



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2018 30 stp

Fakultet for realfag og teknologi
Sonja Monica Berlijn

Maskinlæring brukt på PMU-data – en litteraturstudie

Machine learning applied to PMU-data – a literature
review

Hanna Lunde Steen

Miljøfysikk og fornybar energi
Fakultet for realfag og teknologi

Forord

Denne masteroppgaven ble skrevet våren 2018 ved Fakultet for realfag og teknologi ved Norges miljø- og biovitenskapelige universitet (NMBU). Den utgjør avslutningen på en femårig mastergrad i Miljøfysikk og fornybar energi. Oppgaven har vært både lærerik, interessant og utfordrende, og jeg håper den vil være til nytte.

Først ønsker jeg å takke min hovedveileder, Sonja Berlijn, for all hjelp og meget god veiledning. Jeg vil også takke min biveileder, Kristin Tøndel, som har gitt mange nyttige innspill. Videre vil jeg takke Arne Auen Grimenes for hans støtte og hjelp i oppstarten av min oppgave.

Sist men ikke minst vil jeg takke samboer, familie og venner som har støttet meg både gjennom å lese oppgaven og ved å rett og slett være der.

Ås, 11. mai 2018

Hanna Lunde Steen

Sammendrag

En rekke utfordringer i kraftsystemet vil bli mer gjeldende fremover; økende kompleksitet, økende forbruk, strengere krav om forsyningssikkerhet, og hensyn til miljø og bærekraft. Dette gjør at nettet trenger å bli smartere og gi bedre kommunikasjon, overvåkning og automatisk reaksjon på hendelser. For å oppnå dette trengs mye data om nettets tilstand, hvor sensoren PMU (phasor measurement unit) er en enhet som kan bidra til dette. Den gir høyfrekvente data som er tidsstemplet og synkroniserte, som legger til rette for å kunne fange opp endringer i kraftsystemet med høy presisjon. En annen teknologi som vil kunne bidra til et smartere nett er maskinlæring. Ved hjelp av store mengder data, kan maskinlæringsmetoder bygge gode modeller som kan gjøre raske og presise prediksjoner. PMU'ens høye samlingsrate gjør at det hentes inn store mengder data som kan brukes som input til maskinlæringsalgoritmer til å sammen utvikle nettet i en smartere retning.

I denne oppgaven har fagområdet "maskinlæring brukt på PMU-data" blitt undersøkt ved hjelp av en litteraturstudie. Det ble først undersøkt hva som er de vanligste maskinlæringsalgoritmene, før disse ble brukt som søkeord sammen med "PMU". Disse søkene ga i alt 110 artikler, som dannet grunnlaget for videre analyse. Fagområdets sammensetning ble undersøkt ved å plote ulike maskinlæringsalgoritmer mot ulike utfordringer i nettet som ble sett på. En rekke histogram over ulike egenskaper ved artiklene ble også presentert, blant annet geografisk opphav og type utgivelsesinstitusjon. Det ble, basert på dette, valgt ut 5 fordypningsartikler som ble gått gjennom i detalj for å gi et dypere innblikk i de mest sentrale underområdene av fagområdet.

Hovedfunnene er at fagområdet er på forskningsstadiet, det vil si TRL-fase 2, den vanligste maskinlæringsmetoden er beslutningstrær og det vanligste problemet å løse er lokalisering/detektering av feil og hendelser i nettet. I fordypningsartiklene ble presentert gode resultater, hvor flere er innenfor det som kan kreves for sikker drift av kraftsystemet. Transientstabilitet ble funnet til å være et sentralt tema, både individuelt og som en del av andre utfordringer.

Basert på funnene foreslås det videre forskning og utvikling på flere områder; deteksjon/lokalisering av feil og hendelser, stabilitet - parallelt med utvikling av probabilitisk drift - og demping. I tillegg anbefales et økt fokus på hvordan en skal klare å heve fagområdet nærmere implementering og til en høyere TRL-fase. Om en skal søke samarbeid med andre fagmiljøer anbefales det å se til USA, Canada og Storbritannia, da disse bidro til flest artikler blant de funnet i litteraturstudien.

Abstract

Several challenges in the power system will be more substantial in the future; increasing complexity, increasing consumption, demand of increased security of supply, sustainability and environmental considerations. This causes a need for the power system to become smarter, through better communication, monitoring and automatic response to events. To achieve this, a lot of data is required that describes the state of the network, and the sensor PMU (Phasor measurement unit) can contribute to this. It provides high-frequency data that is time-stamped and synchronized, which facilitates accurate detection of changes in the power system. Another technology that could contribute to a smarter power system is machine learning. It can, using large amounts of data, build models that can make fast and precise predictions. The high collection rate of the PMU leads to large amounts of data to be collected, that can be used as input to machine learning algorithms.

Based on a literature study, the subject area “machine learning used on PMU data“ has been studied in this thesis. After the most common machine learning algorithms were mapped, they were used as keywords together with “PMU“. These searches gave a total of 110 articles, which formed the basis for further analysis. The subject area’s composition was analysed by plotting different machine learning algorithms against various challenges in the power system that were attempted to be solved. Several histograms of various variables of the articles were presented, for example geographical origin. In addition, 5 central articles were selected to provide a deeper insight into the most central sub-areas of the subject area.

The main findings are that the subject area is at the research stage, i.e. TRL (technology readiness level)-phase 2, the most frequently used machine learning method is decision trees and the most frequent problem to solve is the location/detection of faults and events. In the selected articles good results were presented, where several are within the requirements of safe operation of the power system. Transient stability was found to be a key subject, both individually and as part of other challenges.

Based on the findings from the literature, future research and development is proposed in several areas; detection/location of faults and events, stability - parallel with development of probabilistic operation, and damping. In addition, it is recommended to increase focus on how to develop the field of study closer to implementation and to a higher TRL-phase. If one is to seek cooperation with other academic communities, it is recommended to look to the United States, Canada and the UK, as these contributed to the most articles in the literature study.

Innhold

Forord	i
Sammendrag	iii
Abstract	iv
1 Introduksjon	1
1.1 Innledning	1
1.2 Metodebeskrivelse	2
1.3 Avgrensninger	3
2 Begreper og forkortelser	4
2.1 Begrepsforklaring	4
2.2 Forkortelser	5
3 PMU - Phasor Measurement Unit	6
3.1 Historie	6
3.2 Måleteknikk	7
3.3 Måleenhet	8
3.4 Mikro-PMU	9
3.5 Anvendelser	9
3.5.1 Tilstandsestimering	10
3.5.2 Transientstabilitet	11
3.5.3 Kontroll av kraftsystemet	12
3.5.4 Vern av kraftsystemet	12
4 Maskinlæring	14
4.1 Supervised maskinlæring	15
4.1.1 Regresjon	16
4.1.2 Klassifisering	17
4.2 Unsupervised maskinlæring	21

4.2.1	Clustering	22
4.2.2	Assosiasjon	23
4.2.3	PCA - Prinsipalkomponentanalyse	23
4.3	Reinforcement maskinl�ring	24
4.4	Dyp l�ring	26
4.5	SVM - Support vector machines	27
5	Maskinl�ring brukt p� PMU-data	29
5.1	Litteraturanalyse	29
5.1.1	Vurdering av litteratur	29
5.2	Fordypning	37
5.2.1	Feil/hendelse - unsupervised maskinl�ring	37
5.2.2	Sikkerhetsvurdering - beslutningstr�r og tr�r	38
5.2.3	Transientstabilitet - support vectors	39
5.2.4	Spenning - dyp l�ring	39
5.2.5	Spenning - regresjon	40
5.3	Diskusjon av litteratur	41
5.3.1	Litteraturs�k	41
5.3.2	Fordypningsartikler	42
5.3.3	Videre forskning og utvikling	43
5.4	Oppsummering av funn	46
6	Konklusjon	47
6.1	Forslag til videre forskning	48
6.2	Egenrefleksjon	48
	Bibliografi	50
A	Resultat litteraturs�k	53

Kapittel 1

Introduksjon

1.1 Innledning

I dag er de aller fleste sosiale, økonomiske og politiske aktiviteter avhengige av at kraftsystemet leverer elektrisitet. Kraftsystemet er viktigere i dag enn noen sinne tidligere, og det er viktig å fortsatt opprettholde den gode forsyningssikkerheten Norge har hatt i flere år [21]. Kraftnettets økende kompleksitet, økende forbruk [22], krav om forsyningssikkerhet og hensyn til miljø og bærekraft gjør at vi trenger et stadig smartere nett, med bedre kommunikasjon og informasjonsteknologi. En sentral del av flere smarte nett er at de kan overvåke nettet og automatisk sette inn tiltak for å rette feil eller unngå feil. Da er det nødvendig med høyfrekvent innsamling av data for å kunne plukke opp alle hendelser. Om disse i tillegg er synkroniserte ligger alt til rette for presis oppfanging og analyse av feil. Her seiler PMU frem som en strålende sensor som gir akkurat slike data.

En annen spennende teknologi som får økt fokus for tiden er maskinlæring. Dette gjør det mulig å bygge gode modeller ved hjelp av store mengder data, som raskt kan produsere nøyaktige prediksjoner i nettet, og støtte opp om et smartere nett. PMU gir, med sin høyfrekvente datainnsamling, store datamengder som kan brukes som input til maskinlæringsalgoritmer. Det er denne kombinasjonen, maskinlæring brukt på PMU-data, denne oppgaven har som mål å undersøke. I oppgaven blir det vurdert hvilket nivå fagområdet er på, eksempelvis om det er noe implementert i reelle kraftsystemer eller om det fortsatt er på forskningsstadiet. Videre vil det blant annet analyseres hva som preger fagområdet i form av ulike underområder. Til slutt i analysedelen gjøres det et dypdykk i noen av underområdene for å gi et nærme-

re innblikk i fagområdets muligheter og eventuelle utfordringer. Oppgaven løses gjennom et litteratursøk hvor resultatet brukes som bakgrunn for de overnevnte analysene. For å gi en generell forståelse av aktuell litteratur, gis det en innføring i både PMU og de vanligste algoritmene og metodene innenfor maskinlæring.

1.2 Metodebeskrivelse

Hovedmetoden i denne oppgaven har vært en litteraturstudie. For teoretisk bakgrunn om PMU og maskinlæring ble både bøker og artikler funnet ved hjelp av Google Scholar, hvor boken “Synchronized Phasor Measurements and Their Applications” [25] var hovedgrunnlaget for forståelsen av PMU. Innenfor maskinlæring ble bøkene “The Elements of Statistical Learning” [7] og “The top ten algorithms in Data Mining” [34] mye brukt.

Før litteratursøket på maskinlæring opp mot PMU-data begynte, ble det undersøkt hvilke underkategorier og algoritmer innenfor maskinlæring som er vanligst og mest sentrale. Ut ifra dette ble det bestemt en rekke sentrale søkeord som ble brukt sammen med “PMU” i søkene for å få et representativt utvalg. Det ble søkt i søkemotorene Google Scholar og Web of Science. Størsteparten (72,2 %) av artiklene som ble funnet her var fra kilden IEEE. Da søk i disse to søkemotorene ga tilstrekkelig antall artikler, i alt 110, med en rekke ulike vinklinger, ble det ikke utført ytterligere litteratursøk etter dette.

Etter endt søk ble litteraturens sammensetning undersøkt tanke på hvilke algoritmer som blir brukt, hvilke problemer som blir løst, hvilket nivå det undersøkte fagområdet er på og geografisk utstrekning av opprinnelsesland. Dette ble gjort ved hjelp av ulike histogram, samt plott av hvilket problem som ble løst opp mot hvilken algoritme som ble brukt. På bakgrunn av dette ble 5 artikler valgt til grundigere analyse. Dette ble gjort for å gi et mer detaljert bilde av hva slags problemer som løses innenfor fagområdet, og eventuelt hvilke utfordringer som finnes. For å få mer innblikk i hva som er mest lovende å forske videre på og utvikle mot implementering ble fagpersoner fra NTNU og Statnett SF kontaktet og intervjuet.

1.3 Avgrensninger

Da dette er en masteroppgave, er det begrenset med tid tilgjengelig for innhenting av litteratur, som er den aktuelle datainnhenting i denne oppgaven. For å få en mest mulig representativ oversikt over fagområdet "maskinlæring på PMU-data", ble det satt en rekke avgrensninger for hva som skulle inkluderes i litteraturstudien. Disse er beskrevet i punktene nedenfor:

- Kun transmisjonsnett.
- Kun arbeid etter 2000, for å få et aktuelt overblikk.
- Se på kraftnett-applikasjoner, ikke f.eks. mot generatorer.
- Utelukke plassering av PMU.
- Utelukke mikro-PMU.
- Utelukke PMU opp mot cyber-sikkerhet.
- Utelukke maskinlæring for kun databehandling/strukturering - artikkelen må ha et tydelig mål/problem å løse med dataene. Dette inkluderer:
 - Utelukke maskinlæring for å håndtere missing data/bad data.
 - Utelukke dimensjonsreduksjon og feature selection når dette er eneste metode.

Kapittel 2

Begreper og forkortelser

2.1 Begrepsforklaring

Fagområde - faglig område innenfor for eksempel forskning, i denne oppgaven menes området “maskinl ring p  PMU-data“ n r “fagomr de“ brukes.

Feature selection -   velge ut de forklarende variablene i et datasett som forklarer m lvariabelen best.

Forklarende variabel - Den eller de variablene som er med   beskrive m lvariabelen. For eksempel kan en persons alder og blodtrykk beskrive risiko for hjerteinfarkt, hvor alder og blodtrykk er forklarende variabler og risiko for hjerteinfarkt er m lvariabel.

Hyperplan - Et geometrisk underrom med en dimensjon mindre enn rommet det befinner seg i. Eksempelvis en linje (1 dimensjon) i et plan (2 dimensjoner), eller et plan i et tredimensjonalt rom.

Instans/datapunkt - Et tilfelle av typen data en ser p . For eksempel en enkelt m ling fra en PMU.

K-fold kryssvalidering - En m te   estimere forutsigelsesfeil hvor en deler opp datasettet i K deler, hvor alle bortsett fra en brukes til   trene en modell [7]. Den som er igjen brukes til   teste modellen og estimerer en forutsigelsesfeil. Dette gj res igjen en gang med hver av de K delene utelukket fra trening og brukt til testing en gang hver, til en da sitter igjen med K estimater av forutsigelsesfeil. Deretter tas snittet av disse, som er kryssvalideringsestimatet av forutsigelsesfeil.

Kraftnettet - Når dette eller “nettet” brukes mener transmisjonsnettet.

Metode - En gruppe algoritmer med lik/lignende virkemåte.

Målvariabel - Se forklarende variabel.

Node - I beslutningstrær: punkt hvor data skilles ytterligere i to eller flere nye noder ved å stille et spørsmål. I nevrale nett: punkt som får aktivering fra tidligere node eller input-variabel og sender aktivering videre basert på vekten gitt den bestemte noden.

Underområde - Fagkombinasjon innenfor det utvalgte fagområdet (her “maskinlæring på PMU-data”), for eksempel “sikkerhetsvurdering og beslutningstrær”.

Variabel - Egenskapene som kan variere i hvert datapunkt. For eksempel strøm og spenning i en PMU-måling.

2.2 Forkortelser

DFT - Diskret Fourier transformasjon

GPS - Global positioning system

PCA - Principal component analysis

PMU - Phasor measurement unit

SCADA - Supervisory control and data acquisition

SCDR - Symmetrical component distance relay

SIPS - Systemplan for integritetsbeskyttelse (System integrity protection schemes)

SVM - Support vector machines

TRL - Technology readiness level

TSO - Transmission system operator, systemansvarlig nettselskap

WAMS - Wide-area monitoring system

Kapittel 3

PMU - Phasor Measurement Unit

Prosesen med å gå fra et tradisjonelt nett til et mer moderne og smart nett er godt i gang. Gjennom for eksempel sanntidsovervåkning, bedre kommunikasjon i nettet og WAMS (Wide-area monitoring systems) har nettet blitt stadig smartere [18]. I kjernen av dette finner vi “måleenhet for fasevinkler“, eller Phasor Measurement Unit (PMU) på engelsk, som er en essensiell sanntidssensor for fremtidens smarte nett. Kraftnettet drives stadig nærmere sine grenser og blir stadig mer komplekst [19], og for å fortsatt kunne gi god forsyningsikkerhet er det viktig å kunne overvåke nettet nøye og med høy frekvens for å raskt fange opp feil eller forstyrrelser. En PMU har meget høy samplingsrate på 30 målepunkter per sekund [30], og gir med det et godt grunnlag for smartere overvåkning av nettet. I følgende seksjon vil PMUs historie, virkemåte og anvendelser bli beskrevet for å gi et godt grunnlag for forståelse av videre analyse og diskusjon i oppgaven. For grundigere og mer teknisk beskrivelse av PMU henvises det til [26].

3.1 Historie

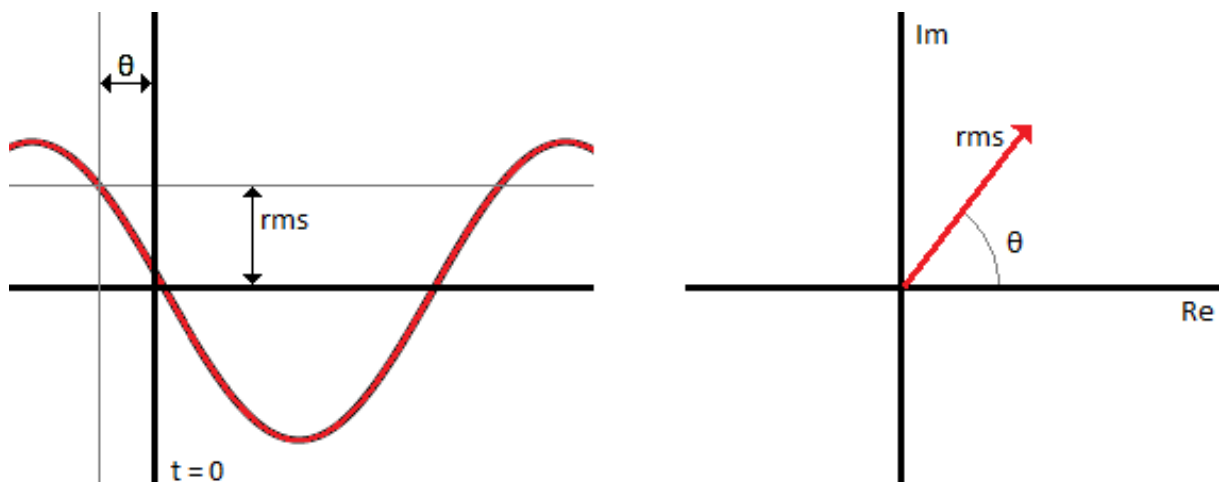
Dagens PMU-teknologi kan sies å stamme fra Symmetrical component distance relay (SCDR), et relévern utviklet tidlig på 1970-tallet [25]. Den gang var ikke mikrodatamaskinene kraftige nok til å kjøre algoritmene for avstandsvernet. Løsningen på dette ble SCDR, som bruker symmetriske komponenter av strøm og spenning til å forenkle beregningene. Over tid ble mikrodatamaskiner kraftigere, og SCDR var ikke lenger nødvendig. SCDR's effektive måte å måle symmetriske komponenter, og dens positive strøm- og spenningssekvenser, har vist

seg å være nyttig i mange kraftanalyseprogrammer.

I 1965 var det et katastrofalt strømbrudd i Nord-Amerika, det ga motivasjon for å utvikle bedre systemer for å overvåke kraftsystemets tilstand i sanntid [25]. Mye forskning ble gjort, og en av teknologiene som kom ut av dette var PMU. Den ble utviklet ved Virginia Tech på 1980-tallet, og de første prototypene ble produsert i 1991 av Macrodyne, i samarbeid med Virginia Tech [18].

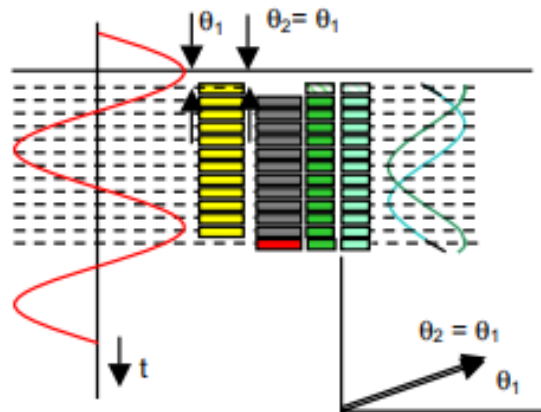
3.2 Måleteknikk

En PMU måler en fasevektor (phasor), som er en måte å presentere en sinuskurve på. Lengden til fasevektoren tilsvarer rms-verdien til sinuskurven og vinkelen tilsvarer vinkelen mellom maksverdien til sinuskurven og akse (t=0), gitt som θ . Dette er illustrert i figur 3.1. En spennings-fasevektor kan uttrykkes slik: $V = |V|e^{j\theta}$, hvor V er spenning [1].



Figur 3.1: Illustrasjon av en fasevektor. Til høyre er sinusbølgen, og til venstre er fasevinkelen. Im står for imaginær akse, og Re for reell akse.

Fasevektorene beregnes basert på tidsstemplet data, og en av de mest brukte metodene for dette er Diskret fourier transformasjon(DFT) [25]. En mer beregningseffektiv metode er rekursiv DFT, hvor bidraget fra siste måling legges til den estimerte fasevektoren samtidig som den eldste målingen trekkes fra. Frekvensen til sinuskurven kan variere med en liten andel, men om den har konstant, nominell frekvens vil rekursiv DFT produsere en konstant fasevektor. Rekursiv DFT er illustrert i figur 3.2.

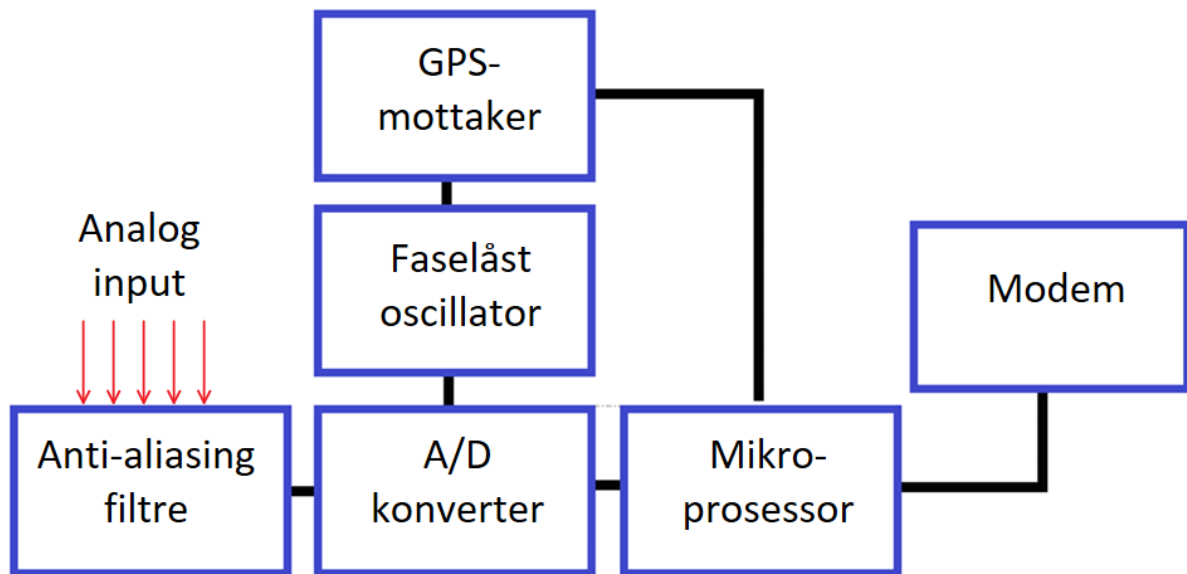


Figur 3.2: Illustrasjon av rekursiv DFT, hvor gule punkter er registrerte data, grå er estimater, grønne er cosinus-andelen og gult er sinus-andelen til fasevektoren. Illustrasjonen er hentet fra [25].

3.3 Måleenhet

Dagens PMU har en GPS(Global positioning system)-mottaker, denne tar inn GPS-signaler med frekvens en puls per sekund(pps) [25]. GPS-signalet har en nøyaktighet på over $1\mu s$, som er mer enn godt nok for at estimatene gir et godt grunnlag for analyse av kraftsystemets tilstand.

GPS-signalet brukes til å generere tidsstemplede datapunkter i en faselåst sløyfe [25], dette betyr at datapunktene vil ha konstant fase i forhold til et referansesignal [23]. Analoge data, hentet inn fra strøm- og spenningstransformatorer [30], blir tidsstemplet ved hjelp av de genererte tidsstemplede datapunktene [25]. Videre brukes disse dataene til å beregne fasevektorene til fasespenning og -strøm ved hjelp av rekursiv DFT. Ved å kombinere de kalkulererte fasevektorene lages positive sekvens-målinger, som til slutt tidsstemples. Et funksjonelt blokkdiagram av en PMU er gitt i figur 3.3.



Figur 3.3: Illustrasjon av en PMU. Mikroprosessor beregner de tidsstemplede fasevektorene ut ifra input fra det digitale signalet fra A/D konverteren og GPS-signalet. Modemet viderefører resultatet slik at det kan brukes til videre analyse av kraftsystemet. Basert på [25].

3.4 Mikro-PMU

PMU blir hovedsakelig anvendt i transmisjonsnett, men det er også blitt utviklet PMU'er spesielt tilpasset distribusjonsnett [18]. Disse kalles μ PMU og gir mer nøyaktige målinger, som er nødvendig i distribusjonsnett hvor avvik i fasevinkel er mindre. Vanlige PMU'er har typisk en nøyaktighet på 1° , mens en μ PMU kan ha en nøyaktighet ned mot 0.01° .

3.5 Anvendelser

Et stadig mer komplekst nett krever nøyere overvåking for å opprettholde smidig drift. Data fra PMU gir grunnlag for flere løsninger for både overvåking, kontroll og vern, og utbredelsen av ulike anvendelser i ulike deler av verden er vist i tabell 3.1. Noen anvendelser vil beskrives ytterligere i følgende delkapittel.

Tabell 3.1: Oversikt over utbredelser av ulike anvendelser i ulike deler av verden. T = Testfase, P = planleggingsfase. Hentet fra [3].

PMU-anvendelse	Nord-Amerika	Europa	Kina	India	Brasil	Russland
Etterforstyrrelseanalyse	✓	✓	✓	P	T	✓
Stabilitetsovervåkning	✓	✓	✓	P	P	✓
Overvåkning av termisk overlast	✓	✓	✓	P	P	✓
Restaurering av kraftsystemet	✓	✓	✓	P	P	P
Modellvalidering	✓	✓	✓	P	T	✓
Tilstandsestimering	P	P	P	P	P	P
Sanntidskontroll	T	T	T	P	P	P
Adaptivt vern	P	P	P	P	P	P
Wide area-stabilisator	T	T	T	P	P	P

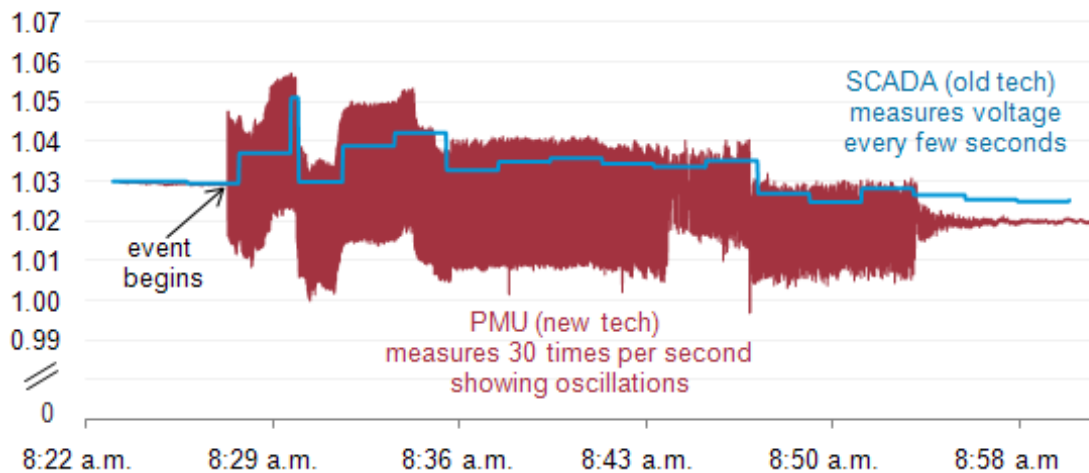
3.5.1 Tilstandsestimering

For å drifte et kraftnett er det essensielt å ha oversikt over tilstanden til nettet. Ved å vite tilstanden og kunne se tegn på kommende utfordringer kan nettet driftes mer effektivt og sikkert. Tilstandsestimering går ut på å ta inn målinger fra nettet og gi ut estimerte tilstandsvariabler [18]. Dagens mest utbredte teknikker for tilstandsestimering ble utviklet på 1970-tallet og er avhengige av å måle aktiv og reaktiv effektlyt samt spenningsverdier på stasjoner/samleskinner [25].

Det finnes to typer tilstandsestimering; statisk og dynamisk. Det mest utbredte er statisk, hvor en antar at systemet er statisk, altså ikke endrer seg, mellom målinger [26]. Denne antagelsen er viktig da det i tradisjonelle systemer for tilstandsestimering tok tilstrekkelig lang tid mellom hver datainnhenting til at systemet potensielt kunne ha forandret seg. I dynamisk tilstandsestimering antar en at systemet forandrer seg kontinuerlig. En må da enten ha høyfrekvente målinger for å fange opp forandringene, eller modellere endringene basert på de data man har.

Tradisjonelt har SCADA(Supervisory control and data acquisition) gitt input til systemer for tilstandsestimering, hvor dataene ikke har vært synkronisert [18]. Her er også datainnhenting så treg at det dreier seg om statisk tilstandsestimering. Ved å bruke data fra PMU kan nøyaktigheten til tilstandsestimering forbedres mye, og en kan bevege seg mot dynamisk tilstandsestimering. Dette er illustrert i figur 3.4 hvor en får et mer dynamisk bilde av en situasjon med PMUs høye samplingsfrekvens. En annen fordel er også at det ikke er nødvendig med en PMU på hver stasjon/samleskinne for å kunne gi et fullstendig estimat av systemet [25]. Ved å ta i bruk PMU-data til tilstandsestimering kan man enten legge det til det eksiste-

rende systemet, og bruke PMU-data i tillegg til SCADA, eller sette opp et helt nytt system som kun baserer seg på PMU. Ofte vil det være mest aktuelt å legge til PMU-data i et eksisterende system. Det innebærer hovedsakelig at estimatene også baseres på vinklene til fasevektorene, da systemet allerede har målinger på spenningsstørrelsen. Dette gir flere fordeler og gir problemstillingen symmetri [26]. Videre er det også flere utfordringer med en slik implementering, for eksempel at en ikke kan være sikker på når de tidsstemplede dataene fra en PMU er plassert i forhold til de konvensjonelle dataene fra SCADA.



Figur 3.4: PMU- og SCADA-måling av spenningsforstyrrelser, hvor spenningsforstyrrelse er på y-aksen. Hentet fra [6].

3.5.2 Transientstabilitet

I kraftnettet oppstår en transient når systemet går fra en stabil tilstand til en annen [33]. Årsaker til dette kan for eksempel være lynnedslag i eller nær en kraftlinje, eller en omkobling ved hjelp av brytere (vanligste årsak). Under transientene utsettes systemet for meget forhøyede strømmer som potensielt kan skade komponenter i kraftsystemet. Et kraftsystems transientstabilitet er systemets evne til å opprettholde synkron drift under en større feil eller forstyrrelser [24]. Ved store nok forstyrrelser på nettet kan det medføre tap av synkronisering, dersom forsøk på å rette opp feilen ikke er vellykket. Dette er en form for transientustabilitet og utvikles meget raskt etter forstyrrelsen inntreffer. Transienter har en ikke-lineær form og utvikles meget raskt (μs - ms). Dette gjør at transientstabilitet både er en viktig og vanskelig utfordring å både vurdere og kontrollere. Med PMUs raske samplingsrate kan dens målinger brukes til å fange opp slike transienter.

3.5.3 Kontroll av kraftsystemet

Tradisjonelt har kontroll av kraftsystemet vært basert på lokale data [26]. For eksempel i form av tilbakemeldingskontroll eller at kontrollhandlinger utføres basert på matematiske modeller av systemet. Med inntoget av PMU som gir nær-sannetidsmålinger har det blitt skapt muligheter for kontroll basert på målinger av systemet som kontrolleres. Ventetiden på PMU-dataene er uproblematisk, særlig fordi dataene er tidsstemplet og dermed gir et bilde av systemet meget kort tid før selve analysen. Eksempler på typer kontroll PMU kan støtte opp om er [26][18]:

- HVDC-systemet
- Eksiteringskontroll
- Kraftsystemsstabilisatorer
- FACTS-kontroll (Fleksible AC-transmisjonssystemer)
- Frekvenskontroll
- Lokalisering og måling av frekvensovertoner

Også inn mot spenningskontroll kan PMU bidra. Stabil spenning i nettet er viktig for å hindre skade på utstyr som for eksempel overoppheting av motorer og generatorer [18]. Ved å opprettholde stabil spenning vil en også kunne holde transmisjonstapene i ledningene så lave som mulig. I det ekstreme kan en få spenningskollaps og fullstendig utfall av nett. Dette kan unngås ved å bruke PMU-data til å opprettholde og forbedre systemets spenningsstabilitet.

3.5.4 Vern av kraftsystemet

Vern av kraftsystemet er viktig for å unngå strømbrydd og skade på komponenter. PMU-data har gitt løsninger på flere krevende problemer innenfor vern [26]. Differensialvern er et vern hvor PMU har gitt betydelige forbedringer, og ekte differensialvern har blitt muliggjort ved hjelp av PMU-data. Eksempler på utfordringer forbedret differensialvern kan hjelpe med å løse er:

- Vern av seriekompenserte linjer

- Vern av multiterminale linjer

Innen vern av funksjoner med relativt treg responstid, som for eksempel backup-vern til distansevern, har PMU vært spesielt nyttig [26].

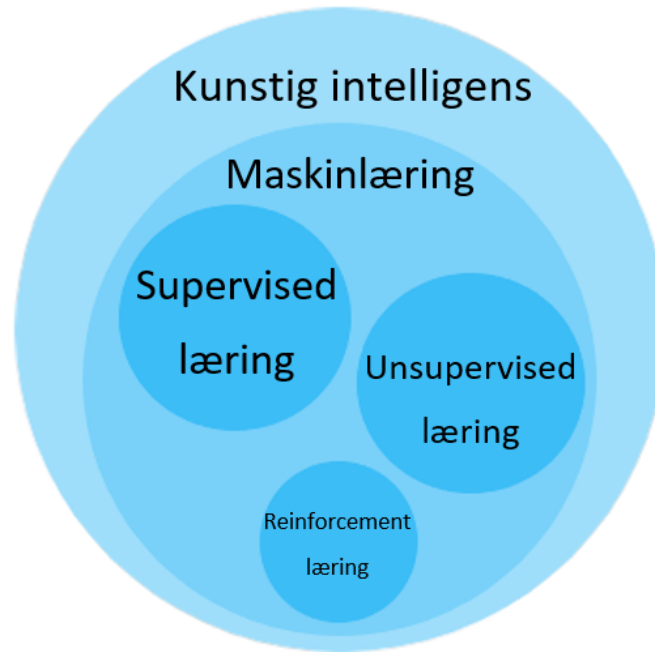
Når kraftsystemet utsettes for høyt stress kan siste utvei være å føre deler av nettet over i øydrift, som betyr at det drives uten tilkobling til øvrig nett. Området som går over i øydrift skal ideelt sett ha balanse mellom produksjon og forbruk, og ofte må enten produksjon eller forbruk justeres for å oppnå dette [26]. Overgang til øydrift gjøres ofte gjennom systemplan for integritetsbeskyttelse (SIPS), som baserer seg på omfattende modeller med forventede verdier av ulike variabler, som for eksempel planlagte og ikke planlagte utfall og lastnivåer. Ofte kan de reelle forholdene være ulike de SIPS setter som antagelser, og systemet vil ikke være optimalt for de rådende forholdene. Her kan nær-sanntidsdata fra PMUs støtte eller erstatte SIPS ved å gi aktuell informasjon om forholdene i kraftsystemet. PMU-data kan for det første brukes til å bestemme om systemet er på vei mot en ustabil situasjon og om øydrift vil være nødvendig. For det andre kan PMU-dataene optimalisere hvilket geografisk område som bør gå i øydrift. For eksempel basert på hvordan produksjon og forbruk kan optimaliseres slik at øydriften vil være i balanse, samtidig som resten av systemet også forblir i balanse.

Kapittel 4

Maskinlæring

Maskinlæring kan defineres på mange måter, og fellesnevneren for de fleste er at en maskin har “evnen til å lære av historiske data ved å finne en sammenheng og dermed komme med en forutsigelse“ [4]. Maskinlæring faller inn under kunstig intelligens, og ofte brukes de to begrepene om hverandre. Det finnes også kunstig intelligens som ikke inneholder maskinlæring, og da for eksempel er et sett med regler som er kodet av mennesker. Ofte inneholder kunstig intelligens maskinlæring, som støtter opp om å kunne lage så smarte systemer som mulig. Forholdet mellom maskinlæring og kunstig intelligens, samt en oversikt over maskinlæring er gitt i figur 4.1. Et annet begrep det er nyttig å definere er stordata. Dette er store mengder data, som ofte er det som gis som input til en maskinlæringsalgoritme og er det algoritmer lærer av. Det er ikke essensielt å ha stordata for å drive maskinlæring, men store datamengder kan ofte støtte opp om funn og gi bedre grunnlag for forutsigelser som ligger nærmere virkeligheten. Videre vil det i denne oppgaven bli brukt begrepene hyperplan, instans/datapunkt, variabel, forklarende variabel og målvariabel når ulike metoder beskrives. Disse er forklart i kapittel 2.

Når maskinlæring skal anvendes er det snakk om å trene opp en modell for det som skal ses på. For å vite hvor god modellen er, er det viktig å ha data både å teste på og å validere med. Derfor deles data ofte inn i et treningsdatasett og et testdatasett. Da vil maskinlæringsalgoritmen først trene på treningsdatasettet, før en tester nøyaktigheten til modellen med testdatasettet. Når en evaluerer resultatene sine er det viktig å være kritisk. To vanlige utfordringer innen maskinlæring er overfitting og underfitting. Overfitting er når modellen er overtilpasset testdatasettet, som kan skje hvis modellen justeres til å passe for godt til testdatasettet. Da vil modellen gi kunstig gode resultater, og vil ikke prestere like godt når den



Figur 4.1: Oversikt over forholdet mellom kunstig intelligens, maskinlæring og de ulike type-
ne maskinlæring

møter nye data. Underfitting kan skje om en har for lite data eller for få variabler, slik at modellen ikke klarer å fange opp den oppførselen den er ment til å modellere. Dette kan skje om datasetter er fordårlig eller om det brukes en algoritme som passer dårlig til problemet eller formen på datasettet.

Maskinlæring deles som oftest opp i de tre kategoriene supervised læring, unsupervised læring og reinforcement læring, som sett i figur 4.1. Disse tre kategoriene vil dekket i de følgende delkapitlene, sammen med dyp læring og support vector machines, som begge både kan kategoriseres som supervised og unsupervised maskinlæring.

4.1 Supervised maskinlæring

Når det er snakk om supervised læring, læres modellen opp med en tilhørende fasit. En av variablene i datasettet er altså den variabelen som skal gjettes på. For eksempel kan en ha et datasett over hvilke karakterer 100 studenter har fått i tidligere fysikk- og matematikkfag på NMBU, inkludert faget FYS377. Da kan supervised maskinlæring brukes til å forutsi hvilken karakter en student vil få i FYS377 basert på karakterer i tidligere fag. Matematisk kan det

fremstilles enkelt som

$$y = f(X)$$

Hvor y er målvariabelen og X er en vektor med de forklarende variablene. Målet med supervised maskinlæring vil være å bestemme funksjonen f slik at den best mulig beskriver forholdet mellom y og X .

Målvariabelen er ofte én av to typer variabler [7]:

- Kontinuerlige/kvantitative variabler
- Diskrete/kvalitative/kategoriske variabler

Ut ifra dette er det delt inn i to hovedgrupper; regresjon(kontinuerlige variabler) og klassifikasjon(diskrete variabler). I de neste to seksjonene vil disse beskrives, samt sentrale algoritmer innenfor begge.

4.1.1 Regresjon

Regresjon defineres her som: *Statistisk teknikk som beskriver forholdet mellom en variabel og en eller flere andre variabler*. Forholdet kan være både lineært og ulineært, hvor denne delen videre vil presentere lineær regresjon. Det finnes flere andre former for regresjon, blant annet "Partial least squares regression", "Principal component regression", "Ridge regression", "LASSO" og "Elastic net". Disse vil ikke bli dekt i denne oppgaven.

4.1.1.1 Lineær regresjon

Algoritmer for lineær regresjon ble utviklet før datamaskinens tid, men er fortsatt et viktig verktøy innen maskinlæring [7]. De kan ofte prestere bedre enn mer avanserte algoritmer, særlig når det er få treningsscenarier, lite data og lav signal-til-støy-ratio. En lineær regresjons-modell har formen

$$f(X) = \beta_0 + \sum_{j=1}^p X_j \beta_j \quad (4.1)$$

Hvor X_j er den j 'te inputvariabelen i vektoren X og β_j er en koeffisient tilhørende den j 'te inputvariabelen. For å finne funksjonen f som best forutsier y må en optimalisere koeffisi-

entene β_j . Dette gjøres ved å minimalisere en residual-sum, altså en sum av avvikene mellom den gjettede og sanne verdien. Den mest populære optimaliseringsmetoden er minste kvadraters metode, hvor en tar kvadratet av avviket og summerer det:

$$RSS(\beta) = \sum_{i=1}^N (y_i - f(x_i))^2 \quad (4.2)$$

Hvor RSS står for residual sum of squares, N er totalt antall datapunkter, y_i er den sanne verdien, $f(x_i)$ er den gjettede verdien og x_i er en vektor med alle variablene i det i 'te punktet. Ved å minimere ligning 4.2 vil en da få den funksjonen f som er best tilpasset datasettet.

4.1.2 Klassifisering

Innenfor klassifikasjon lages en modell som kan plassere datapunkter i ulike klasser basert på verdien til datapunktets variabler. For eksempel kan en klassifisere håndskrevne tall i klassene 0-9 ved at modellen lærer seg hvilke piksler som er farget i hvert av tallene. En klassifiseringsmodell kan være bygd opp på ulike måter, og kan basere seg på blant annet regresjon eller beslutningstrær. Disse to tilnærmingene innenfor klassifisering vil beskrives i de følgende seksjonene.

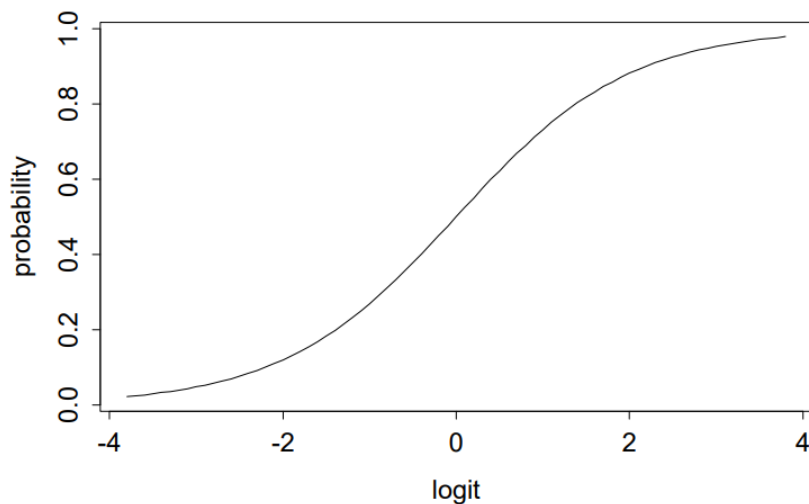
4.1.2.1 Logistisk regresjon

Logistisk regresjon anvendes hovedsakelig på binære klassifikasjonsproblemer, men kan også generaliseres til klassifiseringsproblemer med flere mulige klasser. Det brukes her regresjon inn mot klassifisering. Grunnprinsippet er at en bruker regresjon til å estimere hvilken sannsynlighet, p , det er for de forskjellige utfallene, og dette legges til grunn for klassifiseringen. De to klassene gis ofte verdiene 0 og 1, men det som ligger bak for eksempel kan være om en pasient overlever(1) eller dør(0) etter å ha fått en kreftdiagnose. I regresjonen estimeres altså sannsynligheten for at det ene utfallet finner sted, og da samtidig det andre.

Før klassifiseringen finner sted går en fra sannsynlighet, p , til odds [27]. Odds er gitt fra formel 4.3.

$$odds = \frac{p}{1-p} \quad (4.3)$$

Videre tar en den naturlige logaritmen til oddsen, som kalles *logit*; eller log-odds. Logit av



Figur 4.2: Transformasjonen fra sannsynlighet, p , til logit av sannsynlighet gjennom formel 4.4. Hentet fra [27]

sannsynligheten er her gitt som η , vist i formel 4.4.

$$\eta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (4.4)$$

Logit-funksjonen er en transformasjon, og vises i figur 4.2. For å bygge en logistisk regresjonsmodell antas det videre at logit av sannsynlighetene, $\text{logit}(p)$, er en lineær kombinasjon av de forklarende variablene for hvert i 'te datapunkt:

$$\text{logit}(p_i) = x_i \beta \quad (4.5)$$

Hvor x_i er en vektor av forklarende variabler og β er en vektor av tilhørende koeffisienter for de i 'te datapunktet. β bestemmes gjennom trening av modellen på treningsdatasettet. Ved å kombinere formel 4.4 og 4.5 får en formel 4.6

$$p = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \quad (4.6)$$

Basert på den estimerte sannsynligheten for de ulike utfallene kan en nå klassifisere ved at en setter en terskel for å skille klassene [32]. Gitt at en har to klasser vil en naturlig terskel være på $p = 0.5$, illustrert i figur 4.3. Fra eksempelet om en pasient overlever eller ikke ville en da klassifisert at en pasient vil overleve om sannsynligheten er over 0.5.



Figur 4.3: Eksempel på sannsynlighetskurve som kan brukes til klassifisering. Her er terskelen mellom to klasser satt til $p = 0.5$.

4.1.2.2 K-nearest neighbour

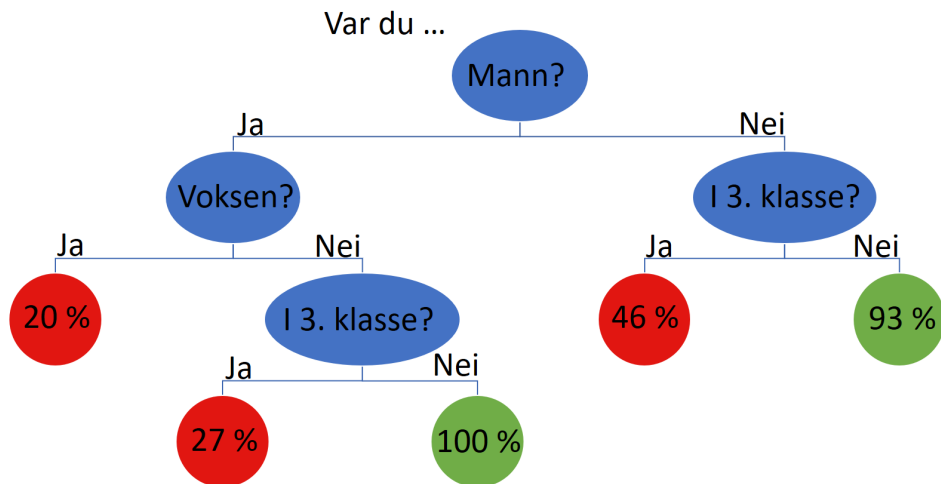
Denne algoritmen krever ingen trening, kun et treningssett, og er minne-basert [7]. Når en ny instans skal klassifiseres undersøkes de k nærmeste datapunktene. Disse "stemmer" over hva den nye instansen skal klassifiseres som, og den klassen som har majoriteten blant dem er det den nye instansen klassifiseres som. Om de er likt trekkes det tilfeldig hva den blir. De sentrale utfordringene er hvilken funksjon en skal bruke for avstand og hvor mange naboer, k , som skal inkluderes. Tross sin enkle virkemåte har k -nearest neighbour vist seg å være en kraftig algoritme som for eksempel kan klassifisere håndskrevne tall.

4.1.2.3 Beslutningstrær

Beslutningstrær har blitt brukt lenge og flere av de sentrale algoritmene ble utviklet på 80- og 90-tallet, for eksempel C4.5(1993) og CART(1984) [13]. Hovedprinsippet i beslutningstrær er at en deler opp datasettet i ulike deler hvor et datapunkt klassifiseres i den klassen som er mest frekvent i den delen. Delene genereres ved å stille ulike spørsmål og dermed dele opp datasettet. Hvert spørsmål burde dele opp datasettet i grupper med så lite variasjon i klasser som mulig. Dette gjøres ved hjelp av en urenhetsfunksjon som en ønsker å minimere i hver oppdeling [16]. Resultatet av dette er et beslutningstre, illustrert med et eksempel i figur 4.4. Der treet starter (øverst) kalles roten, hvert punkt hvor spørsmål stilles kalles en node

og de endelige endepunktene/klassene kan kalles blader. En av de store fordelene med beslutningstrær er at de er meget logiske og gir en modell som er enkel å forstå. Modellene har en tendens til å bli relativt enkle, som førte til at disse typene algoritmer opplevde synkende popularitet en stund, men den siste tiden har teknikker som random forest gitt dem en renessanse [32]. Denne vil nå beskrives, og deretter to populære algoritmer; CART og C4.5.

Sannsynlighet for at du døde på Titanic:



Figur 4.4: Beslutningstre som viser sannsynligheten for å dø dersom du var på Titanic. Det er et resultat av å kjøre en beslutningstre-algoritme med et datasett over hvem som overlevde og ikke. Med 3. klasse menes hvilket område av båten passasjerer befant seg på. Basert på [20]. Der treet starter (øverst) kalles roten, hvert punkt hvor spørsmål stilles kalles en node og de endelige endepunktene/klassene kan kalles blader (de røde og grønne).

Random forest Dette er en ensemble-teknikk hvor grunnprinsippet er å ta mange enkle beslutningstrær for sammen å kunne gjøre bedre forutsigelser enn ett enkelt tre [2]. En random forest skapes ved generering av tilfeldige beslutningstrær basert på et tilfeldig utvalg av dataene. Utvalget trekkes tilfeldig fra det totale datasettet og legges tilbake før det trekkes et nytt utvalg til neste beslutningstre. Når en ny instans skal klassifiseres kjøres den gjennom alle beslutningstrær og deretter "stemmes" det over hva den endelige klassifiseringen vil være. Dette skjer ved at den klassen instansen har endt opp i i flest beslutningstrær er den random forest-algoritmen plasserer den i. Om en har kontinuerlige variabler kan en ta et gjennomsnitt, eventuelt vektet, av alle estimater og gi dette som det endelige estimatet.

CART - Klassifiserings- og regresjonstrær Dette er en gruppe algoritmer for et beslutningstre som deler opp binært og rekursivt [34]. Den kan også analysere kontinuerlige variabler både som forklarende variabler og målvariabel. Her brukes en urenhetsfunksjon kalt Gini-

index'en for å lage treet [16]. Ved hjelp av denne bygges et stort tre, som deretter beskjæres for minimering av estimert feilklassifiseringsfeil. Etter beskjæringen har algoritmen produsert en rekke "nestede", beskjærte trær [34]. Det originale treet beskjæres altså litt etter litt, og for hver beskjæring finnes et "nestet" tre som potensielt kan være det optimale treet. Algoritmen vil ikke selv foreslå hva som vil være det beste treet, men dette kan finnes ved å teste på uavhengige data eller ved hjelp av kryssvalidering.

C4.5 Denne algoritmen er en utvidelse av en annen algoritme, ID3 [34]. Også i C4.5 gros et beslutningstre, her ved å hjelp av en informasjonsbasert, heuristisk urenhetsfunksjon [16]. Også her beskjæres treet, men her med mål om å ende opp med det optimale treet. Beskjæringen foregår fra bladene og oppover, hvor en node erstattes av et blad dersom dette ikke medfører økt estimert feilklassifisering [34]. En betydelig forskjell fra CART er at CART deler binært i hver node, mens det i C4.5 kan være flere utfall i hver node.

C4.5 kan også uttrykkes som en regel-klassifikator hvor hver instans klassifiseres ved å sjekke det opp mot et sett regler [34]. Instansen klassifiseres da i den klassen hvor det først oppfyller alle kriteriene i regelen. Reglene produseres fra det ubeskjærte treet ved å følge alle mulige ruter ut fra roten og til hvert enkelt blad. Deretter forenkles hver regel ved å fjerne kriterier som bidrar til økt feilklassifisering eller bidrar lite til rett klassifisering. Til slutt velges et sett av disse forenklete reglene som gir minst feil på treningssettet, samtidig som det velges en standardklasse som instanser klassifiseres som om de ikke passer til noen av reglene. En ulempe med regel-klassifikator er at den krever mye tid og minne etterhvert som datasettet øker. Regel-klassifikatorer øker med en faktor 300, mens beslutningstrær kun øker med en faktor 44 ved tidobling av antall instanser.

4.2 Unsupervised maskinlæring

Innenfor denne delen av maskinlæring har en ikke en målvariabel og formålet er heller å finne sammenhenger i datasettet. Ved å finne den underliggende strukturen kan en se mønstre og/eller gruppere datasettet [4]. Videre kan unsupervised maskinlæring deles inn i clustering og assosiasjon. Disse to, noen vanlige algoritmer innenfor hver og prinsipalkomponentanalyse vil beskrives dypere i de følgende seksjonene.

4.2.1 Clustering

Clustering går ut på å gruppere lignende datapunkter ved å finne grupper med minimal intern varians. Instansene i en gruppering, eller “klynge“, ligner altså mer på hverandre enn de ligner på andre instanser [7]. En kan også dele inn datasettet hierarkisk, hvor en lager klynger som er sortert etter en gitt skala. Clustering kan deles inn fire hovedtyper [12]:

- **Distansebasert:** Her finner en klynger ved å se på hvor nære datapunkter er hverandre i rommet. Et eksempel er hierarkisk clustering.
- **Fordelingsbasert:** Her lages det klynger basert på hvor sannsynlig det er at punktene i en klynge tilhører samme fordeling, for eksempel normal eller gauss. Et eksempel på en slik algoritme er Expectation-maximiseringsalgoritmen.
- **Tetthetsbasert:** Klyngene finnes her basert på tettheten av datapunkter, hvor ulike tetthetsområdet setter som ulike klynger. Eksempler på dette er algoritmene DBSCAN og OPTICS.
- **Centroid/klyngesenter-basert:** Her bestemmes klyngene ut ifra et datapunkts nærhet til et punkt som settes som klyngesenter (“centroid”). Et eksempel på dette er k-means clustering, som vi bli beskrevet senere i denne seksjonen.

4.2.1.1 k-means clustering

Her velges klyngene ut ifra avstanden til et bestemt klyngesenter [7]. Algoritmen fungerer slik:

1. Algoritmen gjetter et gitt antall klyngesentre, for eksempel 3.
2. Alle datapunktene grupperes ut ifra hvilket klyngesenter de er nærmest ut ifra kvadrert euklidsk avstand.
3. Nytt klyngesenter settes som gjennomsnittet av alle datapunktene som tilhører den klyngen.
4. Punkt 2 og 3 repeteres til det konvergerer mot en endelig gruppering.

Siden en bruker euklidsk avstand gjør dette at algoritmen er avgrenset til kvantitative/ kontinuerlige variabler [7]. To av hovedutfordringene ved k-means clustering er å velge antall klynger og hvor de initialiserende klyngesentrene skal plasseres.

4.2.2 Assosiasjon

Assosiasjon er et område innenfor unsupervised maskinlæring hvor en forsøker å beskrive datasettet ved hjelp av regler. Regler kan for eksempel finnes som grupper av variabler som har samme verdi til samme tid [7]. Assosiasjon brukes som regel på binære datasett, altså som kan ha mange variabler men hver variabel har en av to verdier. Dette kalles gjerne "handlekurv"-analyse. Dette kan da beskrives som at hver variabel i et datasett er en vare i en butikk og har enten verdien 1 (ble kjøpt) eller 0 (ble ikke kjøpt). Hver instans er da en kunde, hvor kundens handlekurv representeres av 1 eller 0 på hver variabel ut ifra om kunden kjøpte de forskjellige varene. En kan da finne regler som beskriver hvilke variabler/varer som ofte kjøpes sammen. Dette kan være nyttig for planlegging av hyller eller salgskampanjer.

En meget populær algoritme for assosiasjon er apriori, som leter etter større og større sett med variabler av lik verdi for å finne de settene som opptrer oftest og da kan være gode regler [34].

4.2.3 PCA - Prinsipalkomponentanalyse

PCA er en av mest brukte og eldste (utviklet første gang i 1901) metodene for å redusere dimensjonene til et datasett slik at en fortsatt beholder informasjonen som finnes i datasettet [11]. For å gjøre dette finner en nye variabler, kalt prinsipale komponenter(PC), som er lineære funksjoner av de originale variablene. Disse nye variablene velges slik at de i synkende grad forklarer så mye varians i data settet som mulig. Altså vil den første variabelen beskrive mest varians, den neste nest mest og så videre. For å finne disse variablene løses et egenverdi/egenvektor-problem.

Gitt at en har et datasett bestående av n instanser, hver med p variabler, dette gir en $n \times p$ -matrise, X . Her er kolonne j vektoren x_j , som består av alle instansers verdi av den j 'te variabelen. For å finne første prinsipalkomponent finner en den lineære kombinasjonen av kolonner som gir mest varians, gjennom datasettets kovariansmatrise. Dette viser seg å være

et egenverdi/egenvektor-problem på formen

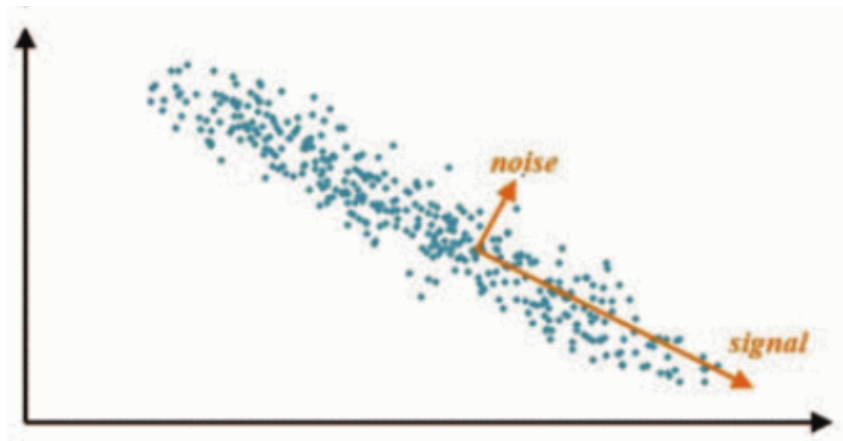
$$\mathbf{S}a_1 = \lambda_1 a_1$$

Hvor \mathbf{S} er kovariansmatrisen, λ er egenverdien og a er egenvektoren og vektoren som gir den lineære kombinasjonen av kolonner med høyest varians:

$$\sum_{j=1}^p a_{1j} x_j = \mathbf{X}a_1$$

Hvor $\mathbf{X}a_1$ er første prinsipalkomponent. λ_1 gir også variansen til den lineære kombinasjonen a_1 den tilhører. Dermed ønsker en å maksimere λ_1 for å finne den lineære kombinasjonen med mest varians. Videre skal neste prinsipalkomponent stå ortogonalt på den forrige, og ha null korrelasjon med alle de foregående prinsipalkomponenter. Dette brukes som begrensning når videre prinsipalkomponenter skal finnes.

Det er vanlig å ikke beregne alle prinsipalkomponentene, men å sette en grense for hvor stor andel av total varians en krever at de skal forklare. En vanlig grense er 70 %, hvor de som forklarer mindre blir ekskludert. Ofte begrenses det og til to eller tre prinsipalkomponenter med hensyn til visualisering, som i figur 4.5.

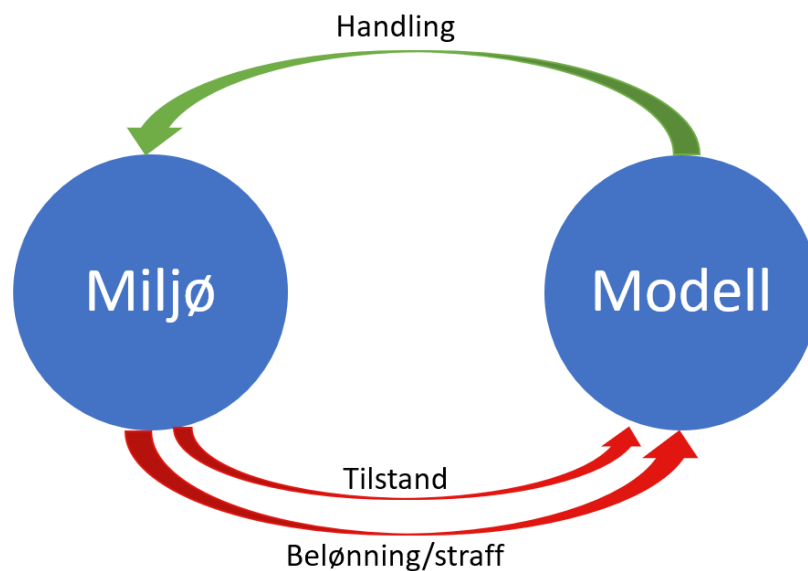


Figur 4.5: Eksempel på PCA hvor de første to prinsipalkomponentene er inkludert, hentet fra [35]

4.3 Reinforcement maskinlæring

Reinforcement maskinlæring kan kalles å “lære med en kritiker”, hvor supervised maskinlæring til sammenligning kan kalles å “lære med en mester” [8]. I reinforcement maskinlæring

utvikles modellen stadig ettersom den får tilbakemeldinger fra miljøet den lærer fra/i på om det den gjør er riktig eller ikke - den får enten belønning eller straff. Basert på dette analyserer den hvordan modellen kan utvikles for å til slutt alltid få belønning. Tilbakemeldingen forteller altså modellen hva den skal gjøre, men ikke hvordan. Dette må maskinlæringsalgoritmen selv finne ut av ved å prøve og feile og analysere tilbakemeldingene i sammenheng med hva den gjør. Reinforcement-algorithmens mål vil være å maksimere den akkumulerte, fremtidige belønningen. Dette gjøres gjennom en "return-funksjon", hvor det finnes tre typer: *Endelig horisont-modellen*, *rabatt-retur (uendelig horisont-modellen)* og *gjennomsnittlig belønningsmodell* [8]. Mer om disse, og dypere beskrivelse, matematisk grunnlag og beskrivelse av noen algoritmer innen reinforcement maskinlæring finnes i [8].

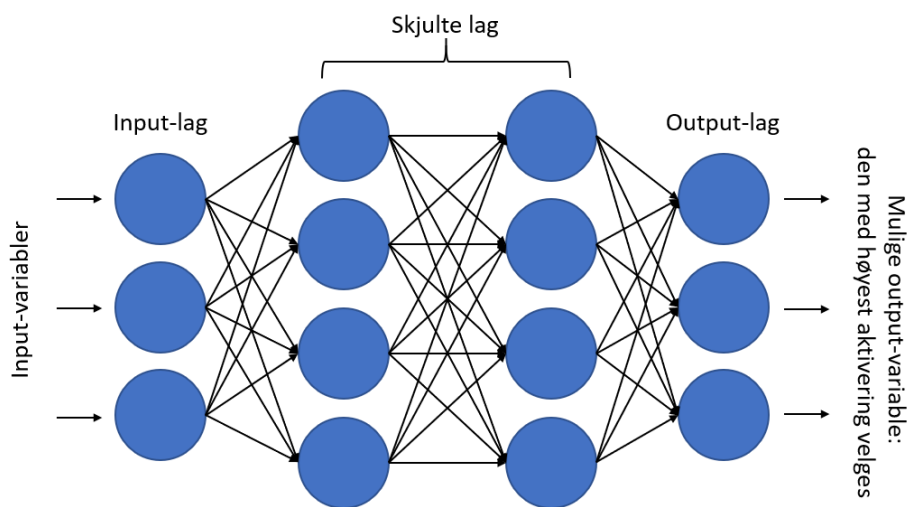


Figur 4.6: Illustrasjon av grunnprinsippet i reinforcement maskinlæring. En modell endrer sin tilstand i et miljø, får tilbakemelding (belønning/straff) og gjør videre en handling/endring i modellen som igjen vurderes av miljøet.

Denne typen maskinlæring brukes for eksempel til å lære roboter å gå, som er inspirert av menneskelig adferd [4]. Også til å lære sjakk har det vist seg nyttig, hvor tilbakemeldinger på hvor godt et trekk var eller rett og slett på om en vant eller ikke bidrar til å lære seg de beste strategiene. En av utfordringene ved reinforcement maskinlæring er at det krever mye minne, da alle tidligere tilstander og deres tilbakemelding må lagres for å kunne gi gode resultater. Med voksende minnekapasitet blir heldigvis denne utfordringen mindre og mindre betydelig og denne formen for maskinlæring har potensiale til å bli mer populær.

4.4 Dyp læring

De metodene og algoritmene som har vært beskrevet så langt baserer seg på matematiske sammenhenger. Dyp læring er derimot inspirert av hvordan den menneskelige hjernen fungerer og lager kunstige nevrale nett. I dyp læring består modellen av flere lag - input-laget, output-laget og et eller flere skjulte lag mellom disse [4]. Når en snakker om nevrale nett menes det som oftest nettverk med kun ett skjult lag, mens innen dype nett er det snakk om flere skjulte lag. Hvert lag består av flere noder, hvor det i input-laget er en node per variabel og i output-laget er en node per mulige output. De skjulte lagene mellom kan ha et annet antall noder. Mellom alle noder er det vekter, disse sier i hvor stor grad aktivering av den første noden skal føre til aktivering i den andre noden. Ved hjelp av dette går et signal gjennom de forskjellige lagene og gir forskjellig aktivering til de ulike nodene, og til slutt velges den output-noden med mest aktivering. En starter altså med et sett variabler i input-laget, beveger seg mot output-laget gjennom de skjulte lagene. Dette er illustrert i figur 4.7. En av styrkene til nevrale nett er at det klarer å lære seg hvilke variabler og variasjoner i variabler den skal være sensitiv på. Dette gjør at algoritmer innen dyp læring har betydelig forbedret *state-of-the-art* innen blant annet stemmegjenkjenning og objektgjenkjenning [14].



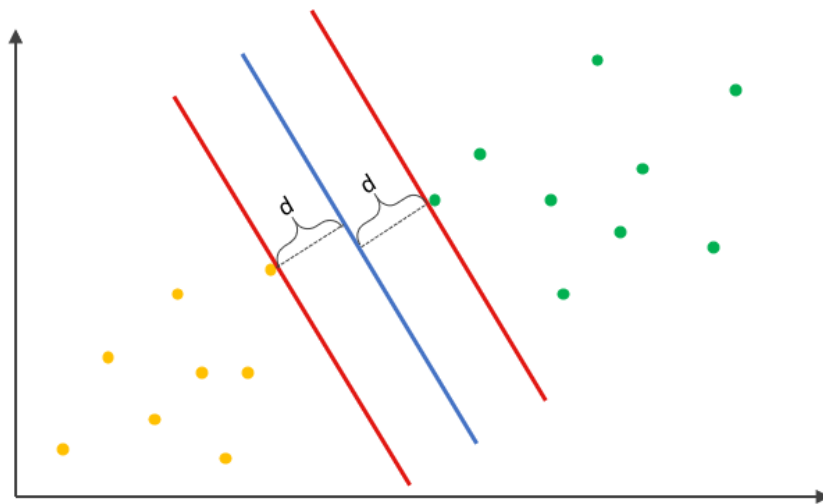
Figur 4.7: Illustrerer hvordan et nevralt nettverk er bygd opp. Svarte piler illustrerer vektene mellom nodene. Det kan være flere enn to skjulte lag.

Vektene mellom nodene bestemmes ved hjelp av at det genereres en gradientvektor som, for hver av vektene, sier noe om hvor mye feilen vil øke eller minske dersom den aktuelle vekten økes [14]. Denne brukes da til å justere vektene mot minimal feil. Et nevralt nettverk trenes ved å gi nettverket et treningssett med målvariablene inkludert. Så kjøres dette treningssettet gjennom nettverket, som justerer vektene for å få minst mulig feil i forhold til den gitte målvariablen for hver instans.

Dyp læring er en metode som kan brukes uten å ha mye domenekunnskap, og gode resultater kan oppnås ved å prøve og feile alene [14]. Dette gjør også at modellene ikke blir like intuitive som for eksempel beslutningstrær. Særlig med de skjulte lagene får man en såkalt “svart boks” som produserer et resultat, men en vet ikke helt hvordan. En negativ side ved dyp læring er at en trenger mye datakraft og mye data for å trene modellen [4]. Når den er ferdig trent derimot, vil den operere raskt. Dyp læring, da særlig dype nett med mange skjulte lag, er veldig fleksible. Et vanlig problem med disse er derfor overfitting. Dette er viktig å ta hensyn til når slike modeller bygges og trenes, og det er da viktig med gode valideringsmetoder for å unngå overfitting.

4.5 SVM - Support vector machines

Support Vector Machines - SVM regnes som en av de mest robuste og nøyaktige maskinlæringsmetodene vi har i dag [34]. Den krever ikke mange eksempler å trene på, og tåler mange dimensjoner/variabler. En SVM-algoritmes virkemåte er at den finner det hyperplanet som best mulig skiller klassene den er satt til å skille mellom. Dette gjør den ved å ta det planet med høyest mulig avstand til de nærmeste punktene i hver enkelt klasse. I et tilfelle med to klasser vil det si at en tar de to punktene som er nærmest hverandre i hver klasse, og finner linjen som maksimerer avstandene til begge disse, se figur 4.8. I flere dimensjoner er prinsippet det samme, men det er flerdimensjonelle plan som skiller i stedet for en linje. Når hyperplanet som skiller klassene er funnet er SVM-modellen ferdig, og videre vil nye instanser plasseres i en klasse avhengig av hvilken side av linjen/hyperplanet den er på. Navnet SVM kommer fra at en kaller de punktene nærmest hverandre i hver klasse for støttevektorer (Support Vectors), i den forstand at det er disse som “støtter opp om” hvor det delende hyperplanet skal være.



Figur 4.8: Illustrasjon av hvordan SVM bygger en modell. Avstanden d til de to nærmeste punktene fra hver klasse skal maksimeres for å finne hyperplanet som best skiller de to klassene.

Kapittel 5

Maskinlæring brukt på PMU-data

5.1 Litteraturanalyse

I denne delen vil sammensetningen av litteraturen som er funnet på fagområdet “PMU og maskinlæring“ analyseres. Formålet er tredelt:

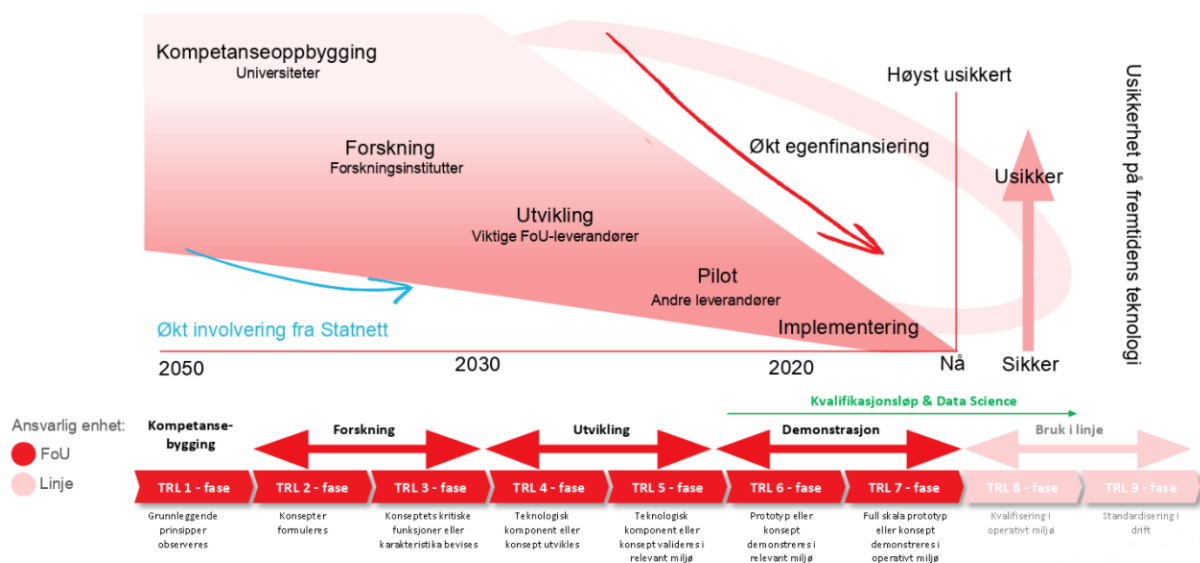
- Undersøke hvor utviklet fagområdet er med tanke på TRL-fase.
- Undersøke hvor satsningen på det er geografisk.
- Identifisere hvilke områder som er mest undersøkt for å senere dykke dypere inn i disse og kunne foreslå mulig videre satsningsområder for det nordiske kraftsystemet.

Dette gjøres ved at sammensetningen av artikler analyseres ved hjelp av grafer som illustrerer aktuelle sammenhenger. Det ble funnet totalt 110 artikler, hvor det er blitt forsøkt å finne et representativt utvalg. I tillegg A er en fullstendig oversikt over alle artiklene som ble funnet i litteraturstudien.

5.1.1 Vurdering av litteratur

For å undersøke hvor utviklet fagområdet er vil det ses på hvilke typer institusjoner som har gitt ut artiklene som er funnet. Dette brukes til å vurdere hvilken TRL-fase området er i, hvor

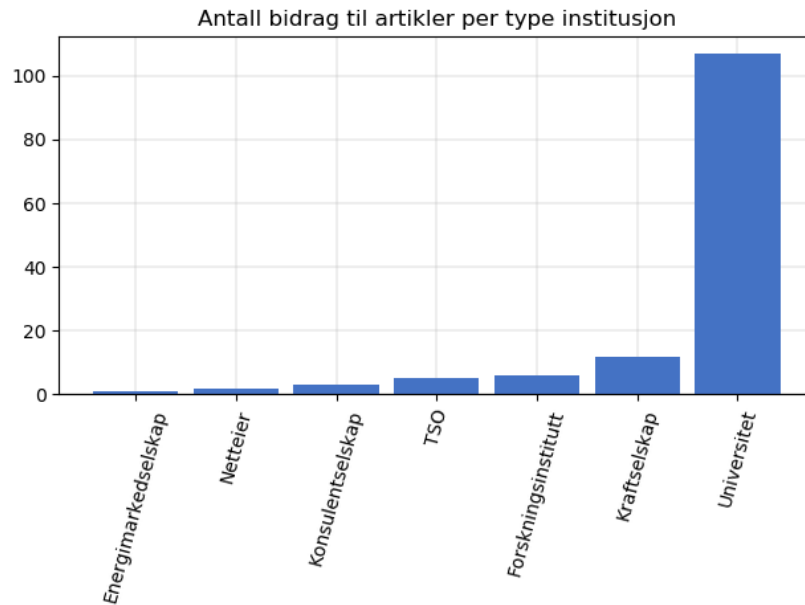
bakgrunnen for ulike TRL-faser er vist i figur 5.1.



Figur 5.1: Figur som viser ulike TRL-faser. Hentet fra [29].

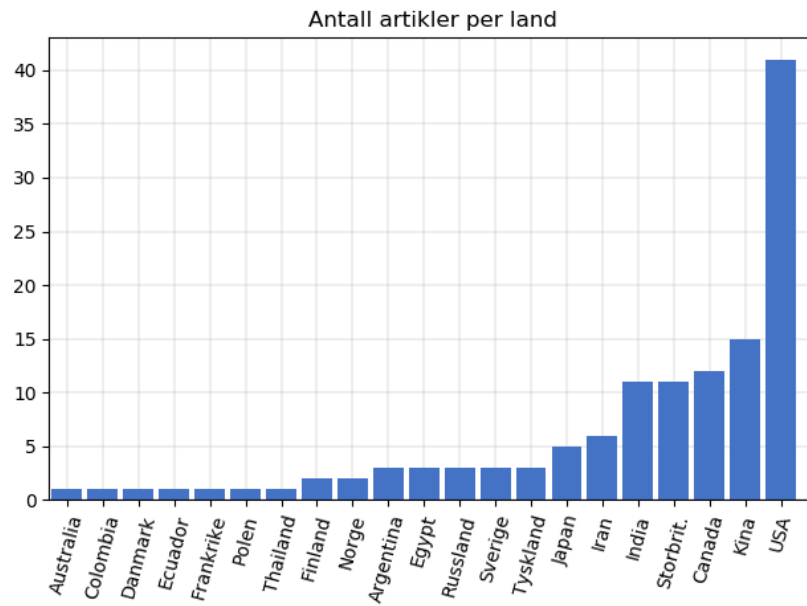
Hvor mange av de forskjellige typene institusjoner som har bidratt til artiklene er vist i figur 5.2. Her ser man en tydelig overvekt av artikler som er gitt ut av universiteter. Dette tyder på at fagområdet fortsatt hovedsakelig befinner seg i TRL-fase 2. At det finnes totalt 23 bidragsytere, 16,9 %, som er TSO'er, kraftselskap, netteiere, energimarkedselskap og/eller konsulentselskap indikerer at det er en utvikling i retning mot TRL-fase 3 og 4. I denne figuren får imidlertid hver type institusjon ett poeng for hver artikkel den har bidratt til. Dette betyr hvert poeng til en av institusjonene ikke betyr at de alene har gitt ut en artikkel, men en TSO kan for eksempel ha bidratt til en artikkel et universitet har gitt ut. Ved gjennomgang av artiklene ble det funnet at kun én artikkel ble gitt ut av en av de tidligere oppramsede institusjonstypene, da en netteier. Det var i tillegg 5 artikler hvor en av de tidligere oppramsede var hovedbidragsyter til artikkelen. Alt i alt peker dette på at fagområdet i meget stor grad fortsatt er på forskningsstadiet, TRL-fase 2, men har begynt utviklingen mot implementering.

Videre er den geografiske utstrekningen av fagområdet undersøkt, og blir illustrert i figur 5.3. Her ser en at det både er mange forskjellige land som bidrar, i alt 21, og samtidig ett land, USA, som skiller seg ut med 32,3 % av bidragene. Dette er ikke overraskende da PMU ble utviklet i USA. Universitet som utviklet den, Virginia Tech, er også blant universitetene med flest bidrag. Også fire andre land, Kina, Canada, Storbritannia og India, bidrar betydelig med 11 eller flere bidrag. Om en ser på enkeltpersoners bidrag er det tre som skiller seg ut med å bidra til 6 eller flere artikler: I. Kamwa (École de Technologie Supérieure), S. R. Samantaray (Indian Institute of Technology) og V. Vittal (Arizona State University). Disse hører til institusjoner i henholdsvis Canada, India og USA, som igjen viser at det er en geografisk bredde innenfor feltet.

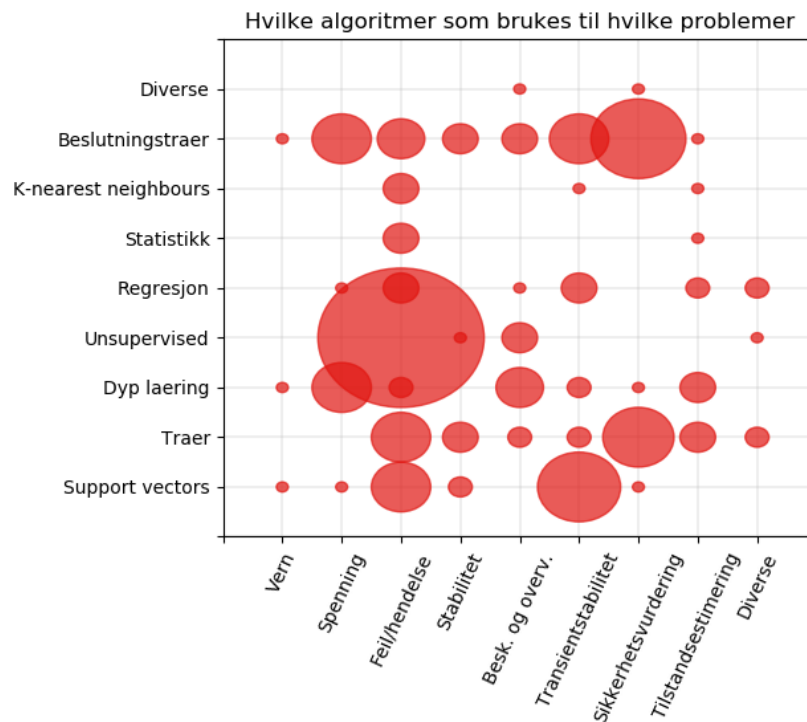


Figur 5.2: Histogram over hvilken type institusjon som har gitt ut artikler. Hver type institusjon har fått ett poeng for hver artikkel den har bidratt til.

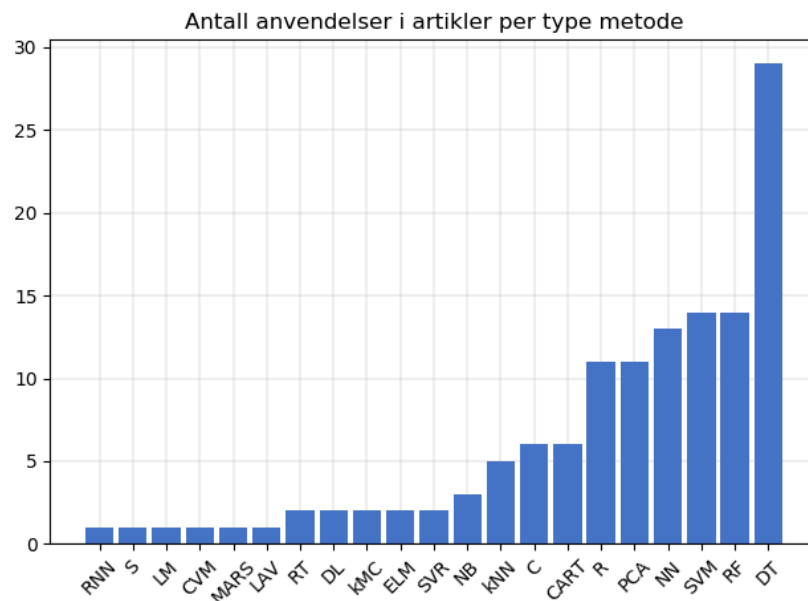
Videre er innholdet i artiklene undersøkt nærmere gjennom å se på hvilken metode som er brukt opp mot hvilket problem som er løst, dette er vist i figur 5.4. Her ser en at kombinasjonen unsupervised maskinlæring og feil/hendelser er den som er blitt undersøkt definitivt flest ganger. Et annet tydelig funn er at blant maskinlæringsalgoritmer er beslutningstrær meget populært, som også fremgår i figur 5.5. Blant problemene er feil/hendelseproblemene tydelig de vanligste å undersøke, som også fremgår i figur 5.6. Utover disse to kategoriene er det også noen andre vanlige kombinasjoner: Transientstabilitet og support vectors, sikkerhetsvurdering og trær og spenning og dyp læring. Dette stemmer godt med figur 5.6 og 5.5, hvor disse problemene og algoritmene er blant de vanligste i litteraturgrunnet.



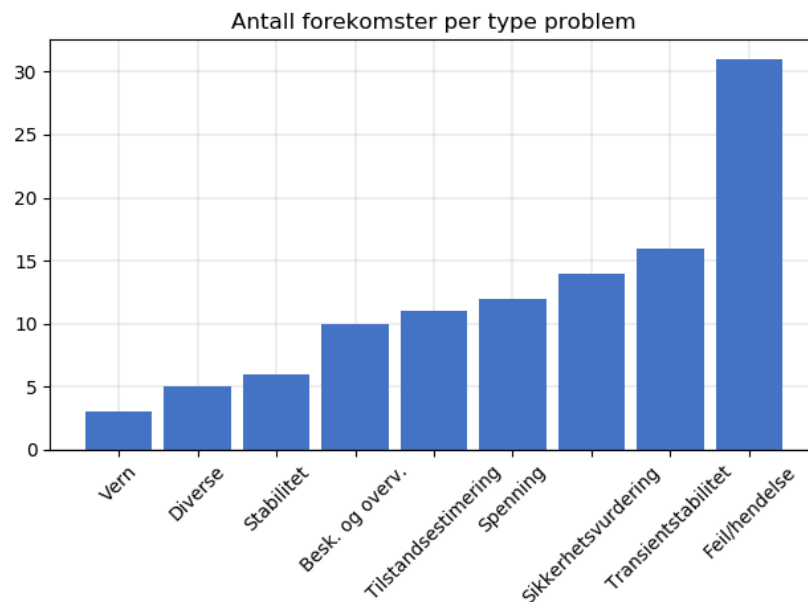
Figur 5.3: Histogram over hvilket land institusjonen som har gitt ut artikler befinner seg i. Hvert land har fått et poeng for hver artikkel en institusjon der har bidratt til.



Figur 5.4: Oversikt over hvilke algoritmer som brukes på hvilke problemer. Sirkelens senter viser hvilken problem-algoritme kombinasjon den viser til og radiusen viser hvor mange tilfeller det var av denne kombinasjonen. En rutes bredde/høyde tilsvarer 10 tilfeller.



Figur 5.5: Histogram over de ulike metodene/algorithmene som brukes med PMU-data. Forkortelser er beskrevet i tabell 5.1



Figur 5.6: Histogram over de ulike gruppene av problemer som forsøkes løst med PMU-data og maskinlæring. Gruppene er beskrevet i tabell 5.2

Det er i utgangspunktet tre kombinasjoner som er klart mest vanlige (sikkerhetsvurdering og trær slås sammen med sikkerhetsvurdering og beslutningstrær da de er relativt like), med 6 artikler eller flere som dekker dem. Dette er:

1. Feil/hendelse - unsupervised maskinlæring
2. Sikkerhetsvurdering - beslutningstrær og andre trær
3. Transientstabilitet - support vectors

I tillegg til dette er den nevnte kombinasjonen spenning og dyp læring interessant både for å se bredde, og siden den er den tredje største kombinasjonen utenfor de to vanligste kategoriene. Til slutt er også en artikkel hvor Norge, ved Statnett SF, har bidratt, valgt ut for å se på hva som gjøres her. Derfor er også følgende to kombinasjoner valgt for videre undersøkelse i fordypningsdelen:

4. Spenning - dyp læring
5. Spenning - regresjon (norsk artikkel)

Innefor de 5 utvalgte kombinasjonene vil det være fokus på de vanligste problemene/ algoritmene innenfor de forskjellige gruppene når fordypningsartikler velges, for eksempel spenningsstabilitet innenfor spenning og SVM innenfor support vectors.

Med tanke på området som skal vurderes til videre forskning vil de 4 vanligste problemene å løse velges til videre vurdering og diskusjon; feil/hendelse, transientstabilitet, sikkerhetsvurdering og spenning. Dette fordi de og omfatter de 5 utpekte områdene til fordypningsdelen. Hvilke av disse fire som vil anbefales for videre forskning og utvikling vil diskuteres under seksjon 5.3.3.

Tabell 5.1: Oversikt over forkortelser og grupperinger innenfor maskinl ring

Forkortelse	Betydning	Gruppe
DT	Beslutningstr�r (Decision trees)	Beslutningstr�r
kNN	K-nearest neighbour	K-nearest neighbour
SVM	Support vector machine	Support vectors
SVR	Support vector regression	Support vectors
CVM	Core vector machine	Support vectors
CART	Classification and regression trees	Tr�r
RF	Random forest	Tr�r
RT	Regression trees	Tr�r
DL	Dyp l�ring	Dyp l�ring
NN	Nevrale nett	Dyp l�ring
RNN	Recurrent neural networks	Dyp l�ring
ELM	Extreme learning machine	Dyp l�ring
C	Clustering	Unsupervised
PCA	Prinsipalkomponentanalyse (Principal component analysis)	Unsupervised
kMC	k-means clustering	Unsupervised
R	Regresjon	Regresjon
MARS	Multivariat adaptive regression splines	Regresjon
LAV	Least absolute value	Statistikk
NB	Naive bayes	Statistikk
S	Supervised	Diverse
LM	Line�r modell	Diverse

Tabell 5.2: Oversikt over gruppering av ulike problemer som løses ved hjelp av maskinlæring og PMU-data.

Problem	Gruppe
Sikkerhetsvurdering	Sikkerhetsvurdering
Tilstandsestimering	Tilstandsestimering
Transientstabilitet	Transientstabilitet
Feillokalisering	Feil/hendelse
Feildeteksjon	Feil/hendelse
Feilanalyse	Feil/hendelse
Hendelsesklassifisering	Feil/hendelse
Øydriftdeteksjon	Feil/hendelse
Angrepsdeteksjon	Feil/hendelse
Katastrofeindikatorer	Feil/hendelse
Spenningskontroll	Spenning
Spenningsstabilitet	Spenning
Spenningsikkerhet	Spenning
Differensialvern	Vern
Distansevern	Vern
Dynamisk signatur	Stabilitet
Stabilitetsvurdering	Stabilitet
Wide area monitoring	Beskyttelse & overvåkning (Besk. og overv.)
Beskyttelsesskjema	Beskyttelse & overvåkning (Besk. og overv.)
Operasjonsstøtte	Beskyttelse & overvåkning (Besk. og overv.)
Kraftsystemsvingninger	Beskyttelse & overvåkning (Besk. og overv.)
Control scheme	Beskyttelse & overvåkning (Besk. og overv.)
Damping	Diverse
Visualisering	Diverse

5.2 Fordypning

I denne delen er det blitt valgt ut en artikkel på hver av de fem områdene pekt ut i forrige del. Først ble kombinasjonen blant de underliggende metodene/problemene av eventuelle grupper som hadde flest artikler valgt ut. Deretter ble det valgt ut en artikkel blant disse på bakgrunn av følgende vurderinger:

- Hvor mange artikler forfatteren og/eller institusjonene bak den artikkelen hadde bidratt til innenfor fagområdet.
- Hvor ny artikkelen var.
- Hvor god artikkelen i seg selv var, hovedsakelig basert på presentasjon av resultater og tilhørende statistisk analyse.
- Om institusjoner utover universiteter og forskingsinstitutt hadde bidratt.

I tabell 5.3 vises hvilke konkrete kombinasjoner som ble valgt innenfor her av underområdene og hvilken begrunnelse som ligger til grunne for utvelgelsen av hver enkelt artikkel. Videre presenteres nå den utvalgte artikkelen fra hvert av de utvalgte områdene. I neste seksjon vil disse diskuteres.

5.2.1 Feil/hendelse - unsupervised maskinlæring

Her ble artikkelen [35] valgt: “Real-time event detection and feature extraction using PMU measurement data“. I denne blir en metode som bruker PCA til å detektere hendelser (unormal oppførsel) i kraftnettet, samt blir blant annet lokasjonen av hendelsen og typen hendelse, presentert. Det blir også utviklet visualisering for å vise dette. Til PCA brukes vinkel, spenning og frekvens hentet fra PMU’er. Ut ifra dette beregnes egenverdier og egenvektorer med PCA, som videre brukes til å indikere om det er stabil drift, signifikante hendelser eller ekstreme hendelser i nettet. Det ble funnet at med mange PMU’er (over 1000 i USA) var dette et beregningsmessig tungt og trengt problem. Ved å bruke “partitional PCA“ (PPCA), som blant annet bruker clustering for å effektivisere, ble beregningene betydelig raskere (fra nesten 1 minutt til under 1 sekund). De klarte primært å lokalisere hendelser, og ga noen indikasjoner til hvordan en kunne analysere plottene deres for å bestemme type, men det ble

Tabell 5.3: Oversikt over hvilke konkrete kombinasjoner innenfor underområdene som ble valgt og hvilken begrunnelse som ligger til grunne for utvelgelse av artikkelen innenfor hver av kombinasjonene.

Underområde	<i>Feil/hendelse - unsupervised maskinlæring</i>	<i>Sikkerhetsvurdering - beslutningstrær</i>	<i>Transientstabilitet - support vectors</i>	<i>Spenning - dyp læring</i>	<i>Spenning - regresjon</i>
Konkret kombinasjon	Feillokalisering - PCA	Sikkerhetsvurdering - Beslutningstrær	Transientstabilitet - Support vector machine	Spenningsstabilitet - Nevrale nett	Spenningsstabilitet - Regresjon
Utvelgelsesbegrunnelse for artikkel	Fra USA, ny (2015), andre artikler innenfor gruppen enten skrevet på tynt grunnlag/dårlig skrevet eller fra land med færre totale bidrag.	En med random forest valgt, hvor det også ble sammenlignet med beslutningstrær. Har kraftselskap blant bidragsyttere, en av de mest frekvente forfatterne, og utgivelsesinstitusjonen hadde flere (4) bidrag. Noe ny artikkel (2010)	Hadde kraftselskap blant bidragsytterne, flere av forfatterne har bidratt til 2 artikler.	Har tydelige gode resultater, forfatterne har flere publikasjoner, få andre gode alternativer i gruppen.	Valgt da Norge er blant bidragsytterne, og modeller ble testet på det norske nettet. Gir et blick på hva som gjøres innenfor fagområdet i Norge.

ikke eksplisitt vist hvordan dette gjøres på en systematisk måte. Det tolkes at deres resultat virker hovedsakelig som beslutningsstøtte/analysestøtte i den formen som ble presentert.

5.2.2 Sikkerhetsvurdering - beslutningstrær og trær

Her ble artikkelen [28] valgt: “Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame“. Her ble random forest brukt med blant annet frekvens og vinkel hentet fra PMU til å gjøre en sikkerhetsvurdering gjennom å klassifisere om et system er stabilt eller ikke. Sikkerhetsvurdering defineres i artikkelen som hvor godt et system står imot all slags mulige feil og forstyrrelser. Innenfor dette finner en både transientstabilitet og spenningsstabilitet, som også ses på i andre artikler som er blitt valgt ut her. I gjeldende artikkel ble først et enkelt beslutningstre testet, med en resulterende nøyaktighet på ca. 80 %. Dette er alt for lavt i kraftsystem-sammenheng, og random forest ble derfor forsøkt videre. Da ble en pålitelighet på 99,9 % (3 niere) presentert, som er den samme påliteligheten som moderne kraftsystem er designet for og opereres etter. Under arbeidet ble det sett på hvilke variabler som hadde mye innvirkning/høy korrelasjon inn mot sikkerhetsvurderingen. Videre fant de også at ustabile tilfeller i datasettet deres (andel originalt 24 %) måtte økes ved hjelp av replisering for å kunne klassifisere også ustabile til-

feller godt. Til slutt poengteres det at denne typen modell er rask å trene, og kan oppdateres årlig/månedlig/ukentlig med prognoser.

5.2.3 Transientstabilitet - support vectors

Her ble artikkelen [9] valgt: “Support Vector Machine-Based Algorithm for Post-Fault Transient Stability Status Prediction Using Synchronized Measurements“. Her ble flere ulike variabler undersøkt for om en kunne forutse transientustabilitet ved å bruke dem som input i en support vector machine. De undersøkte variablene var spenning, rotorvinkler og generatorhastighet. Det ble funnet at rotorvinkler ga dårligst resultater, mens generatorhastighet ga noe dårligere resultater enn spenning, som var best. Spenningsendringer kommer også raskere etter feil og kan dermed gi raskere varsel om eventuell transientustabilitet. Av disse grunnene ble spenning valgt for videre analyse, som ga meget gode resultater. Ved å undersøke flere ulike typer feil presenterte de resultater mellom 97,73 % og 100 %, hvor de beste resultatene var når stabil tilstand skulle klassifiseres. Ved å endre topologien uten å la algoritmen trene opp en ny modell gikk nøyaktigheten noe ned (lavest 96 %), og ytterligere noe ned når realistiske usikkerheter i målingene ble lagt til (95,825 %). Det ble også sett på hvordan andel induksjonsmaskiner i lasten i nettet ville påvirke SVM-modellens nøyaktighet. Det ble funnet at nøyaktigheten ble høyere jo større andelen var, og nøyaktigheten var 100 % fra 20-100 % induksjonsmaskiner. Også tidsbruk ble undersøkt, hvor algoritmen brukte 0,067 ms, som gir god tid til måling, prosessforsinkelser og kommunikasjon for å håndtere en eventuell ustabilitet (bør håndteres på $< 1s$). Til slutt ble modellen testet på Venezuelas nett, med 100 % nøyaktighet. Per 2010 skulle modellen testet for fremtidig implementasjon i Venezuelas nett.

5.2.4 Spenning - dyp læring

Her ble artikkelen [17] valgt: “On-line Voltage Stability Load Index Estimation Based on PMU Measurements“. Målet var å estimere “Voltage stability load bus index“(VSLI) ved hjelp av et type nevralt nett, “Echo State Network“. Det nevrale nettet brukte spenning og vinkel målt ved hjelp av PMU, og estimerte da VSLI. VSLI bergenes vanligvis fra en formel utledet fra thevenin-ekvivalenten til systemet fra den aktuelle samleskinnen. Det var verdien regnet ut med denne formelen som var sammenligningsgrunnlaget for det nevrale nettet. De presenterte gode resultater med høy nøyaktighet både under normale forhold og med feil/forstyrrelser.

5.2.5 Spenning - regresjon

Her ble artikkelen [15] valgt: “PMU-based voltage instability detection through linear regression“. I denne artikkelen ble det brukt lineær regresjon på spenning- og strømmålinger fra PMU for å finne følsomheter for ustabil spenning. Målet er å kunne oppdage langtidskollaps av kraftsystemet og dermed kunne sette inn grep for å unngå det. Ved hjelp av “rullende vindu“ ser en på en og en tidsperiode med data, og flytter dette “vinduet“ framover og ser på hvordan spenningen varierer i forhold til effekt ved hjelp av lineær regresjon. Dette testes først på et enkelt system med jevnt lastøkning, både med og uten prosesstøy, og det vises at modellen her klarer å fange opp ustabilitet i spenning både som følge av høy last og som følge av systemkomponenter som OLTC (On-Load Tap Changer) og OEL (Over Excitation Limiter). Det ses også på en reell hendelse i det norske kraftsystemet, hvor en ved å bruke reelle PMU-målinger ser at også da klarer modellen å fange opp ustabiliteten i spenningen. I det reelle eksempelet ble alvorlig ustabilitet unngått ved å koble ut last, og det foreslås at en kan ha en følsomhetsgrense som varsler når last burde kobles ut. Fordelen med denne fremgangsmåten opp mot tidligere tilnærminger til samme problem er at man her er uavhengig av både topologi og dynamisk modell av kraftsystemet.

5.3 Diskusjon av litteratur

Etter å både ha sett på litteraturen samlet og på utvalgte artikler, vil nå selve litteratursøket, fordypningsartiklene og vurdering av videre forskning og utvikling diskuteres.

5.3.1 Litteratursøk

Den største utfordringen ved litteratursøket var at det var mange artikler på dette feltet, og derfor ble endel begrensninger satt, slik beskrevet i innledningen, seksjon 1.3. Dette gir flere usikkerhetsmomenter rundt om det endelige utvalget artikler som ble funnet er et representativt utvalg gitt de begrensningene som ble satt. Disse kan deles i to:

- Maskinlæring kan ha blitt brukt på PMU uten at dette har blitt sagt eksplisitt i abstract i en eller begge av disse formene:
 - Algoritmen/metoden navngis ikke
 - Det sies ikke eksplisitt at data blir hentet fra PMU

Ved disse hendelsene vil ikke søkemotoren kunne plukke opp alle artikler som faktisk er maskinlæring og PMU sammen. Det var også tilfeller hvor en artikkel kom opp i søket, men hvor et av disse hendelsene var tilfellet, og artikkelen ble ikke tatt med videre grunnet overflod av artikler.

- En begrensning var å ekskludere maskinlæring som driver datarensing, detektering av bad/ missing data, feature selection eller lignende på PMU-data. Det kan ha blitt vurdert til at dette var tilfellet, hvor en algoritme i realiteten både gjorde dette og samtidig bygde en modell basert på dataene. Dermed kan aktuelle artikler ha blitt ekskludert. Dette kan ha gitt kunstig få artikler med PCA som metode, som ofte var brukt til data-prosesser og ble ekskludert.

Søket ble avsluttet grunnet tidsbruk og at tilstrekkelig antall artikler var blitt funnet. Dermed var det kombinasjoner av søkeord, søkemotorer og andre kilder som ikke ble undersøkt. Dette kunne gitt andre artikler, og et annet utvalg. De søkeordene som ble brukt ble vurdert til å være de mest relevante og dette de vanligste algoritmene/metodene innenfor maskinlæring. Derfor vurderes det at utvalget av artikler som ble funnet er relativt representativt. Til

slutt har det store antallet artikler ført til å funnene i denne oppgaven i større grad oversikt over fagområdet enn en dybdeanalyse av hva fagområdet inneholder. Dette fordi det ble for mange artikler til å kunne sette seg grundig inn i hver, og det ble derfor valgt å fokusere på oversikten og kun se grundigere på noen utvalgte fordypningsartikler.

5.3.2 Fordypningsartikler

Alle de valgte artiklene har fått gode resultater, enten vist ved nøyaktighetsmål eller vist visuelt. Noen artikler presenterte hele 100 % nøyaktighet, og ingen dårligere enn 95,8 %. Den laveste var i samme artikkel som et tilfelle av høyeste nøyaktighet, som viser at den samme modellen under andre omstendigheter presterer meget bra. Det er dog vanskelig å sammenligne de ulike artiklenes resultat da noen viser resultatet visuelt og andre angir nøyaktighet. Det er også betydelig variasjon i hvor godt det testes og hvor mange instanser som er brukt til å teste den aktuelle metoden. Dette gjør at noen resultater er mer overbevisende enn andre, hvor kun visualisering viser at metoden eksempelvis klarer å fange opp en hendelse. Alle artiklenes resultat vurderes dog helhetlig til å være gode og viser at de valgte algoritmene og metodene fungerer godt til å løse ulike problemer i kraftnettet.

Flere av de undersøkte artiklene ser ikke på tidsbruk av algoritmene. Tid kan ofte være kritisk, for eksempel i sammenheng med transienter, hvor transientustabilitet krever korrigerende tiltak innen meget kort tid, $< 1s$ [9]. En bør derfor måle, estimere og sende ut signal om eventuell handling på maksimalt 0.5 s. Også innenfor sikkerhetsvurdering har en dette tidspresset, da det ble definert at transientstabilitet er en del av problemet. Likevel er ikke tidsbruk blitt sett på i artikkelen. Kun i 2 av de undersøkte artiklene ble tidsbruken sett på, og kun satt i sammenheng med reell drift i en av dem. Dette tyder på at de utviklede løsningene fortsatt er på et forskningsstadium, hvor implementering og hensyn til reelle tilstander ikke alltid blir inkludert i analysen av metodens prestasjon. Det er også en mulighet at tid ikke er en kritisk variabel i alle problemer, men det kan likevel argumenteres for at det alltid bør være et hensyn å ta for å raskest mulig kunne oppdage feil/forstyrrelser i nettet.

Transientstabilitet er flere ganger en del av et større problem som løses, i tillegg til å løses i seg selv. Det er en definert del av sikkerhetsvurdering og det brukes transientstabilitetsdata i for å detektere hendelser med PCA. Dette vitner om at transientstabilitet er et tema som naturlig knyttes til PMU. Dette er også tydelig fra figur 5.6, hvor transientstabilitet er den gruppen med problemer som er løst nest flest ganger. Her har PMU et klart fortrinn med høy frekvens på målinger og synkroniserte målinger for å kunne fange opp forandringer. Også

innenfor den største gruppen - feil/hendelse - har vi sett et eksempel på at transientstabilitet virker. PMU gir altså data som er meget nyttige for å se på transienter, noe som også har vært tydelig i generell litteratur om PMU. Dette har gjort at det flere ganger var et poeng å kutte ned beregningstiden, som også vil være førende for hvilke algoritmer og hvordan de skal brukes når en senere skal se på transientstabilitet ved hjelp av maskinlæring.

5.3.3 Videre forskning og utvikling

I det nordiske markedet dras frekvenskvalitet frem som en fremtidig utfordring [31], grunnet blant annet raskere, oftere og økt endring i produksjon. Mer uforutsigbar (fornybar) kraft er også vanskeligere å forutsi, og det er ventet mer feil i prognoser. Dette gjør at å opprettholde god frekvenskvalitet blir mer utfordrende. Av de nordiske TSO'ene blir mer avanserte systemer for overvåkning og kontroll pekt ut som en del av løsningen på denne utfordringen. I litteraturen har det blitt funnet at PMU og maskinlæring kan være en potensiell del av denne løsningen, da frekvens blant annet er viktig inn mot sikkerhetsvurdering. Når fremtidig forskning og utvikling innen maskinlæring brukt på PMU skal diskuteres videre, vil det vurderes hvorvidt de ulike områdene vil kunne støtte opp om bedre frekvenskvalitet.

Tidligere i dette kapittelet ble følge fire kombinasjoner valgt ut som potensielle områder til videre forskning og utvikling:

1. Feil/hendelse
2. Sikkerhetsvurdering
3. Transientstabilitet
4. Spenning

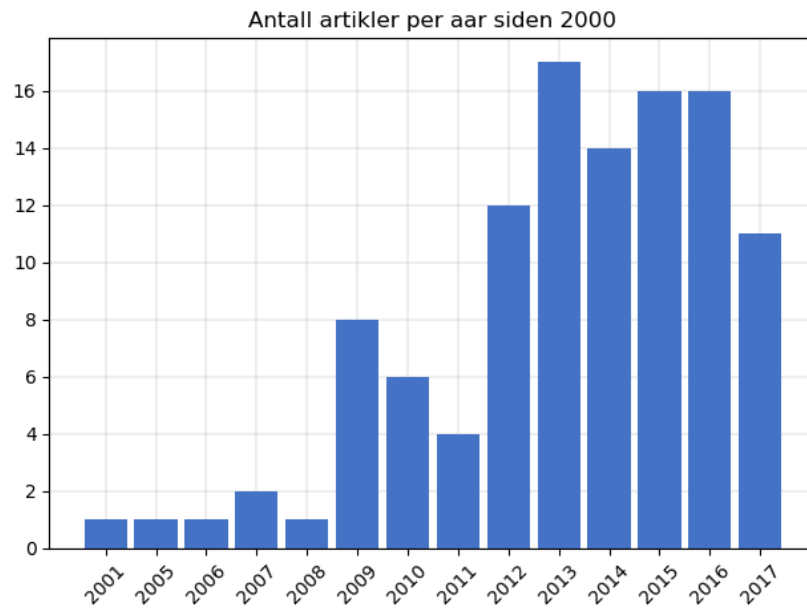
Disse ble valgt fordi de dekker de kombinasjonsområdene med flest artikler innenfor seg, samtidig som problemene de løser er de fire vanligste problemene å se på. For å få mer innsikt i hva som har mest potensiale for det nordiske kraftnettet har flere fagfolk på området, fra både NTNU og Statnett SF, blitt kontaktet. Følgende diskusjon vil være basert på og inspirert av møter og korrespondanse med disse [10][5].

Deteksjon og/eller lokalisering av feil og hendelser er et område som potensielt kan forbedres ved hjelp av PMU. Dersom PMU-enheten i seg selv kan bli oppgradert slik at den har

programvare i seg som kan detektere feil vil dette kunne gjøres lokalt. Dette vil kunne være raskere enn å først samle data i et prosesseringspunkt. Denne programvaren kan da kunne være maskinlæringsbasert. Videre er det også naturlig med løsninger som ser på systemet og finner feil/hendelser og eventuelt da også lokaliserer de, slik det ble gjort i fordypningsartikkelen på feltet. Om dette kan bli implementert vil det kunne brukes i drift og varsle om feil, potensielt også type feil. På sikt kan det også tenkes å utvikle systemer som kan foreslå hva som skal gjøres for å rette opp feilen, og eventuelt også automatisk sette i gang enkle tiltak. Bedre oppfangning av feil og/eller hendelser kan føre til færre feil som gir konsekvenser for drift av kraftnettet. Dette kan i sin tur gi bedre forsyningssikkerhet, og potensielt også bedre spennings- og frekvenskvalitet. Dette vil igjen føre til økonomiske besparelser. Om en ser på kun problemet “deteksjon og/eller lokalisering av feil og hendelser“ uavhengig av hvilken maskinlæringsalgoritme som er brukt er dette tydelig mest forsket på. Dette tyder på at dette er et problem som er interessant å løse, men som ikke har noen åpenbar algoritme som passer det, men hovedsakelig blir klassifisering og unsupervised maskinlæring anvendt. Det er likefremt forsket mye på området, som kan tyde på at det er modent for å videreutvikles til neste TRL-fase.

Kombinasjon 2-4 vil diskuteres samlet som muligheter for å bedre estimere og forutsi stabilitet og sikkerhet i kraftnettet. Innenfor område 4 fokuseres det altså på spenningsstabilitet da dette er det vanligste problemet innenfor gruppen. I dag drives nettet etter et N-1-prinsipp, som gjør at nettet drives med gode marginer i forhold til sine fysiske begrensninger. Der som en hadde gått over til probabilistisk drift ville en få mye bedre utnyttelse av nettet og kunne unngå utbygging og kanskje heller brukt midler på for eksempel vedlikehold og utvikling av smartere løsninger. Ved probabilistisk drift er det interessant å vite hvor nær nettet er ustabilitet, eventuelt hvor sannsynlig det er, for å kunne drifte nettet etter probabilistiske prinsipper. Også systemsikkerhet vil en kunne bruke til å vurdere hvorvidt systemet vil kunne takle feil og forstyrrelser. Her vil PMU og maskinlæring kunne bidra med bedre algoritmer og modeller til å estimere og forutsi stabilitet og sikkerhet, slik litteraturstudien har vist eksempler på. Om dette blir implementert kan en se for seg at en for eksempel varsler når systemet nærmer seg ustabilitet, og også kan peke på hva som er kilden til ustabilitet. Også med N-1-drift kan bedre oversikt over både stabilitet og sikkerhet kunne bidra positivt. I begge tilfeller vil det kunne bidra til færre feil, bedre forsyningssikkerhet, frekvenskvalitet og spenningskvalitet. Forfatteren sitter dog med et inntrykk av, etter møte med fagpersoner, at gevinsten vil være større ved probabilistisk drift.

Stabilitet og sikkerhet er kanskje området som kan ha potensiale til å gi størst gevinster, men vil også være omfattende å implementere. De store gevinstene vil først komme når PMU



Figur 5.7: Histogram over hvor mange artikler som ble utgitt hvert år siden 2000.

innlemmes i flere systemer og funksjoner, og helst som støtte til probabilistisk drift. Dette vil kreve mye, og vil ta tid i en konservativ bransje. Da kan det tenkes at det vil være mer realistisk å implementere lokasjon/deteksjon av feil, som kan fungere som en varslers for netteiere. Det vurderes også at alle områdene vil kunne støtte opp om bedre frekvenskvalitet, gjennom å for eksempel gi bedre overvåkning og oppfangning av feil.

Etter de mest lovende områdene nå er blitt diskutert vil også et område med lite forskning trekkes fram. Etter intervjuet med Statnett [10] satt forfatteren igjen med et inntrykk av at dempingskontroll er et interessant problem å se på med PMU. Dette ble det funnet få artikler på, 4 totalt, noe som tyder på at maskinlæring så langt har blitt lite brukt til å se på det. Da dette virker som noe det er interesse for kan det være interessant å forske mer på.

Til slutt kan det synes som det er generelt utfordrende å heve dette fagområdet til implementering, da det er gjort mye forskning på området, men forstøtt er i en relativt lav TRL-fase. Det er mulig at noen underområder i realiteten er i en høyere TRL-fase enn andre, noe en vil kunne avdekke om de områdene med mest forskning blir undersøkt videre. Om en ser på utviklingen av artikler siden 2000, vist i figur 5.7 er det først de siste 6 årene at det har vært mest forskning publisert. Det er altså store mengder forskning på området, men ikke over så lang tid. Det kan være noe av forklaringen på den lave TRL-fasen. Videre forskning burde nå, i tillegg til å ha som mål å utvikle fagområdet videre, også se på hvordan det kan heves til et høyere TRL-nivå og utvikles mot mulig implementering.

5.4 Oppsummering av funn

Etter å ha gjennomgått litteratur på fagområdet maskinlæring på PMU-data har det blitt identifisert en rekke trekk som oppsummeres her.

- Fagområdet er i stor grad fortsatt på forskningsstadiet (TRL-fase 2), og har begynt utviklingen mot implementering.
- Noe geografisk bredde finnes, USA er klart størst, Kina, Canada, Storbritannia og India er andre betydelige bidragsytere.
- Unsupervised og feil/hendelse er klart vanligst gruppekombinasjon. Beslutningstrær er klart vanligste metode og feil/hendelse er vanligste gruppe.
- De utvalgte artiklene presenterte alle gode resultater, men det var variasjon i hvor godt disse ble testet og om det kun var visualisert eller vist gjennom parametere som for eksempel nøyaktighet. Det var varierende grad av fokus på metodens tidsbruk, men mer hvor transientstabilitet var innblandet. Transientstabilitet ble funnet som et vanlig problem å se på med maskinlæring på PMU-data.
- Det har blitt poengtert at grunnet overflod av artikler og begrenset tid kan interessante og relevante artikler ha blitt oversett. Det ble Like vel vurdert dithen at det ble funnet et utvalgt artikler som er representativt for fagområdet.
- Stabilitet og sikkerhet vurderes å ha godt potensiale, men også til å være krevende å implementere. Det vil antagelig gi best gevinst om en går over til probabilistisk drift. Da pekes lokalisering/deteksjon av feil/hendelser seg ut som det mest lovende området. Begge har potensiale for mindre feil, bedre forsyningssikkerhet, økonomiske besparelser og bedre frekvens- og spenningskvalitet. Anbefales også å forske mer på dempingskontroll, da det er gjort lite på området, men virker å være en sentral utfordring. Avslutningsvis anbefales og et fokus på å se på hvordan fagområdet kan heves til en høyere TRL-fase.

Kapittel 6

Konklusjon

I denne masteroppgaven har det blitt utført en litteraturstudie for å undersøke hva som preger fagområdet "maskinlæring på PMU-data". Det ble først gitt en innføring i både PMU og maskinlæring, før funnene fra litteraturstudien ble presentert. Funnene innebærer fagområdets modenhet(TRL-fase), geografiske spredning, hvor mye ulike kombinasjoner av problemer og maskinlæringsalgoritmer er forsket på i forhold til hverandre, og en presentasjon av artikler fra de mest sentrale underområdene. Alt i alt har dette gitt en oversikt over forskning på hvordan maskinlæring kan brukes på PMU-data som ikke har blitt funnet i tidligere litteratur.

Det ble funnet at fagområdets modenhet er i TRL-fase 2, og i stor grad fortsatt er på forskningsstadiet, med tung overvekt av universiteter som står bak forskningen. Geografisk er det noe bredde, hvor USA står for klart mest forskning på fagområdet. Det ble funnet 4 underområder som blir forsket mest på og var interessante å se videre på; feil/hendelse - unsupervised maskinlæring, sikkerhetvurdering - beslutningstrær og andre trær, transientstabilitet - support vectors, spenning - dyp læring. Beslutningstrær ble funnet til å være den maskinlæringsmetoden som ble mest brukt, og feil/hendelse var den typen problemer som ble løst mest hyppig.

Fra hvert av de 4 vanligste underområdene ble det valgt ut én artikkel innenfor den kombinasjonen av problem og metode som var vanligst i det aktuelle underområdet. I tillegg ble en artikkel hvor Norge, ved Statnett SF, hadde bidratt også valgt ut. Alle de utvalgte artiklene presenterte gode resultater. Det var varierende grad av fokus på tidsbruk som en viktig faktor i resultatene. Det ble også funnet at transientstabilitet er et vanlig problemet å se på ved hjelp av maskinlæring på PMU-data.

6.1 Forslag til videre forskning

Basert på seksjon 5.3.3 er det flere forskningsområder som peker seg fram. Det som vurderes til å være nærmest å kunne utvikle mot implementering er deteksjon og/eller lokalisering av feil og hendelser, ved hjelp av klassifisering eller unsupervised maskinlæring. Stabilitet vurderes også som lovende, men da i et større bilde hvor også probabilistisk drift utvikles parallelt. Videre anbefales det også forskning på å bruke maskinlæring og PMU-data til dempingskontroll, da dette virker interessant for Statnett SF, men har blitt lite forsket på. Generelt sett foreslås også et økt fokus på hvordan fagområdet kan heves til høyere TRL-fase og utvikles næremere implementering. Dette kan både være i form av å inkludere et slikt fokus i videre forskning, og i form av egen forskning dedikert til denne problemstillingen.

Om det nå settes i gang nye forskningsprosjekter innenfor maskinlæring brukt på PMU-data kan det være interessant å samarbeide med institusjoner fra land hvor det har blitt forsket mer på dette. Da dras USA, Canada og Storbritannia fram som mulige land å søke samarbeid i. Virginia Tech (USA), Hydro-Québec (Canada) og Queen's University Belfast (Storbritannia) er de institusjonene som har bidratt til flest artikler i sine land, og kan være gode kandidater å søke samarbeid med.

6.2 Egenrefleksjon

Da denne oppgavene ble valgt forventet forfatteren å lære mye om PMU og dens anvendelser. Denne forventningen har blitt innfridd, men mengden av forskning gjort på hvordan maskinlæring brukes på PMU har overrasket. Dette har gjort at læringsutbyttet har vært på et mer oversiktlig plan enn det var forventet på forhånd. Forfatteren sitter igjen med et ønske om å ha lært mer om PMU i løpet av studiene på NMBU, da dette synes å være en lovende sensor som er interessant å ha kjennskap til. Derfor trekkes særlig bakgrunnskapittelet om PMU, kapittel 3, frem som stoff forfatteren kan se for seg at inkluderes i fag om kraftsystemet, eksempelvis FYS377. Også det følgende kapittelet, kapittel 4, kan være nyttig for å få oversikt over en gruppe metoder som blir stadig mer aktuell for bruk til kraftsystemapplikasjoner.

Dette er første masteroppgave gjort på dette temaet og den har gitt et oversiktsbilde over fagområdet som ikke er funnet i tidligere litteratur. Forfatteren sitter igjen med en tanke om at det ville vært interessant å se videre på noen av de underområdene som har blitt oppdaget. Dette vil kunne være utgangspunkt for fremtidige masteroppgaver, hvor forfatteren ser

for seg at en kan se på hvordan PMU brukes inn mot å lokalisere og/eller detektere feil og hendelser, sikkerhetsvurdering, transientstabilitet og/eller spenningsstabilitet.

Bibliografi

- [1] Alexandra von Meier, A. M. (2012). Micro-synchrophasors: a promising new measurement technology for the ac grid. In *i4Energy Seminar*.
- [2] Boulesteix, A.-L., Janitza, S., Kruppa, J., and König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):493–507.
- [3] Chakrabarti, S., Kyriakides, E., Bi, T., Cai, D., and Terzija, V. (2009). Measurements get together. *IEEE Power and Energy Magazine*, 7(1).
- [4] Dietrichson, F. S., Friis, J. W., Hole, J., Krossholm, K. K., Måseidvåg, S., and Steen, H. L. (2017). Maksinlæring - the times they are a-changin'n.
- [5] Duong, D. T. (2018). Mail-korrespondanse.
- [6] eia (2012). New technology can improve electric power system efficiency and reliability. Hentet 12. februar 2018 fra <https://www.eia.gov/todayinenergy/detail.php?id=5630>.
- [7] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.
- [8] Glorennec, P. Y. (2000). Reinforcement learning: An overview. In *Proceedings European Symposium on Intelligent Techniques (ESIT-00), Aachen, Germany*, pages 14–15. Citeseer.
- [9] Gomez, F. R., Rajapakse, A. D., Annakkage, U. D., and Fernando, I. T. (2011). Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. *IEEE Transactions on Power Systems*, 26(3):1474–1483.
- [10] Jacobsen, M.-R. and Schaug-Pettersen, J. (2018). Møte om pmu og maskinlæring.
- [11] Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065):20150202.

- [12] Kaushik, S. (2016). An introduction to clustering and different methods of clustering. Hentet 19.14.18 fra: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>.
- [13] Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4):261–283.
- [14] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.
- [15] Leelaruji, R., Vanfretti, L., Gjerde, J., and Lovlund, S. (2013). Pmu-based voltage instability detection through linear regression. In *Power and Energy Engineering Conference (APPEEC), 2013 IEEE PES Asia-Pacific*, pages 1–6. IEEE.
- [16] Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):14–23. Hentet 26. februar 2018 fra <http://www.stat.wisc.edu/loh/treeprogs/guide/wires11.pdf>.
- [17] Makasa, K. J. and Venayagamoorthy, G. K. (2011). On-line voltage stability load index estimation based on pmu measurements. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–6. IEEE.
- [18] Mohanta, D. K., Murthy, C., and Sinha Roy, D. (2016). A brief review of phasor measurement units as sensors for smart grid. *Electric Power Components and Systems*, 44(4):411–425.
- [19] Moslehi, K. and Kumar, R. (2010). Smart grid-a reliability perspective. In *Innovative Smart Grid Technologies (ISGT), 2010*, pages 1–8. IEEE.
- [20] Ng, A. (2016). Decision trees tutorial. Hentet 22. februar 2018 fra <https://algorithms.wtf/2016/07/27/decision-trees-tutorial/>.
- [21] Norge, E. (2017a). Forsyningssikkerhet. Hentet 16. april 2018 fra <https://energifaktanorge.no/norsk-energiforsyning/forsyningssikkerhet/>.
- [22] Norge, E. (2017b). Kraftproduksjon. Hentet 16. april 2018 fra: <https://energifaktanorge.no/norsk-energiforsyning/kraftforsyningen/>.
- [23] norske leksikon, S. (2009). Faselåst sløyfe. Hentet 5. februar 2018 fra <https://snl.no/fasel>
- [24] Pavella, M., Ernst, D., and Ruiz-Vega, D. (2012). *Transient stability of power systems: a unified approach to assessment and control*. Springer Science & Business Media.

- [25] Phadke, A. and Thorp, J. (2006). History and applications of phasor measurements. In *Power Systems Conference and Exposition, 2006. PSCE'06. 2006 IEEE PES*, pages 331–335. IEEE.
- [26] Phadke, A. G. and Thorp, J. S. (2017). *Synchronized phasor measurements and their applications*, volume 1. Springer.
- [27] Rodriguez, G. (2007). Logit models for binary data. Hentet 16. februar 2018 fra <http://data.princeton.edu/wws509/notes/c3.pdf>.
- [28] Samantaray, S., Kamwa, I., and Joos, G. (2010). Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame. *IET generation, transmission & distribution*, 4(12):1334–1348.
- [29] SF, S. (2018). Fou i statnett. ppt-presentasjon fra FoU i Statnett.
- [30] Srivastava, A. K. (2014). Phasor measurement (estimation) units. *Washington State University, Pullman, WA*.
- [31] Statnett, Fingrid, E. S. k. (2018). Challenges and opportunities for the nordic power system.
- [32] Udemy (2018). Machine learning a-z™: Hands-on python & r in data science. <https://www.udemy.com/machinelearning/>.
- [33] Van der Sluis, L. (2001). *Transients in power systems*. John Wiley & Sons Ltd.
- [34] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37.
- [35] Xu, T. and Overbye, T. (2015). Real-time event detection and feature extraction using pmu measurement data. In *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on*, pages 265–270. IEEE.

Tillegg A

Resultat litteratursøk

I følgende tabell vises litteraturen som utgjør bakgrunnen for litteraturanalsen i oppgaven.

Tittel	År (publisert)	Metode/ algoritme	Problem forsøkt løst	Institusjon	Type institusjon	Land
Using phasor data for visualization and data mining in smart-grid applications	2015	C	V	University of North Dakota	Universitet	USA
Fault localization in Smart Grid using wavelet analysis and unsupervised learning	2013	C	FL	University of Denver	Universitet	USA
Fault Diagnosis Based on Cluster Analysis Theory in Wide Area Backup Protection System	2009	C	FA	North China Electric Power University	Universitet	Kina
Unsupervised Clustering on PMU Data for Event Characterization on Smart Grid	2016	C	EC	Washington State University	Universitet	USA
Comprehensive Clustering of Disturbance Events Recorded by Phasor Measurement Units	2013	C	EC	New Mexico State University	Universitet	USA
Large power network signature analysis with PMU signal-dynamic clustering approach	2015	C	DS	Power Grid Corporation of India	Netteier	India
A Fast Stability Assessment Scheme based on Classification and Regression Tree	2012	CART	StA	Texas A&M University	Universitet	USA
Classification Trees for Complex Synchronphasor Data	2013	CART	PS, EC	Virginia Tech, Clemson university, National Institute of Statistical Sciences	Universitet	USA

Classification and regression tree-based adaptive damping control of inter-area oscillations using wide-area signals	2013	CART	D	North China Electric Power University, Virginia Tech, China Electric Power Research Institute	Universitet, forskningsinstitutt	USA, Kina
Multi-Polytope-Based Adaptive Robust Damping Control in Power Systems Using CART	2014	CART	D	North China Electric Power University, Virginia Tech	Universitet	Kina, USA
Transient stability assessment via decision trees and multivariate adaptive regression splines	2017	CART, MARS	TS	The University of British Columbia, BBA Inc.	Universitet, konsultentselskap	Canada
Power System Transient Stability Assessment Based on Big Data and the Core Vector Machine	2016	CVM	TS	Wuhan University, University of Tennessee	Universitet	Kina, USA
Introducing machine learning for power system operation support	2017	DL	OS	RTE, LRI (Laboratoire de Recherche en Informatique)	TSO, forskningsinstitutt	Frankrike
A deep learning-based cyber-physical strategy to mitigate false data injection attack in smart grids	2016	DL	AD	The University of Akron	Universitet	USA
Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements	2009	DT	VSe	Arizona State University, Electric Power Research Institute, American Electric Power Service	Universitet, forskningsinstitutt, Kraftselskap (produksjon + tranmissionsnett)	USA
Fast on-line voltage security monitoring using synchronized phasor measurements and decision trees	2001	DT	VSe	Virginia Tech	Universitet	USA
On-line dynamic voltage instability prediction based on decision tree supported by a wide-area measurement system	2012	DT	VS	Iran University of Science and Technology	Universitet	Iran
Multi-objective optimization of decision trees for power system voltage security assessment	2016	DT	VS	Cleveland State University	Universitet	USA
Decision Tree Algorithm for Real-Time Identification of Critical Voltage Control Areas	2014	DT	VC	Wroclaw University of Technology	Universitet	Polen
Decision Tree Based Online Identification of Critical Voltage Control Area	2012	DT	VC	Islamic Azad University	Universitet	Iran

Adaptive decision-trees-based regional voltage control	2017	DT	VC	Virginia Tech	Universitet	USA
On-line prediction of transient stability using decision tree method — Sensitivity of accuracy of prediction to different uncertainties	2013	DT	TS	University of Manchester	Universitet	Storbritannia
Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability	2013	DT	TS	University of Manchester	Universitet	Storbritannia
Development of Rule-Based Classifiers for Rapid Stability Assessment of Wide-Area Post-Disturbance Records	2009	DT	StA	McGill University, Hydro-Québec	Universitet, Kraftselskap(offentlig prod+dist)	Canada
Event stream processing for improved situational awareness in the smart grid	2015	DT	SE	Mississippi State University, Pacific Gas and Electric (PG&E) Company	Universitet, privat energiselskap	USA
A Systematic Approach for Dynamic Security Assessment and the Corresponding Preventive Control Scheme Based on Decision Trees	2013	DT	SA, CS	Aalborg University, University of Tennessee, Energinet.dk	Universitet, TSO	Danmark, USA
Methodology for a Security Dependability Adaptive Protection Scheme based on Data Mining	2009	DT	SA	Virginia Tech	Universitet	USA
An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees	2007	DT	SA	Arizona State University, Entergy Services, Inc.	Universitet, Kraftproduksjonsselskap (privat)	USA
Design of a Real-Time Security Assessment Tool for Situational Awareness Enhancement in Modern Power Systems	2009	DT	SA	Arizona State University, Salt River Project	Universitet, kraftselskap	USA
Phasor Measurements-Aided Decision Trees for Power System Security Assessment	2009	DT	SA	Zhejiang University	Universitet	Kina
Online dynamic security assessment with missing pmu measurements: A data mining approach	2013	DT	SA	Arizona State University	Universitet	USA
Wide Area System Islanding Contingency Detecting and Warning Scheme	2014	DT	ID	Dominion Resources, Virginia Tech	Kraftselskap, universitet	USA
Decision Trees Using Synchronized Phasor Measurements for Wide-Area Response-Based Control	2010	DT	FD	Midwest Independent Transmission System Operator, Indiana University-Purdue University Indianapolis	TSO, Universitet	USA

A decision tree-based method for power system fault diagnosis by synchronized Phasor Measurements	2012	DT	FA	Sharif University of Technology	Universitet	Iran
Smart grid line event classification using supervised learning over PMU data streams	2015	DT	EC	Washington State University	Universitet	USA
Data-Mining-Based Intelligent Differential Relaying for Transmission Lines Including UPFC and Wind Farms	2015	DT	DiR	Indian Institute of Technology	Universitet	India
Transient stability prediction based on apparent impedance trajectory recorded by PMUs	2014	DT, CART	TS	Virginia Tech	Universitet	USA
PMU based voltage security assessment of power systems exploiting principal component analysis and decision trees	2015	DT, PCA	VSe	Shiraz University	Universitet	Iran
Data-mining model based adaptive protection scheme to enhance distance relay performance during power swing	2016	DT, RF	PS - DR	Indian Institute of Technology, Robert Bosch Engineering and Business Solutions Ltd	Universitet, konsultentselskap	India
Online Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining	2015	DT, RE, SVM	DS	The University of Manchester	Universitet	Storbritannia
Evaluation of Classification Methods for Online Identification of Power System Dynamic Signature	2014	DT, SVM	DS	The University of Manchester	Universitet	Storbritannia
Extreme Learning Machine Approach for Real Time Voltage Stability Monitoring in a Smart Grid System using Synchronized Phasor Measurements	2016	ELM	VS	Velammal College of Engineering & Technology, Kala-salingam University	Universitet	India
Optimized extreme learning machine for power system transient stability prediction using synchrophasors	2015	ELM	TS	State Grid Liaoning Electric Power Supply Co. Ltd., Northeast Dianli University	Strømproduksjons- og forsyning, universitet	Kina
A Robust Initialization Algorithm for k-Means Clustering in Power Distribution Networks with PMU-based Adaptive Protection System	2014	kMC	PS	University of Texas	Universitet	USA
A novel fast transient stability prediction method based on PMU	2009	kNN	TS	Zhejiang University	Universitet	Kina
Power system state recognition using data mining algorithms	2013	kNN	SE	West Virginia University	Universitet	USA
Synchrophasor-based data mining for power system fault analysis	2012	kNN, NB, kMC	FA	KTH Royal Institute of Technology	Universitet	Sverige

PMU based robust state estimation using scaling	2012	LAV	SE	Northeastern University	Universitet	USA
Phase Angles as Predictors of Network Dynamic Security Limits and Further Implications	2014	LM, RF	SA	École de Technologie Supérieure, Hydro-Québec	Universitet, Kraftsel-skap(offentlig prod+dist)	Canada
A Neural Network Based Wide Area Monitor for a Power System	2005	NN	WAM	Missouri University of Science and Technology	Universitet	USA
Online Monitoring of Voltage Stability Margin Using an Artificial Neural Network	2010	NN	VS	University of Manitoba	Universitet	Canada
A real time PMU data and neural network approach to analyze voltage stability	2012	NN	VS	Kumamoto University	Universitet	Japan
Neural network method based on PMU data for voltage stability assessment and visualization	2011	NN	VS	Kumamoto University	Universitet	Japan
Wide area transient stability prediction using on-line Artificial Neural Networks	2008	NN	TS	Ain Shams University, Industrial Education Suez Canal University, University of Hail	Universitet	Egypt, USA
Operation State Prediction in Wind Power Integrated Systems Based on Artificial Neural Network	2015	NN	SE	Chongqing University	Universitet	Kina
A method for determining pseudo-measurement state values for topology observability of state estimation in power systems	2012	NN	SE	Meiji University	Universitet	Japan
Neural network-based power system dynamic state estimation using hybrid data from SCADA and phasor measurement units	2017	NN	SE	Shahid Beheshti University	Universitet	Iran
Artificial neural network based static security assessment module using PMU measurements for smart grid application	2016	NN	SA	Kalasalingam University	Universitet	India
PMU-ANN based real time monitoring of power system electromechanical oscillations	2016	NN	PSO	Malaviya National Institute of Technology	Universitet	India
Communication infrastructure effects on time detection of controlled islanding using robust neural network	2015	NN	ID	Shahid Beheshti University	Universitet	Iran
Adaptive Distance Relaying Scheme in Presence of UPFC using WAMS	2009	NN	DR	Indian Institute of Technology, Mississippi State University	Universitet	India, USA
A Statistical Process Control Approach for Automatic Anti-Islanding Detection Using Synchrophasors	2013	PCA	ID	Queen's University Belfast	Universitet	Storbritannia

Synchrophasor-Based Islanding Detection for Distributed Generation Systems Using Systematic Principal Component Analysis Approaches	2015	PCA	ID	Queen's University Belfast	Universitet	Storbritannia
Detection and Visualization of Power System Disturbances using Principal Component Analysis	2013	PCA	FL	Imperial College London	Universitet	Storbritannia
Fault Localization for Synchrophasor Data using Kernel Principal Component Analysis	2017	PCA	FL	Jilin University	Universitet	Kina
Principal components fault location based on WAMS/PMU measure system	2011	PCA	FL	North China Electric Power University	Universitet	Kina
Real-time event detection and feature extraction using PMU measurement data	2015	PCA	FL	University of Illinois	Universitet	USA
PCA fault location based on wide area measurement systems	2010	PCA	FL	North China Electric Power University	Universitet	Kina
Loss-of-Main Monitoring and Detection for Distributed Generations Using Dynamic Principal Component Analysis	2013	PCA	FD	Queen's University Belfast	Universitet	Storbritannia
Real-Time Multiple Event Detection and Classification Using Moving Window PCA	2016	PCA	FD	Queen's University Belfast	Universitet	Storbritannia
A novel radial basis function neural network principal component analysis scheme for PMU-based wide-area power system monitoring	2015	PCA, NN	WAM	Queen's University Belfast	Universitet	Storbritannia
PMU Based Monitoring of Inter-Area Oscillation in Thailand Power System via Home Power Outlets	2007	R	WAM	King Mongkut's Institute of Technology Ladkrabang, Kyushu Institute of Technology	Universitet	Thailand, Japan
PMU-based voltage instability detection through linear regression	2013	R	VS	KTH Royal Institute of Technology, Statnett SF	Universitet, TSO	Sverige, Norge
Improved method for real-time transient stability assessment of power systems	2012	R	TS	National University of San Juan, University Duisburg-Essen	Universitet	Argentina, Tyskland
Use of PMUs in regression-based power system dynamic state estimation	2012	R	SE	Virginia Tech	Universitet	USA
PMU Based State Estimation by Integrating Correlation	2013	R	SE	Arizona State University	Universitet	USA
Fault Location Using Wide-Area Measurements and Sparse Estimation	2016	R	FL	Northeastern University	Universitet	USA
Event detection method for the PMUs synchrophasor data	2012	R	FD	University of Texas	Universitet	USA

Research on real-time admittance matrix identification based on WAMS and multiple linear regression	2014	R	FA	Wuhan University	Universitet	Kina
Statistical Analysis and Forecasting of Damping in the Nordic Power System	2015	R	D	Lund University, Aalto University, Fingrid, Statnett SF	Universitet, TSO	Sverige, Finland, Norge
Syncrophasor measurement based small signal stability assesment using FFT-CWT approach in Japan-Campus-WAMS	2014	R	D	Kyushu Institute of Technology	Universitet	Japan
Transient stability assessment via decision trees and multivariate adaptive regression splines	2017	R, DT	TS	The University of British Columbia, BBA Inc.	Universitet, konsultentselskap	Canada
A Random Forest-Based Approach for Voltage Security Monitoring in a Power System	2014	RF	VSe	Melentiev Energy Systems Institute(SB RAS), University of Tasmania, TU Dortmund University	Forskningsinstitutt, universitet	Russland, Australia, Tyskland
A novel approach for instability detection based on wide-area measurements and new predictors	2016	RF	SE	École de technologie supérieure	Universitet	Canada
Development of New Predictors Based on the Concept of Center of Power for Transient and Dynamic Instability Detection	2016	RF	SE	École de technologie supérieure	Universitet	Canada
A New Adaptive Dependability-Security Approach to Enhance Wide Area Back-up Protection of Transmission System	2017	RF	SE	Indian Institute of Technology Delhi	Universitet	India
Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame	2010	RF	SA	McGill University, Hydro-Québec	Universitet, Kraftselskap(offentlig prod+dist)	Canada
Ensemble methods of classification for power systems security assessment	2017	RF	SA	Energy Systems Institute of Russian Academy of Sciences, Queen's University Belfast	Universitet	Russland, Storbritannia
On-Line Power Systems Security Assessment Using Data Stream Random Forest Algorithm Modification	2017	RF	SA	ESI SB RAS	Forskningsinstitutt	Russland
Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning	2013	RF	SA	Arizona State University	Universitet	USA
A New Approach to Power System Disturbance Assessment using Wide Area Post Disturbance Records	2017	RF	EC	Indian Institute of Technology Delhi	Universitet	India

Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices	2010	RF	CP	McGill University, Hydro-Québec	Universitet, Kraftselskap(offentlig prod+dist)	Canada
Synchrophasor Data Baseline and Mining for Online Monitoring of Dynamic Security Limits	2014	RF, DT	SA	École de Technologie Supérieure, Hydro-Québec	Universitet, Kraftselskap(offentlig prod+dist)	Canada
On-line Voltage Stability Load Index Estimation Based on PMU Measurements	2011	RNN	VS	Missouri University of Science and Technology	Universitet	USA
Regression Tree for Stability Margin Prediction Using Synchrophasor Measurements	2013	RT	StA	Texas A&M University	Universitet	USA
Study on a Real Time Model Predictive based Power System SIPS	2013	S	PS	University of Manchester	Universitet	Storbritannia
Genetic algorithm based support vector machine for on line voltage stability monitoring	2015	SVM	VS	Indian Institute of Technology Roorkee	Universitet	India
Power System Transient Stability Analysis Based on PMU and Hybrid Support Vector Machine	2016	SVM	TS	Huazhong University of Science and Technology, Zhejiang University	Universitet	Kina
Dynamic Vulnerability Assessment due to Transient Instability based on Data Mining Analysis for Smart Grid Applications	2011	SVM	TS	National University of San Juan, XM "Electricity Market Experts"	Universitet, energimarkedsselskap	Argentina, Colombia
A Hierarchical Method for Transient Stability Prediction of Power Systems Using the Confidence of a SVM-Based Ensemble Classifier	2016	SVM	TS	Beijing Jiaotong University, China Electric Power Research Institute	Universitet, forskningsinstitutt	Kina
Support Vector Machine-Based Algorithm for Post-Fault Transient Stability Status Prediction Using Synchronized Measurements	2010	SVM	TS	University of Manitoba, Manitoba Hydro	Universitet, kraftselskap	Canada
Online ensemble learning for security assessment in PMU based power system	2016	SVM	SA	University of Québec	Universitet	Canada
Transmission line fault detection and localisation methodology using PMU measurements	2014	SVM	FL	National Institute of Technology, MVGR College of Engineering Vizianagaram	Universitet	India

Line outage detection using support Vector Machine (SVM) based on the Phasor Measurement Units (PMUs) technology	2012	SVM	FD	Ain Shams University	Universitet	Egypt
Preventing transmission distance relays maloperation under unintended bulk DG tripping using SVM-based approach	2017	SVM	DR	Texas A&M University	Universitet	USA
Fast Sequence Component Analysis for Attack Detection in Synchronphasor Networks	2015	SVM	AD	Oregon State University, Washington State University, Portland State University	Universitet	USA
Evaluating performance of classifiers for supervisory protection using disturbance data from phasor measurement units	2014	SVM, kNN, NB, RT	EC	New Mexico State University	Universitet	USA
Performance Comparison of Classification Methods for Line Outage Detection	2016	SVM, NB, kNN	FD	Ain Shams University	Universitet	Egypt
SOFM based support vector regression model for prediction and its application in power system transient stability forecasting	2006	SVR	TS	Huazhong University of Science & Technology	Universitet	Kina
Real-time transient stability assessment based on centre-of-inertia estimation from phasor measurement unit records	2014	SVR	TS	CENACE, University Duisburg-Essen, National University of San Juan	Netteier, universitet	Ecuador, Tyskland, Argentina

Tabell A.1: Her vises all litteratur funnet i litteratursøket av maskinlæring brukt på PMU-data.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway