
0.013	50.0	x	x	280.0	x	mineralull	hålrör	ac	stål	0.971
1.607	125.0	x	1972.0	442.0	x	mineralull	hålrör	ac	stål	0.959

Tabell 11: Exempel på riskobjekt från Göteborgs datamängd. Prediktioner är gjorda med den globala modellen.

water_content	pipe_dimension	pipe_location	installation_year	jacket_dimension	drainage	pipe_isolation	system	jacket	mediapipe	score
5.196	500.0	ovanlagd	1983.0	x	x	mineralull	x	x	stål	0.947
9.405	350.0	marklagd	1966.0	x	x	mineralull	hålrör	betong	stål	0.946
6.758	300.0	marklagd	1967.0	x	x	mineralull	hålrör	betong	stål	0.945
93.821	700.0	ovanlagd	1983.0	x	x	mineralull	x	x	stål	0.945
9.405	350.0	marklagd	1967.0	x	x	mineralull	hålrör	betong	stål	0.945
9.215	300.0	marklagd	1985.0	x	Ja	mineralull	hålrör	betong	stål	0.944

PLATTFORM FÖR PREDIKTIVT UNDERHÅLL

Genom att använda maskininlärningstekniken Random Forests för att göra en riskestimering av ledningsobjekt är det möjligt att identifiera vilka faktorer som är mest relevanta för att bedöma skaderisken. Samspelet av faktorer är det som i slutändan leder till en prediktion och inte en summering av enskilda faktorer bidrag. Inkludering av ytterligare riskfaktorer, till exempel olika omgivningsfaktorer, kan förbättra modellens prestanda ytterligare.

Resultatet av detta projekt visar att det går att modellera skadestatistik från fjärrvärmebolag med Random Forest-modeller. En modell tränad på mer än ett bolags data kan sedan användas av enskilda bolag med bättre prediktioner än om dessa modeller enbart hade tränats lokalt.

Ett nytt steg i energiforskningen

Forskningsföretaget Energiforsk initierar, samordnar och bedriver forskning och analys inom energiområdet samt sprider kunskap för att bidra till ett robust och hållbart energisystem. Energiforsk är ett politiskt neutralt och icke vinstutdelande aktiebolag som ägs av branschorganisationerna Energiföretagen Sverige och Energigas Sverige, det statliga affärsverket Svenska kraftnät, samt gas- och energiföretaget Nordion Energi. Läs mer på energiforsk.se.

