

Prosjektnotat

TITTEL			
Uttesting og implementasjon av estimering av trafotemperatur			
WORK PACKAGE	VERSJON	DATO	ANTALL SIDER
WP Pilot	1.0	2021-08	14
FORFATTER(E)	WP-LEDER		GRADERING
Ida Risnes Hansen			Åpen

Innholdsfortegnelse

1	Problembakgrunn	3
1.1	Hotspot-temperaturens effekt på aldring	3
1.2	Estimere hotspot med maskinlæring	4
1	Problembakgrunn	4
1.1	Hot-spot temperaturens effekt på aldring	4
1.2	Estimere hotspot med maskinlæring	5
2	Datagrunnlag	5
2.1	Forklaringsvariabler	5
2.2	Treningsdata	5
2.3	Begrensninger og forbehold	5
3	Metodikk	6
3.1	Maskinlæringsalgoritme: XGBoost	6
3.2	Kryssvalidering	6
3.3	Forklaringsvariabler	6
3.3.1	Basisvariabler	6
3.3.2	Differansen	6
3.3.3	Forsinkede variabler (lag features)	7
3.4	Modellinndeling	7
3.4.1	Stasjonstype	8
3.4.2	Lastprosent og årstid	8
3.5	Hyperparametre	9
4	Resultater	9
4.1	Toppmodeller	9
4.1.1	Mast	10
4.1.2	Kiosk	10
4.1.3	Bygg	10
4.2	Optimalisering	10
5	Diskusjon	10
5.1	Kryssvalidering	11
5.2	Inndeling i last og årstid	11
5.3	Forklaringsvariabler	11
6	Videre arbeid	11
7	Referanser	11

8 Appendiks	12
8.1 Gjennomsnittlige resultater fra kryssevaluering av forskjellige modeller	12
8.1.1 Mast.....	12
8.1.2 Kiosk	13
8.1.3 Bygg	14

1 Problembakgrunn

Transformatorer, eller trafoer, er en uunnværlig del av strømnettet. Agder Energi har mange tusen nettstasjoner med trafoer rundt i distriktet. Vedlikehold og utskifting av disse tar opp mye tid og ressurser. Å finne en effektiv måte å monitorere helsetilstanden til disse er derfor av stor interesse.

En viktig faktor i denne sammenhengen er maks-temperaturen inni vikinglene til trafoen; den såkalte hot-spot temperaturen. Er denne for høy over tid vil levetiden til trafoen bli kraftig redusert.

1.1 Hot-spot temperaturens effekt på aldring

I en rapport fra Møre Trafo (Longva, 1999) er aldringshastighet A_{MT} og hot-spot temperaturen T_{HS} koblet sammen på følgende måte:

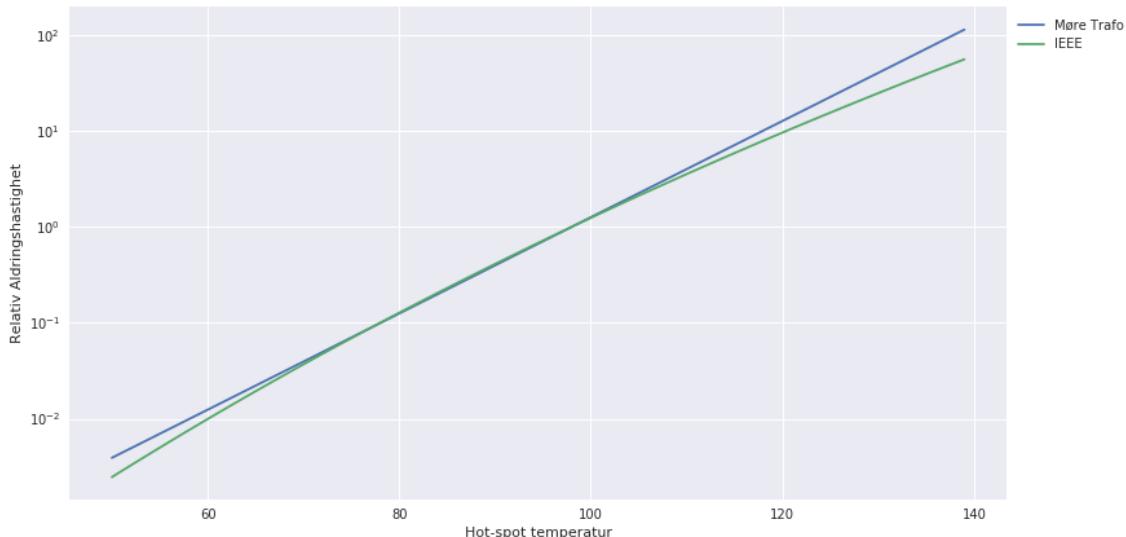
$$A_{MT}(T_{HS}) = 2^{(98-T_{HS})/6}$$

Uttrykket tar utgangspunkt i en antagelse om at en transformator med konstant hot-spot temperatur på 98 grader vil ha en levetid på 14 år. For hver temperaturøkning på 6 grader vil aldringshastigheten dobles, og levetiden halveres. Tilsvarende vil en reduksjon på 6 grader doble levetiden.

En lignende relasjon er gitt i en artikkel i IEEE fra 2008:

$$A_{IEEE}(T_{HS}) = e^{\left(\frac{15000}{383} - \frac{15000}{273+T_{HS}}\right)}$$

Figur 1 viser at de to utrykkene gir lignende sammenhenger.



Figur 1 Den relative aldringshastigheten for ulike hot-spot temperaturer.

Disse er begge empiriske uttrykk, og det er uklart hvor godt de stemmer på generell basis. De er likevel gitt her som et mer konkret eksempel på hvordan sammenhengen mellom hot-spot temperaturen og levetiden til en transformator kan se ut.

Hot-spot temperaturen vil normalt variere over tid. Den gjennomsnittlige relative aldringshastigheten er da gitt

$$A_{MT,avg} = \frac{\int 2^{(98-T_{HS})/6} dt}{\int dt}$$

1.2 Estimere hotspot med maskinlæring

Noen av de viktigste faktorene som påvirker temperaturen til en transformator er belastningen man gir den, temperaturen til omgivelsene og individuelle egenskaper som kjøleevne. Det finnes modeller som kobler disse tingene sammen (se for eksempel Jean-Noël Bérubé 2004). Disse inneholder derimot oftest empiriske konstanter som ikke er lett tilgjengelige for enhver trafokrets, og tar heller ikke høyde for ting som værpåvirkning.

Målet i dette prosjektet har derfor vært å bruke maskinlæring til å trenere modeller som kan predikere trafo-temperaturer basert på lett tilgjengelig data som vær og belastning. Leander Berg Thorkildsen skrev våren 2021 en masteroppgave for Agder Energi hvor han trente en slik modell for transformatorer i stasjonstypen kiosk (Thorkildsen 2021).

Dette er en videreføring hans arbeid, hvor vi har trent modeller for flere stasjonstyper og sett nærmere på hvor generalisert disse er til resten av trafoene til Agder Energi.

2 Datagrunnlag

2.1 Forklарingsvariable

Agder Energi har rundt 8000 nettstasjoner fordelt rundt i distriktet. For de fleste av disse har de tilgang til belastningsdata på timesnivå. Fra meteorologisk institutt har de tilgang til værdata som lufttemperatur, nedbør, luftfuktighet og vind. Det er derfor gitt disse variablene vi ønsker å predikere trafo-temperaturen. Det er ikke gitt at dette er et tilstrekkelig grunnlag for å få en troverdig og stabil modell, men per i dag er det hva modellene baserer seg på.

2.2 Treningsdata

For å trenere modellene er det satt ut temperatur-sensorer fra Disruptive Technologies på rundt 25 trafoer. I dette prosjektet har vi brukt dataen samlet inn av disse mellom november 2019 og juni 2021. De fleste av trafoene har en sensor på toppen og en på bunnen. I treningen av modellene har vi da valgt å bruke data fra kun en av disse slik at det alltid er enten topp-temperaturen eller bunn-temperaturen modellene er trenet på.

2.3 Begrensninger og forbehold

Siden temperatursensorene er plassert på utsiden av transformatorene er det i praksis ikke hotspot-temperaturen man trener modellene til å predikere, men et under-estimat av denne. For å bruke utrykkene for relativ aldringshastighet gitt over, er det derfor nødvendig å legge til en temperaturmargin.

For alle trafokretsene manglet det data fra midten av desember til midten av januar. I tillegg mangler enkelte av trafokretsene data i andre intervaller.

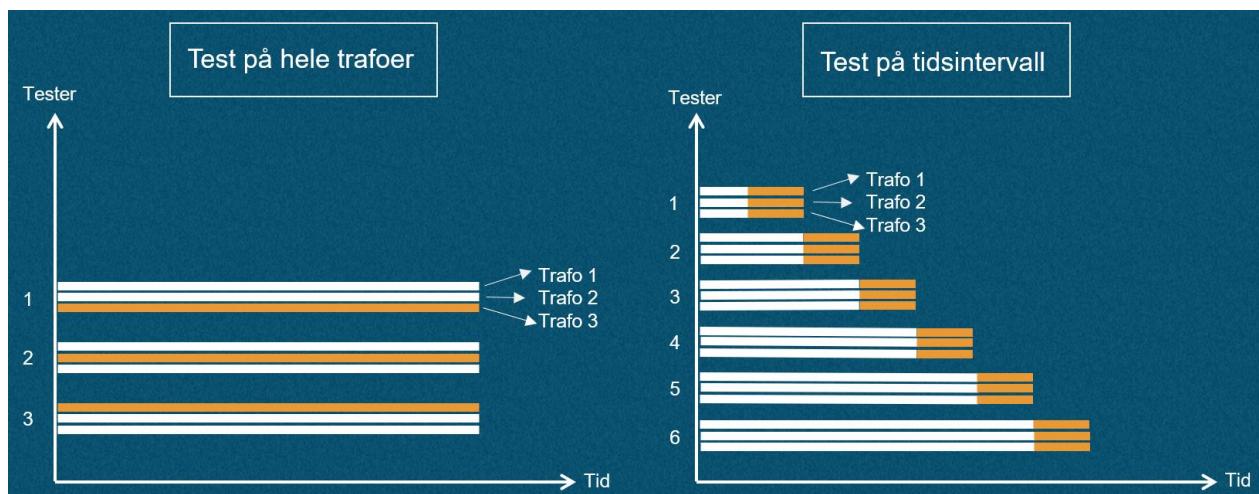
3 Metodikk

3.1 Maskinlæringsalgoritme: XGBoost

Vi har brukt Python sitt XGBoost bibliotek for å trenne modellene.

3.2 Kryssvalidering

Vi kryssvaliderede modellene på to måter. Først ved å teste på hele trafokretser, så ved tidsserie-kryssvalidering.



Figur 2 Kryssvalidering ved test på hele trafokretser, og ved test på tidsintervaller.

3.3 Forklарingsvariable

3.3.1 Basisvariable

Den tilgjengelige last- og værdataen gir følgende forklaringsvariable:

1. Lastprosent (last i prosent av merkeytelsen til trafoen)
2. Lufttemperatur
3. Nedbør
4. Luftfuktighet
5. Vind
6. Måned

Dette er utgangspunktet for alle modellene vi lager.

3.3.2 Differansen

I tillegg til det vi har kalt basisvariablene over, har vi testet å legge til endringen i lastprosenten og lufttemperaturen per time. Her er tanken at hvor fort belastningen eller lufttemperaturen endrer seg spiller inn på endringen i trafo-temperaturen.

For en forklaringsvariabel F blir

$$F_{diff,i} \equiv \frac{\Delta F_i}{\Delta t_i} = \frac{F_{i+1} - F_i}{t_{i+1} - t_i}$$

Hvis det er mer enn tre timer mellom to datapunkter ($\Delta t_i > 3 \text{ timer}$) settes endringen til NaN.

3.3.3 Forsinkede variabler (lag features)

Til slutt har vi testet å legge til forsinkede variabler, eller «lag features». Dette er motivert av at vi tror det vil være en forsinkelse mellom last- og vær-påvirkningen og endringen i trafo-temperaturen. For et gitt datapunkt vil det da være viktig å vite hva lasten og været var tilbake i tid.

For en forklaringsvariabel F vil den forsinkede variablen j timer tilbake i tid være gitt,

$$F_{lag,i} \equiv F_{i-j}.$$

For lastprosent og temperatur testet vi å legge til forsinkede variabler 24 timer tilbake i tid. For resten av værdataen og for differansevariablene testet vi å legge til data opp til 3 timer tilbake i tid.

For å slippe å legge til hver time og dermed redusere antallet forklaringsvariabler prøvde vi også å bruke et rullende gjennomsnitt. Det rullende gjennomsnittet mellom a og b timer tilbake i tid for en forklaringsvariabel F er gitt

$$F_{mean\ lag,i} \equiv \frac{\sum_{j=a}^b F_{i-j}}{b-a}$$

Forklaringsvariabel	Lag	Rullende gjennomsnitt
Lastprosent	1, 2, 3, ..., 24	(4-6), (7-9), (10-12), (12-24)
Lufttemperatur	1, 2, 3, ..., 24	(4-6), (7-9), (10-12), (12-24)
Vind	1, 2, 3, ..., 24	(1-3)
Luftfuktighet	1, 2, 3, ..., 24	(1-3)
Nedbør	1, 2, 3, ..., 24	(1-3)
Måned	1, 2, 3, ..., 24	(1-3)
Endring i lastprosent	1, 2, 3, ..., 24	(1-3)
Endring i lufttemperatur	1, 2, 3, ..., 24	(1-3)

Vi testet ved å legge til flere og flere lag-timer for hver av variablene, og deretter ved å erstatte lag-timene etter 3 timer med rullende gjennomsnitt.

3.4 Modellinndeling

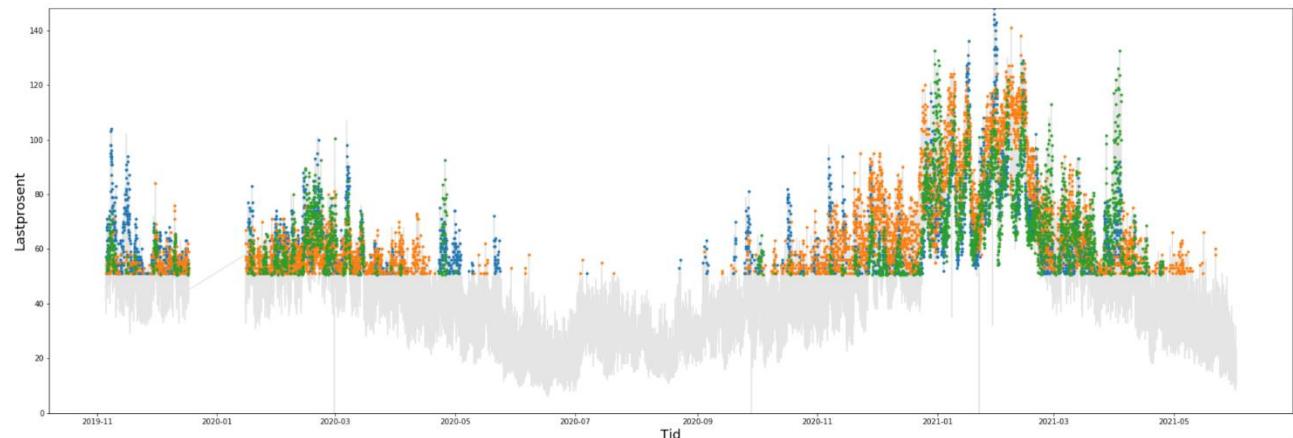
3.4.1 Stasjonstype

Agder Energi har tre hovedtyper av nettstasjoner: mast, kiosk og bygg. Det er naturlig å anta at disse vil påvirkes i ulik grad av de forskjellige forklaringsvariablene. Værvariablene kan for eksempel virke inn på en trafo festet i en mast på en helt annen måte enn for en stående i et godt isolert bygg.

Vi har derfor trent ulike modeller for disse tre typene. Det er data fra 3 master, 6 bygg og 14 kiosker i datasettet. For byggene og kioskene trener vi modellene basert på sensor-temperaturen på toppen av trafokretsen, mens vi ser på sensoren på bunnen av trafoene i mast. Dette valget ble gjort basert på hvor flertallet av temperaturverdiene var høyest.

3.4.2 Lastprosent og årstid

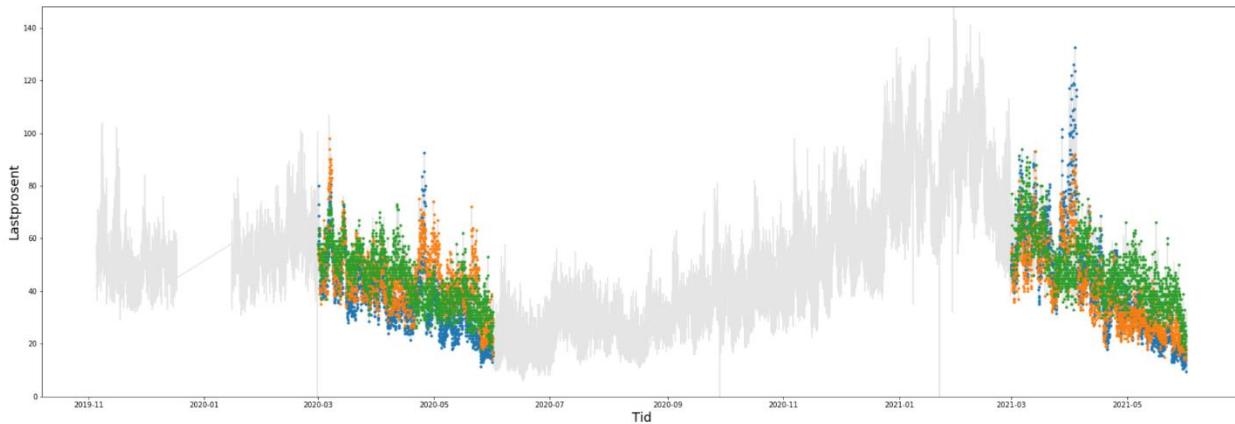
I tillegg til å lage ulike modeller for de forskjellige stasjonstypene har vi testet å dele inn i modeller basert på last og årstid. For last-inndelingen lagde vi en modell for alle datapunktene med belastning under 50% av merkeytelsen til trafoen, og en for datapunktene over 50%.



Figur 3 Eksempel på datagrunnlaget for en modell med belastning over 50%. De ulike fargene representerer de tre ulike trafokretsene i mast.

I årstidsinndelingen lagde vi ulike modeller basert på dataen fra de ulike årstidene:

- Vinter: desember-februar
- Vår: mars-mai
- Sommer: juni-august
- Høst: september-november



Figur 4 Eksempel på datagrunnlaget for en modell for våren (mars-mai). De ulike fargene representerer de tre ulike trafokretsene i mast.

Til slutt testet vi å lage modeller hvor vi kombinerte disse inndelingene og laget en høy-last og en lav-last modell for hver årstid. Dette ga totalt 8 modeller for hver av stasjonstypene.

3.5 Hyperparametre

XGBoost har mange parametre som kan justere detaljene i hvordan algoritmen fungerer (XGBoost Parameters 2020). På grunn av tidsmangel gjennomførte vi bare et kort randomisert søk etter hyperparametre for «topp-modellene» for hver av de tre stasjonstypene. Tabellen nedenfor viser alle verdiene i «gridet».

Parameter	Range
<i>learning_rate</i>	0.01, 0.1, 0.3
<i>min_child_weight</i>	1, 5, 10
<i>gamma</i>	0.5, 1, 1.5, 2, 5
<i>subsample</i>	0.6, 0.8, 1.0
<i>colsample_bytree</i>	0.6, 0.8, 1.0
<i>n_estimators</i>	100, 200, 300, 400
<i>max_depth</i>	3, 4, 5, 6, 8, 10

Av totalt 31104 mulige kombinasjoner rakk vi kun å søke gjennom 2000 av dem for hver stasjonstype.

4 Resultater

4.1 Toppmodeller

Nedenfor er resultatene fra toppmodellene for de tre stasjonstypene. De fullstendige resultatene er lagt til i appendiksen.

4.1.1 Mast

Kryssvalidering	Modellinndeling	Forklaringsvariabler	RMSE	R2
Tid	Last	Differanse	2.86	0.80
Trafo	Ingen	Differanse	3.05	0.75

4.1.2 Kiosk

Kryssvalidering	Modellinndeling	Forklaringsvariabler	RMSE	R2
Tid	Årstid	Differanse	4.55	0.73
Trafo	Last og årstid	Differanse	5.97	0.39

4.1.3 Bygg

Kryssvalidering	Modellinndeling	Forklaringsvariabler	RMSE	R2
Tid	Årstid	Differanse	3.46	0.56
Trafo	Ingen	Basisvariabler	7.97	-1.33

4.2 Optimalisering

Det randomiserte grid-søket ga følgende parametere for toppmodellene for de tre typene.

Parameter	Range	Mast	Kiosk	Bygg
<i>learning_rate</i>	0.01, 0.1, 0.3	0.1	0.1	0.1
<i>min_child_weight</i>	1, 5, 10	5	10	1
<i>gamma</i>	0.5, 1, 1.5, 2, 5	5	0.5	0.5
<i>subsample</i>	0.6, 0.8, 1.0	0.6	0.8	1.0
<i>colsample_bytree</i>	0.6, 0.8, 1.0	0.8	1.0	0.6
<i>n_estimators</i>	100,200,300,400	300	400	300
<i>max_depth</i>	3,4,5,6,8,10	3	4	8

På grunn av tidsmangel rakk vi ikke teste modellene med de optimaliserte parameterne. Siden det randomiserte søket bare testet 1/30 av alle kombinasjonene er det ganske trygt å anta at disse heller ikke er de faktisk beste parameterne, og det kan være hensiktsmessig å gjøre et mer komplett søk.

5 Diskusjon

På grunn av dårlig disponering av tid er dette dessverre en litt mangefull analysedel uten figurer. Forhåpentligvis kan PowerBI-rapporten (Estimering av traftemperatur med sensor) brukes litt mer aktivt for å analysere resultatene for de ulike modellene.

5.1 Kryssvalidering

Vi ser at kryssvalideringen basert på tidsintervaller konsistent gir de beste resultatene. Dette kan være en indikasjon på at modellen gir bedre prediksjoner når den er trent på data fra trafokretsen det testes på. Kryssvalideringen på hele trafostasjoner gjør det det særlig dårlig på bygg og kiosker.

Datasettene for mastene er rimelig lik hverandre; de har lignende fordelinger av lastprosent og temperaturer. Det er også tydelig at disse sensortemperaturene er svært avhengig av lufttemperaturen. Dette kan forklare hvorfor den trafo-baserte kryssvalideringen gjør det greit her med en RMSE på 3.05 grader.

For kioskene og byggen er det større sprik i hvordan forklaringsvariablene er fordelt. Siden datasettene er små når det kommer til antall trafokretser gir dette spraket mye mer varierte resultater. Det er også ting som tyder på at ulike trafokretser responerer forskjellig på lignende forklaringsvariabler.

5.2 Inndeling i last og årstid

Vi ser at modellinndelingene basert på last og årstid i mange tilfeller kan gi en noe bedre RMSE-verdi. Det er likevel ikke store forskjeller, og i snitt er gevinsten relativt liten sammenlignet med å samle all dataen. Lastinndelingen kan gi merkelige resultater i tidsintervaller hvor den oscillerer rundt 50%. Da kan det bli mange hull i datasettene. En årstid-inndeling kan virke mer hensiktsmessig. Spesielt for vintermånedene hvor lastprofilene ofte skiller seg ut.

5.3 Forklaringsvariabler

Mens differanse-variablene forbedret de fleste modellene, ga både de forsinkede variablene og det rullende gjennomsnittet liten gevinst i snitt. Dette kan være en konsekvens av at dataen er for tynn og sprikende til at modellene klarer å fange opp disse mønstrene.

6 Videre arbeid

Det er liten tvil om at det er mulig å forbedre modellene fra dette prosjektet. Både når det kommer til hvilke forklaringsvariabler man velger, inndelingen av modellene og hyperparameter-optimalisering. Samtidig ville et bedre datagrunnlag gi helt andre muligheter til å evaluere modellene. Og nettopp dette er viktig når modellene etter hvert skal ut i produksjon og predikere temperaturen på trafokretser hvor vi ikke har fasiten.

7 Referanser

- Jean-Noël Bérubé, Jacques Aubin, W. McDermid. 2004. «Transformer winding hot spot temperature determination.»
- Longva, Kårstein. 1999. *Fordelingstransformatorer og levetid*. Møre Trafo AS.
- Thorkildsen, Leander Berg. 2021. «Estimation of temperature development in secondary substation transformers.»
2020. «XGBoost Parameters.» *XGBoost Documentation*.
<https://xgboost.readthedocs.io/en/latest/parameter.html>.

8 Appendiks

8.1 Gjennomsnittlige resultater fra kryssevaluering av forskjellige modeller

Nedenfor følger hovedresultatene fra kryssvalideringen av de ulike modellene.

8.1.1 Mast

Trening	Modellinndeling	Forklарingsvariabler	RMSE	R2
Tid	Ingen	Basisvariabler	3.13	0.76
		Differanse	3.02	0.78
		Lag	3.30	0.73
		Rullende gjennomsnitt	3.46	0.70
	Last	Basisvariabler	3.03	0.77
		Differanse	2.86	0.80
		Lag	3.10	0.77
		Rullende gjennomsnitt	3.19	0.76
	Årstid	Basisvariabler	3.08	0.77
		Differanse	2.93	0.79
		Lag	3.16	0.75
		Rullende gjennomsnitt	3.22	0.74
	Last og årstid	Basisvariabler	3.11	0.76
		Differanse	2.93	0.79
		Lag	3.22	0.76
		Rullende gjennomsnitt	3.30	0.75
Trafo	Ingen	Basisvariabler	3.27	0.74
		Differanse	3.07	0.77
		Lag	3.13	0.75
		Rullende gjennomsnitt	3.10	0.75
	Last	Basisvariabler	3.33	0.73
		Differanse	3.17	0.75
		Lag	3.18	0.76
		Rullende gjennomsnitt	3.20	0.75
	Årstid	Basisvariabler	3.37	0.72
		Differanse	3.05	0.76
		Lag	3.23	0.74
		Rullende gjennomsnitt	3.20	0.73

	Last og årstid	Basisvariabler	3.41	0.72
		Differanse	3.17	0.76
		Lag	3.32	0.72
		Rullende gjennomsnitt	3.27	0.73

8.1.2 Kiosk

Trening	Inndeling	Forklарingsvariabler	RMSE	R2
Tid	Ingen	Basisvariabler	5.09	0.66
		Differanse	4.60	0.72
		Lag	5.23	0.64
		Rullende gjennomsnitt	5.23	0.64
	Last	Basisvariabler	5.18	0.64
		Differanse	4.76	0.70
		Lag	5.10	0.64
		Rullende gjennomsnitt	4.92	0.65
	Årstid	Basisvariabler	4.94	0.68
		Differanse	4.55	0.73
		Lag	5.14	0.65
		Rullende gjennomsnitt	5.17	0.64
	Last og årstid	Basisvariabler	5.00	0.66
		Differanse	4.57	0.72
		Lag	4.98	0.65
		Rullende gjennomsnitt	4.84	0.65
Trafo	Ingen	Basisvariabler	7.10	0.33
		Differanse	6.87	0.37
		Lag	7.35	0.28
		Rullende gjennomsnitt	7.56	0.24
	Last	Basisvariabler	7.22	0.31
		Differanse	6.24	0.36
		Lag	6.93	0.35
		Rullende gjennomsnitt	6.45	0.31
	Årstid	Basisvariabler	7.37	0.28
		Differanse	7.11	0.33
		Lag	7.30	0.29
		Rullende gjennomsnitt	7.36	0.28

	Last og årstid	Basisvariabler	7.36	0.26
		Differanse	5.97	0.39
		Lag	6.40	0.41
		Rullende gjennomsnitt	6.04	0.39

8.1.3 Bygg

Trening	Inndeling	Forklарingsvariabler	RMSE	R2
Tid	Ingen	Basisvariabler	3.97	0.42
		Differanse	3.82	0.46
		Lag	4.00	0.41
		Rullende gjennomsnitt	4.05	0.40
	Last	Basisvariabler	3.96	0.42
		Differanse	3.80	0.47
		Lag	4.05	0.40
		Rullende gjennomsnitt	4.05	0.40
	Årstid	Basisvariabler	3.69	0.50
		Differanse	3.46	0.56
		Lag	3.80	0.47
		Rullende gjennomsnitt	3.80	0.46
	Last og årstid	Basisvariabler	3.68	0.50
		Differanse	3.47	0.56
		Lag	3.80	0.47
		Rullende gjennomsnitt	3.85	0.45
Trafo	Ingen	Basisvariabler	7.97	-1.33
		Differanse	8.31	-1.53
		Lag	8.49	-1.65
		Rullende gjennomsnitt	8.83	-1.86
	Last	Basisvariabler	7.98	-1.34
		Differanse	8.34	-1.56
		Lag	8.52	-1.66
		Rullende gjennomsnitt	8.81	-1.85
	Årstid	Basisvariabler	8.24	-1.49
		Differanse	8.16	-1.60
		Lag	8.40	-1.45
		Rullende gjennomsnitt	8.16	-1.45

	Last og årstid	Basisvariabler	8.24	-1.49
		Differanse	8.04	-1.37
		Lag	8.24	-1.61
		Rullende gjennomsnitt	8.19	-1.46