

Sebastian Hjertaas Paulsen
Erik Kolås

Prediksjon av strømpriser ved bruk av nevrale nettverk

Bacheloroppgave i Økonomi og administrasjon
Veileder: Denis Mike Becker
April 2022

Sebastian Hjertaas Paulsen
Erik Kolås

Prediksjon av strømpriser ved bruk av nevrale nettverk

Bacheloroppgave i Økonomi og administrasjon
Veileder: Denis Mike Becker
April 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen



NTNU

Kunnskap for en bedre verden

Forord

Denne bacheloroppgaven er skrevet av to studenter ved NTNU Handelshøyskolen i Trondheim, våren 2022. Oppgaven representerer slutten på tre lærerike år på studiet økonomi og administrasjon. Innholdet i denne oppgaven står for forfatterens regning.

Bacheloroppgaven omhandler prediksjon av strømpriser ved bruk av nevrale nettverk. Vi har dermed fått et innblikk i hvordan vi kan benytte maskinlæring i økonomiske applikasjoner. Det har også gitt oss kunnskap om hvordan analytikere jobber for å analysere kommende strømpriser.

Det har vært både spennende og interessant å skrive om dette temaet. Vi har møtt på flere utfordringer fra å innhente datamateriale, omforme datasettet og til å utforme de nevrale nettverkene vi har benyttet. Det har vært en bratt læringskurve hvor vi har lært veldig mye. Kodingen av oppgaven har gitt oss mye glede og mestringsfølelse etter mye strev og stadige nye utfordringer. Temaet om strømpriser er ikke bare relevant for oss som skriver oppgaven, men er også veldig relevant for folk flest som betaler mye for strøm. Det å kunne få et innblikk i hvorfor strømprisene faktisk svinger som de gjør har vært veldig interessant.

Vi retter en stor takk til vår veileder, Denis Mike Becker, som gjennom våren har gitt oss særdeles god veiledning, raske tilbakemeldinger og konstruktive innspill. Vi setter stor pris på hans hjelp, kunnskap og interesse for temaet i denne oppgaven. Videre vil vi også takke andre medstudenter som har bidratt med gode råd i gjennomføringen av oppgaven i sin helhet.

God lesing!

Sammendrag

I denne bacheloroppgaven har vi gjennomført prediksjoner av strømpriser ved bruk av nevrale nettverk. Vi har ved bruk av programmeringsspråket Python kodet ulike nevrale nettverk til å kunne predikere strømpriser. Nevrale nettverk er ofte illustrert ved bruk av en menneskehjerne, hvor nervetråder er representert som noder. Ved å tilpasse nettverkene til datasettet fikk vi predikert en time med fremtidige strømpriser.

Vårt datasett består av femten år med timebaserte historiske observasjoner. Strømpriser er satt som avhengig variabel, da det er denne vi ønsket å predikere. Videre valgte vi ut de uavhengige variablene som kunne antas å ha størst innvirkning på strømprisene. De variablene vi valgte var de mest relevante råvareprisene, målinger for værdata og en variabel for risiko i markedet.

I oppgaven har det blitt brukt tre ulike typer nevrale nettverk:

- “Feedforward neural network” (ANN)
- “Recurrent neural network” (RNN)
- “Long-short term memory” (LSTM)

Forskjellen på nettverkene er at ANN ikke kan predikere fremtidig sekvensielle data, slik som er mulig for RNN og LSTM. Disse klarer å ta vare på predikerte data og benytte de til videre prediksjon. Vi benyttet likevel ANN i oppgaven for å sammenligne treffsikkerheten mellom nettverkene. I tillegg lagde vi to ulike versjoner av LSTM, for å teste om flere lag og noder kan være med å øke treffsikkerheten.

De ulike nevrale nettverkene fikk alle prediksjoner med en gjennomsnittlig feilmargin på mellom 3,9 og 6,2 øre/kWh, avhengig av hvilken type nettverk som ble brukt. Videre forklarte de uavhengige variablene over 90 prosent av variansen til strømprisene i alle de ulike nevrale nettverkene. Det nettverket som ga de beste resultatene for prediksjon av fremtidige strømpriser var LSTM-nettverket med hele 96 prosent forklart varians og en gjennomsnittlig feilmargin på 4,4 øre/kWh. Nettverkene følger et reaktivt handlingsmønster, noe som gir svekkede muligheter for prediksjoner av ukjente fremtidige strømpriser. Oppgaven vil gå mer i dybden om de nevrale nettverkene og implikasjonene som følger av resultatene.

Abstract

In this bachelor thesis, we have carried out predictions of electricity prices using neural networks. Using the programming language Python, we have coded various neural networks to be able to predict electricity prices. Neural networks are often illustrated using a human brain, where nerve fibers are represented as nodes. By adapting the networks to the dataset, we predicted an hour of future electricity prices.

Our dataset consists of fifteen years of hourly historical observations. Electricity prices are set as a dependent variable, as this is what we wanted to predict. Furthermore, we selected the independent variables that could be assumed to have the greatest impact on electricity prices. The variables we chose were the most relevant commodity prices, measurements for weather data and a variable for risk in the market.

Three different types of neural networks have been used in the thesis:

- “Feedforward neural network” (ANN)
- “Recurrent neural network” (RNN)
- “Long-short term memory” (LSTM)

The difference between the networks is that ANN cannot predict future sequential data, as is possible for RNN and LSTM. These are able to take care of predicted data and use it for further prediction. We still used ANN in the task to compare the accuracy between the networks. In addition, we created two different versions of LSTM, to test whether more layers and nodes can help increase accuracy.

The various neural networks all received predictions with an average margin of error between 3.9 and 6.2 øre/kWh, depending on the type of network used. Furthermore, the independent variables explained over 90 percent of the variance in electricity prices in all the various neural networks. The network that gave the best results for predicting future electricity prices was the LSTM network with as much as 96 percent explained variance and an average margin of error of 4.4 øre/kWh. The networks follow a reactive pattern of action, which provides weakened opportunities for predictions of unknown future electricity prices. The thesis will go more in depth about the neural networks and the implications that follow from the results.

Innholdsfortegnelse

1. Innledning	1
1.1 Innledning	1
1.2 Problemstilling.....	2
1.3 Avgrensninger.....	2
2. Teori.....	3
2.1 Maskinl�ring	3
2.2 Nevrale nettverk.....	3
2.2.1 Kunstige nevrale nettverk	3
2.2.2 Recurrent Neural Network.....	4
2.2.3 Long Short-Term Memory.....	5
2.3 Tidligere forskning rundt problemstillingen.....	5
3. Data og metode	8
4. Analyse	11
4.1 Visualisering av variabler	11
4.1.1 Str�mpriser.....	11
4.1.2 De uavhengige variablene.....	13
4.1.3 Forhold mellom str�mpriser og uavhengige variabler.....	15
4.2 Utforming av nevrale nettverk	17
4.2.1 Beskrivelse av nettverkene	17
4.2.2 ANN.....	17
4.2.3 RNN og LSTM	18
4.3 Predikerte resultater av nevrale nettverk.....	18
4.3.1 Beskrivelse.....	18

4.3.2 ANN.....	20
4.3.3 LSTM.....	21
4.3.4 RNN	22
4.3.5 Stacked LSTM	23
5. Diskusjon	24
5.1 Sammenligning og diskusjon av predikerte resultater	24
5.2 Evaluering og forbedring av modellene.....	26
6. Konklusjon og videre arbeid.....	29
6.1 Konklusjon.....	29
6.2 Videre arbeid.....	29
7. Referanseliste	31
8. Vedlegg.....	34

Liste over figurer

Figur 1: Utforming av et nevralt nettverk	4
Figur 2: Illustrasjon av RNN og LSTM.....	4
Figur 3: Strømpriser visualisert i timer	11
Figur 4: Strømpriser visualisert i uker	12
Figur 5: Sammenligning av strømpriser fra 2008 og 2021	13
Figur 6: Sammenligning av strømpriser fra 2010 og 2021	13
Figur 7: De uavhengige variablene i hvert sitt diagram.....	14
Figur 8: Sammenligning av strøm-, olje-, gass- og kullpriser	16
Figur 9: Sammenligning av strømpriser og temperatur	16
Figur 10: ANN resultater for høye strømpriser.....	20
Figur 11: ANN resultater for lave strømpriser.....	20
Figur 12: LSTM resultater for høye strømpriser.....	21
Figur 13: LSTM resultater for lave strømpriser.....	21
Figur 14: RNN resultater for høye strømpriser	22
Figur 15: RNN resultater for lave strømpriser	22
Figur 16: Stacked LSTM resultater for høye strømpriser	22
Figur 17: Stacked LSTM resultater for lave strømpriser	23
Figur 18: Sammenligning av ukentlige prediksjoner.....	25
Figur 19: Prediksjon av den første timen av 2022	26
Figur 20: Trening av LSTM.....	27
Figur 21: Trening av Stacked LSTM	27

Liste over tabeller

Tabell 1: Deskriptiv statistikk av årlige strømpriser.....	12
Tabell 2: Korrelasjonsmatrise	15

1. Innledning

1.1 Innledning

Strøm, og spesielt strømprisene, har fått mye oppmerksomhet i media det siste året. “Strømkrisen er ikke over” (Blaker, 2022), “[f]lere bør vurdere fastpris på strøm” (Kroepelien, 2022), og “[r]egningen for desember-strømmen er klar: - Som å betale 100 kroner for et brød” (Solli, 2021), er bare noen få av artiklene skrevet av norske medier rundt årsskiftet 2021/2022. Men hvorfor har mediene gitt dette temaet så mye oppmerksomhet? Året 2020 bydde på mye snø og nedbør i fjellene som resulterte i svært høy produksjonskapasitet av strøm ved vannkraftverkene. I tillegg var det en relativt mild vinter og vår som medførte til en lavere etterspørsel for oppvarming av landets husholdninger. Kombinasjonen av høy strømproduksjon i vannkraftverkene, og lav etterspørsel etter strøm blant folket, førte til at vannmagasinene forble fulle. I tillegg var eksporten av strøm til utlandet redusert grunnet lavere kapasitet på kablene til andre land (Energi Norge, 2022b). Dette resulterte i svært lave strømpriser. På noen tidspunkter var strømprisen til og med negativ, hvis man ikke tok med nettleie og avgifter i betraktningen (Hovland, 2020). Men situasjonen i det påfølgende året skulle vise seg å være en helt annen.

Gassprisene ble firedoblet fra fjoråret, CO₂-prisene doblet (Aurdal, 2021), og kullprisene ble tredobbelt (Lang-Ree og Sæter, 2021). Dette førte til høyere internasjonale strømpriser, spesielt i land som produserer strøm av nevnte råvarer. I tillegg fikk Norge to nye kabler til Europa, noe som gjorde det svært attraktivt å eksportere strøm. Det var en kald vinter, og behovet for strøm til oppvarming av hus i Norge var stort. Og på toppen av det hele var det uvanlig lite nedbør i andre halvår, noe som førte til lavere produksjon av strøm i vannkraftverkene (Energi Norge, 2022b). Alle disse faktorene førte til ekstreme strømpriser som svært få hadde forutsett. De fleste husstander hadde spotprisavtaler og kjente krisen på lommeboka, mens de fire prosent med fastpris kunne glede seg over et godt valg (Energi Norge, 2022a).

Vi ser altså at strømpriser er et veldig dagsaktuelt tema, og det å kunne forutse strømprisene kan medføre store besparelser. For eksempel kan man installere batterier i boliger som lades opp når strømprisene er lave, og bruke strømmen fra batteriene når prisene er høye. Prediksjon av

strømprisene vil også gjøre det lettere å regne på om strømsparende investeringer er lønnsomme å gjennomføre. Dette kan være alt fra installasjon av solcellepaneler, til det å isolere boligen på nytt for å spare kostnader fra oppvarming. Samfunnet vårt skifter over til å være mer avhengige av elektrisitet. Hele 65 prosent av nybilsalget i 2021 var elbiler (Rustad, 2021). Offentlig transport som busser og ferger skifter også gradvis over til elektrisitet framfor fossilt drivstoff. En stor fordel med prediksjon av strømprisene vil også være å kunne lade transportmidlene når prisene er lave.

1.2 Problemstilling

I denne oppgaven vil vi predikere strømpriser ved å bruke ulike typer nevrale nettverk. Ved å analysere treffsikkerheten til de ulike nettverkene vil vi finne den mest optimale nettverkskonfigurasjonen til å predikere strømprisene for vårt datamateriale. Vår problemstilling er da formulert som følgende:

“Prediksjon av strømpriser ved bruk av nevrale nettverk”

1.3 Avgrensninger

I denne oppgaven vil vi trene opp tre ulike former for nevrale nettverk til å predikere strømprisene i Midt-Norge. De ulike nettverksstrukturene er “feedforward neural network”, “recurrent neural network” og “long-short term memory”. Variablene vi mater inn i nettverkene er kull, gass, olje, uran, gjennomsnitts temperatur og nedbørsdata i Trondheim, og VIX-indeksen. Vi trener opp de nevrale nettverkene ved å bruke strømpriser fra starten av 2007 til utgangen av 2019, og tester prediksjonsevnen til nettverkene fra starten av 2020 til utgangen av 2021.

2. Teori

2.1 Maskinlæring

Maskinlæring er en betegnelse for kunstig intelligens hvor maskinen lærer over tid og finner mønster i store datamengder. Det kan brukes ved å sette opp en statistisk modell og benytte datasett hvor maskinen kan lære seg eller trene på å finne forskjellige mønstre. Dermed vil den lære seg selv opp, og når det kommer inn ny data vil maskinen bruke det man har lært til å predikere mønsteret i den nye dataen (Tidemann og Elster, 2022).

Selv hvor komplisert maskinlæring kan oppfattes, er det ikke noe nytt som har oppstått de siste årene. Allerede på 1940- og 50-tallet kom de første teoriene om maskinlæring. De første publiserte artiklene tilhører forskerne Alan Turing, Warren McCulloch, Walter Pitts og John von Neumann. Alan Turing stilte det viktige spørsmålet «Kan en datamaskin tenke?». Det første kunstige nevralt nettverk ble utviklet i 1958 av Frank Rosenblatt (*Artificial neural network*, 2022).

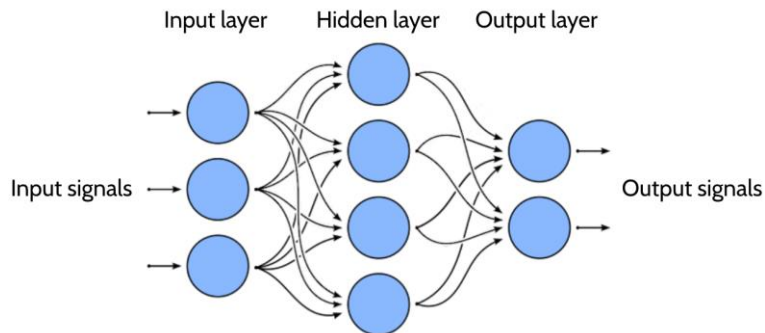
Det finnes forskjellige former for maskinlæring, en av de vanligste tilnærmingene er kunstige nevralt nettverk, som vi skal benytte oss av i denne oppgaven.

2.2 Nevrale nettverk

2.2.1 Kunstige nevralt nettverk

Et nevralt nettverk er en modell gitt som en representasjon av hvordan en menneskehjerne fungerer (Dvergsdal, 2019). En menneskehjerne er bygd opp av nerveceller som tar inn informasjon som man både lagrer og lærer av. I et nevralt nettverk vil nervecellene være representert som noder eller nevroner som skal ta inn et sett med data og lære av disse. Nevronene får en vekt eller form for verdi eller viktighet, og ved at man trener det nevralt nettverket vil disse nevronene lære seg hvordan de skal vektlegges for å finne de riktige mønstrene i et nytt datasett. Ved hjelp av skjulte lag mellom input- og output-laget vil disse nevronene koble sammen et nettverk der informasjon går inn og behandles av nevronene så de kan lage best mulig prediksjon. Disse skjulte lagene med nevroner har også ulike

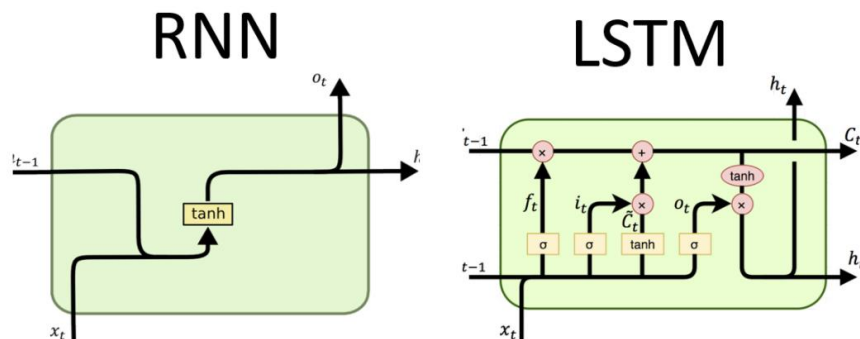
aktiveringsfunksjoner for å skille dataene som sendes ut. Slike nevralt nettverk kan brukes til klassifisering (gjenkjenning av bilder og tale) og prediksjon (predikere aksje- eller strømpriser).



Figur 1: Utforming av et nevralt nettverk

2.2.2 Recurrent Neural Network

RNN er en form for kunstig nevralt nettverk som ofte brukes når man skal benytte seg av data i form av en tidsserie. RNN kan i motsetning til vanlige nevralt nettverk benytte sekvensielle data og trenes til å huske historiske mønstre. Disse historiske mønstrene brukes deretter til å predikere framtidige data. De mest brukte aktiveringsfunksjonene i et RNN er sigmoid, tanh og relu (Saeed, 2021).



Figur 2: Illustrasjon av RNN og LSTM

2.2.3 Long Short-Term Memory

Long Short-Term Memory (LSTM) er en utvidelse av RNN. Forskjellen på de er at LSTM vil klare å ta vare på tidligere informasjon i større grad. Et vanlig RNN vil ikke kunne huske langsiktige avhengigheter i sekvensielle data. Ved å bruke LSTM vil man heller la dataene gå i en loop nevronene vil sende informasjon videre, glemme det eller sende de tilbake. Dermed klarer man å unngå de langsiktige avhengighetsproblemene vanlig RNN har.

2.3 Tidligere forskning rundt problemstillingen

Deepak Singhal og K.S. Swarup skrev artikkelen “Electricity price forecasting using artificial neural networks” (Deepak og K.S, 2011) i 2010. Singhal og Swarup ønsket å predikere strømpriser i nær framtid (short-term market) ved hjelp av nevralt nettverk. En av metodene de brukte var recurrent neural network. Her brukte de historiske strømpriser fra to måneder i forveien for Pennsylvania, New Jersey og Maryland som treningsdata. I tillegg brukte de forventet etterspørsel av strøm, tidligere etterspørsel etter strøm og drivstoffpriser som input. Modellen deres traff relativt bra på flere områder, men det var flere ting de merket seg som kritiske. De så at valg av input-variabler var veldig viktig. Ettersom strømprisene blir påvirket av veldig mange ulike variabler, så måtte de også ha i bakhodet at resultatene modellen førte til var et produkt av hvilke- og hvor mange variabler de valgte å ta med. De fant også ut at modellen hadde god treffsikkerhet på de dagene hvor man så en jevn utvikling i strømprisen. Store økninger eller nedganger i pris var komplisert for modellen å få med seg. Disse kjappe økningene eller nedgangene er ofte direkte knyttet opp til en økning eller nedgang i en annen variabel (gasspris, nedbørmengde, oljepris osv.). Hvis de ikke hadde med den spesifikke variabelen som økte/sank i modellen, så kom prediksjonen til å bli upresis. Singhal og Swarup mente at disse var de viktigste faktorene for endringer i strømprisen:

1. Tilbud og etterspørsel av strøm er ulik.
2. Volatilitet i drivstoffpriser.
3. Usikkerhet i etterspørselen etter strøm.
4. Usikkerhet i strømproduksjonen ved vannkraftverk. Påvirkes av nedbørmengde.
5. Feil ved maskineri og lignende som fører til stans i strømproduksjonen.

6. Problemer som gjør at all tilgjengelig strøm ikke kan bli tilbudt til kundene. For eksempel feil ved strømkabler til land man vil eksportere til.
7. Oppførselen til tilbydere og etterspørrere på strøm-markedet. Oppførselen kan knyttes direkte opp til aktørenes forventninger til fremtidig pris og etterspørsel.
8. Markedsmanipulasjon. Store aktører kan påvirke prisene gjennom for eksempel lovgivning, kjøpsavtaler og produksjonsavtaler.

Videre kan vi se på en artikkel skrevet av Carter Bouley i februar 2020 kalt “Recurrent Neural Networks for Electricity Price Prediction” (Bouley, 2020).

Her snakker Bouley om daglige, ukentlige og sesongbaserte mønstre i strømforbruket. Han velger å legge fokus på den daglige bruken. Her snakker han om at strømbruken er høy på morgenen når alle står opp, lavere når alle drar på jobb, høyere når alle kommer hjem fra jobb, og igjen lav når alle legger seg for å sove. Dette blir også reflektert gjennom strømprisene. Når etterspørselen etter strøm er høy er prisen høy, og når etterspørselen er lav så er prisen lav.

Bouley mener derfor at det kunne vært kostnadssparende for forbrukerne hvis man klarte å flytte strømforbruket sitt til tidspunkter i løpet av døgnet hvor både etterspørselen og prisen på strøm var lavere. Han ønsket å finne ut dette ved å predikere fremtidige strømpriser. Han hentet inn historiske priser for gass, olje, uran og kull. I tillegg hentet han inn historiske strømpriser og værdata. Han restrukturerte så dataen slik at han kunne se 168 timer tilbake i tid (1 uke), og 24 timer fram i tid. Bouley testet flere ulike nevralt nettverk som LSTM, GRU, ARIMA og RNN. Han konkluderte med at Recurrent Neural Network ga de beste resultatene.

I artikkelen blir det også nevnt flere interessante bruksområder for strømprediksjonen. For eksempel kan vaskemaskiner og tørketromler skru seg på når man ligger og sover. Enten ved hjelp av smartplugger som er tidsinnstilt, eller nyere hvitevarer som er koblet opp med Wi-Fi. Bouley prøvde også å simulere en situasjon med et batteri. Dette var produktet Powerwall 2 fra Tesla som er et populært batteri til bruk for husstander. Batteriet ble satt opp slik at det ble ladet når strømprisene ble predikert lavest i løpet av døgnet, og så at strømmen i batteriet ble solgt ut igjen i strømmettet når strømprisen ble predikert høyest. Det hadde resultert i en inntekt på 280

pund, og ladekostnader på 151 pund, for hele 2019. I såpass liten skala med kun ett batteri høres det kanskje ikke veldig lukrativt ut, men hvis man skalerer det opp så kan det være meget interessant. Et av de største problemene er at en Tesla Powerwall har utsalgspris på 5600 dollar, noe som helt sikkert kan virke avskrekkende for en del privatkunder. I tillegg er strømprisene veldig volatile, så både kostnadene og inntektene er usikre. Men, det er nettopp denne volatiliteten prediksjonene skal klare å utnytte for å skape profitt.

3. Data og metode

For å predikere strømpriser ved å bruke nevralt nettverk er vi nødt til å lage en modell ved å benytte data som det kan lages og knyttes mønster til. Siden vi skal predikere strømpriser ønsker vi å benytte variabler som kan knyttes opp mot strømproduksjon. Strøm i Europa produseres som regel ved hjelp av vannkraft, kull, gass, olje og uran. Dermed trenger vi variabler som kan representere disse produksjonskildene. I tillegg vil temperaturen kunne påvirke etterspørselen til strøm. Vi har valgt å benytte oss av en timebasert modell, og vi ønsker data fra starten av 2007 helt til slutten av 2021. For å gjøre analysen lettere å tolke vil vi benytte oss av samme valuta på alle økonomiske variabler, hvor vi har valgt priser i euro.

Vår avhengige variabel er strømpriser. Ved hjelp av Eikon sine databaser har vi hentet ut strømpriser fra Nordpool-Norway for region N03/Midt-Norge, som vil si det geografiske området som dekker Trøndelag og Møre og Romsdal. Via Eikon har vi også fått tak i alle 24 priser for hvert døgn, altså hver enkelt timespris. Dessverre har man ikke tilgjengelige data for alle helger og helligdager, så gjennom Python og Pandas har vi klargjort dataen ved hjelp av funksjoner som «df.reindex» og «df.fillna» har man etterfylt alle dager som manglet med den siste observerte verdien i datasettet. For lørdager og søndager blir de fylt opp med den siste prisen man hadde på fredagen. Da vi må ha strømprisen som både en avhengig og en uavhengig variabel for å bruke den i nettverket vårt har vi laget «laggede» verdier. Strømprisen har blitt skiftet 168 (24 x 7) ganger, for å muliggjøre at nettverket kan se en uke tilbake i tid. Følgelig fjernes den første uken med observasjoner fra den avhengige variabelen, slik at den får samme struktur som de uavhengige variablene. Vår avhengige variabel «Strompris» har 131 328 observasjoner i perioden 8. januar 2007 kl. 00:00, til 31. desember 2021 kl. 23:00. De laggede verdiene for strømprisene er nummerert med 1 til 168.

Gjennom Eikon sin database har vi hentet ut prisene for kull. Kullprisene er indeksen til ICE(Intercontinental Exchange) for dagsprisen til kull hentet ut av Richard Bay i Sør Afrika. Disse prisene ligger inne i dollar, og er blitt konvertert ved bruk av Python biblioteket “currency_converter”. Bibliotekets funksjon «CurrencyConverter» konverterer prisene til euro med den historiske valutakursen for hver observasjon. For å kjøre funksjonen gjennom datasettet mer effektivt brukes funksjonene “df.apply” og “lamda x”. Alle råvaredata vi har hentet ut har

kun dagspriser, så i likhet med dagene som manglet i strømprisene, har vi brukt «df.fillna» og «df.reindex» for å gi datasettet et timebasert oppsett og fylt de manglende verdiene med siste observerte verdi. Eksempelvis vil prisen 1. juli 2020 kl. 00:00 bli brukt til å fylle ut prisene for de resterende timene den dagen. Variablene for kull har fått navnet «Coal_Euro» og har 131 328 observasjoner.

Gassprisene er hentet fra Eikon sin database. Gassprisene er fra en nederlandsk indeks, og er hentet fra ICE. Prisene ligger inne i euro på refinitiv natural gas. Som nevnt er også gass registrert med døgnbaserte data, og “df.fillna” og “df.reindex” er benyttet for å få det ønskede timebaserte oppsettet som den uavhengige variabelen. Variabelen for gass har vi navngitt “Gas_Euro”.

Den uavhengige variabelen Olje har vi gitt navnet “Oil_Euro”. Dataen er oppgitt i dollar, fra Federal Reserve Economic Data. For å få dataen omgitt til riktig valuta brukes “CurrencyConverter” med samme fremgangsmåte som beskrevet ovenfor. Da dataen er dagsbasert brukes funksjonene “df.fillna” og “df.reindex” for å kunne fylle inn alle manglende data og for å få et 24-timers oppsett.

Det viste seg å være krevende å få hentet ut fullstendig datamateriale over hele perioden for prisene av uran. Vi fikk hentet ut døgnbaserte uran priser fra 01.07.2007 til 31.08.2021 gjennom NYMEX. Og vi benyttet gjennomsnittsverdiene fra Trading Economics for de manglende månedene, for å få et mer nøyaktig datamateriale enn om det hadde vært unnlatt. Prisene ble hentet ut i dollar og vi benyttet derfor funksjonen “CurrencyConverter” til å konvertere til euro. Deretter reindexerte vi datamaterialet til en 24-timers syklus og etterfylte timene med siste gyldige observasjon.

Temperatur- og nedbørsdataen er hentet fra “seklima.no”. Vi har gitt variablene navnene “Temperatur” og “Rainfall”. Temperaturdataen er timebasert og har derfor ønsket mengde datapunkter. Nedbørsdatamaterialet er døgnbasert og må reindexeres til et 24-timers oppsett. Verdiene som manglet er for begge fylt inn ved hjelp av funksjonen “df.fillna”, som gir de manglende datapunktene verdien fra forrige gyldige data. Værstasjonen vi har valgt er på Voll i Trondheim. Bakgrunnen for dette er at dette var en av få værstasjoner som har registrert data sammenhengende gjennom hele den ønskede perioden.

Vi har benyttet VIX-indeksen for å få med en uavhengig variabel for markedets risiko over perioden. Vix-indeksen gir en indikasjon på volatilitet i S&P 500-indeksen, som vi tenker i en viss grad kan overfører til europeiske og norske forhold. For å gjøre klar datamaterialet til analysen fjernet vi alle kolonner med unntak av “Adj Close”. Observasjonene var døgnbaserte slik at vi var nødt til å fylle inn de manglende timene med siste gyldige observasjon som forklart tidligere.

Til slutt klargjorde vi datamaterialet til analyse ved å samle alle de uavhengige variablene i en dataframe kalt “dfX” og strømprisen i en dataframe vi kalte “dfy”. Ved å reindeksere og etterfylle manglende data fikk alle variablene en lik struktur med 131 328 observasjoner over perioden 8. januar 2007 kl. 00:00, til 31. desember 2021 kl. 23:00. Det nevrale nettverket vil prøve å fange opp sammenhenger i datamaterialet ved å gå gjennom en og en observasjon og justere vektene for hvor godt prediksjonen treffer sammenlignet med den avhengige variabelen. Det at vårt datamateriale ikke har noen manglende data og innehar nøyaktig like mange observasjoner, vil kunne bidra til å lage bedre nevrale nettverk som er i stand til å lage mer nøyaktige prediksjoner av strømprisene.

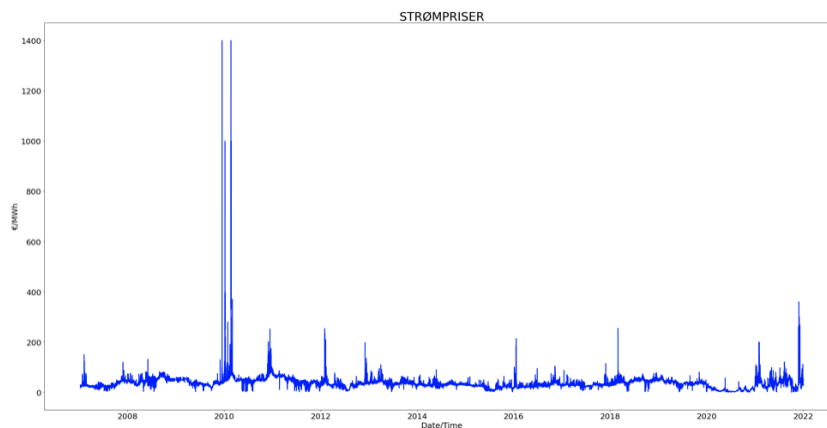
Vårt valg av datamateriale vil påvirke i hvilken grad funnene i vår analyse kan overføres til andre forhold. Prediksjonevnen til de nevrale nettverkene er avgrenset geografisk til strømprisene i Midt-Norge, og tidsmessig til perioden fra 2007 til 2021. Bruksområdet til våre nettverk kan da ikke uten videre overføres til andre forhold enn disse. Allikevel vil funnene i analysen kunne være av interesse for hvordan nevrale nettverk på generelt grunnlag kan bygges opp og anvendes på lokale forhold ved bruk av relevante variabler og et oppdatert datamateriale.

4. Analyse

4.1 Visualisering av variabler

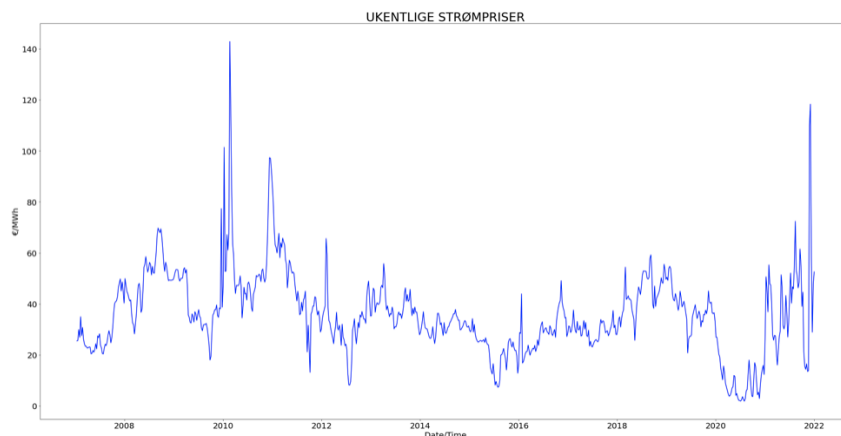
4.1.1 Strømpriser

I vår oppgave hvor vi ønsker å predikere strømpriser er det viktig å ha litt innblikk i hvordan ikke bare strømprisene, men også prisene/verdiene til våre andre valgte variabler endrer seg over tid. Ved hjelp av Python får vi plottet opp hvordan prisene forandrer seg over tid og vi kan også lett få fram forskjellige statistiske data som gjennomsnitt, standardavvik osv. Strømprisene vist i figur 3 kan være vanskelig å tolke nøye, mye på grunn av noen store spikes i strømmprisen rundt år 2010. Her er det også greit å vite hvorfor strømmprisen i Midt-Norge hoppet ekstremt opp rundt 2010. Grunnen i dette tilfellet var en kombinasjon av bitende kulde, lav beholdning i vannmagasinene, brudd på strømkablene mellom Norge og Nederland og at en del svenske kjernekraftverk var ute av drift (NRK, 2010).



Figur 3: Strømpriser visualisert i timer

Ved å benytte oss av ukentlige strømpriser vil det være lettere å se mønstrene i strømmprisen. Som man ser på figur 4 så går prisene sesongbasert opp og ned og danner en form for årlig syklus.



Figur 4: Strømpriser visualisert i uker

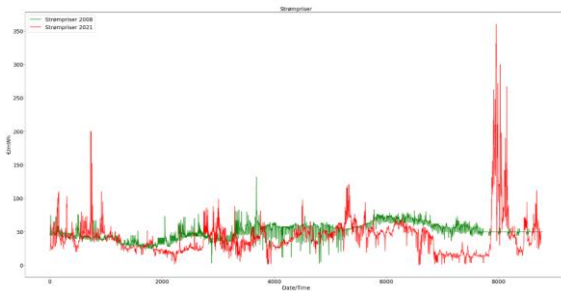
Grunnen her er at strømprisene er som regel lavere om sommeren enn den er i vintermånedene. Man ser fortsatt tydelig spiken som oppstår i 2010, men man ser også en voldsom høy pris i slutten av 2021. Her gikk prisen fra å være under 20 €/MWh, til å bli så høy som 120 €/MWh. I tillegg var den høyeste timesprisen da 360 €/MWh, noe som er usedvanlig høyt for denne regionen. Mangel på strømproduksjon gjennom året på grunn av tørt og vindstille vær i store perioder er en stor grunn til at strømprisen vokste så mye mot slutten av 2021. Som man ser i tabell 1 er det kun tre år hvor strømprisen har oversteget 300 €/mWh, noe som tilsier at det må ligge noe spesielt bak disse høye prisene. Man kan også se en tendens til at jo større spikes et år har jo høyere standardavvik vil man ha det året.

Faktisk kom strømprisen i minus i løpet av 2021, noe som ville tilsi at forbrukerne fikk betalt for å bruke strøm, om man ser bort fra nettleie, påslag og ulike avgifter (Bach og Høgseth, 2021). Dette skyldes masse vind i Nederland, Tyskland og Danmark, og ved hjelp av kablene Norge har til disse landene ble det et stort energioverskudd i noen timer. Denne prisen var midt på natta i en helg, så man fikk dessverre ikke altfor stor glede av den lave prisen på -0,01 øre, spesielt når prisen per kilowatt time var oppe i nesten 1 krone igjen neste dag.

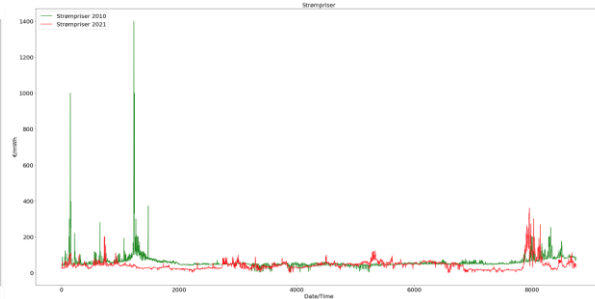
Tabell 1: Deskriptiv statistikk av årlige strømpriser

Strømpriser	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Gjennomsnitt	29,47	50,32	39,33	58,14	47,18	31,49	38,92	31,53	21,26	28,72	29,49	44,14	38,48	9,47	41,11
Standardavvik	9,83	11,82	24,34	45,17	15,45	13,83	7,79	5,25	7,77	10,45	5,56	10,38	7,88	6,98	27,15
Minimum	4,09	2,39	1,78	1,82	1,45	3,92	1,38	9,95	1,14	11,28	5,19	1,92	1,38	0	-0,01
Maksimum	150,3	131,73	1400,1	1400,11	93,11	253,92	109,55	89,97	63,99	214,25	114,7	255,02	80,75	57	360

Strømprisen var stort sett ganske lav i 2021 og på perioder om sommeren var den nesten gratis. Likevel ser man i figur 5 og 6 at spiken som oppstod i 2021 er ganske stor sammenlignet med den vanlige strømprisen i et år som 2008, men den er mye mindre sammenlignet opp mot spikene i 2010.



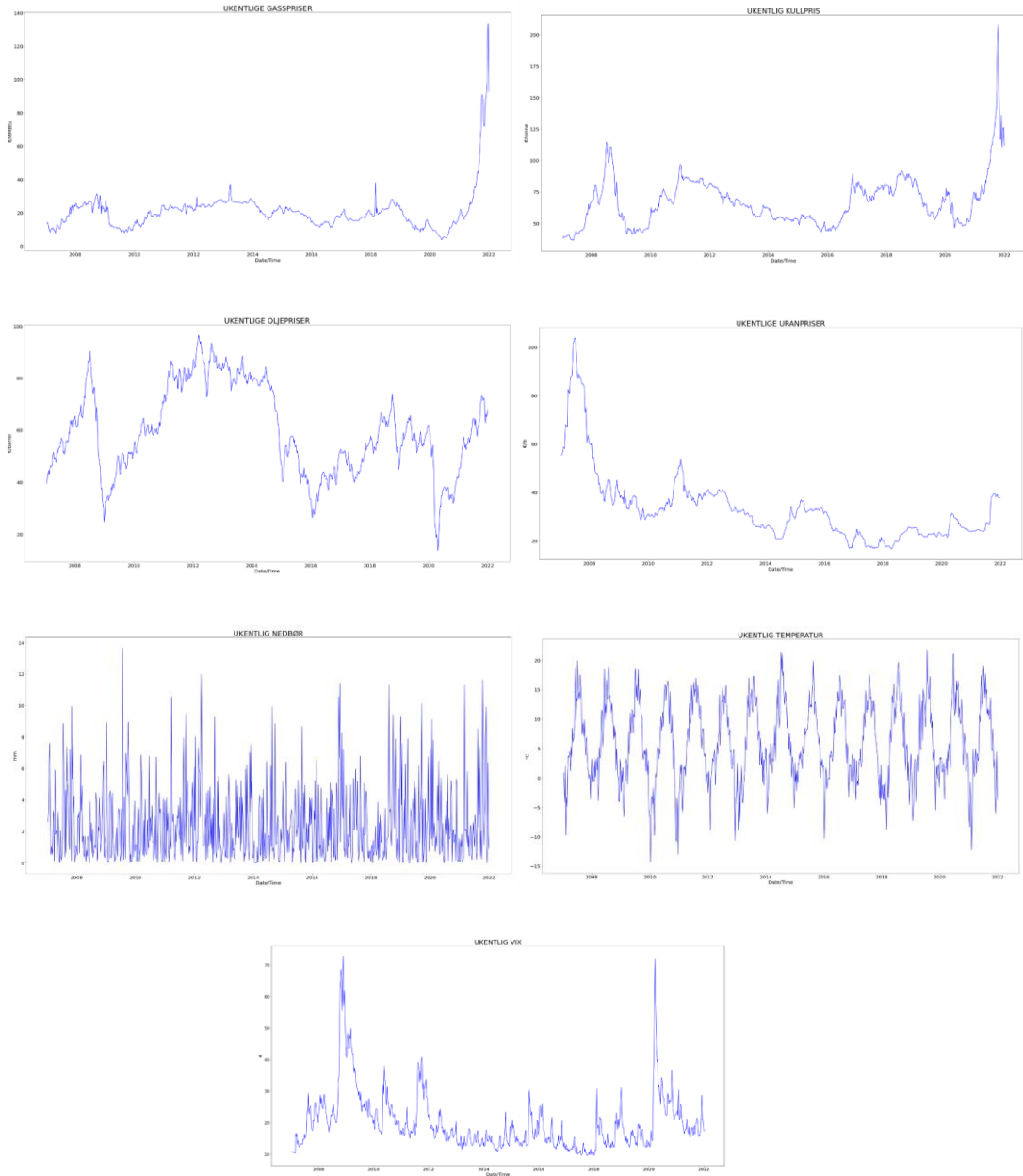
Figur 5: Sammenligning av strømpriser fra 2008 og 2021



Figur 6: Sammenligning av strømpriser fra 2010 og 2021

4.1.2 De uavhengige variablene

Det nevrale nettverket skal bestå av strømprisene og sju andre variabler som vi tenker kan være med å påvirke strømprisen. Nedenfor er de ukentlige prisene til disse sju variablene illustrert i figur 7 med hvert sitt diagram. De blir illustrert i ukentlige priser for å gi et bedre bilde av helheten til prissvingningene eller andre variasjoner i datamateriale. I likhet med strømprisen har også flere av disse en spike mot slutten av 2021. Strømmen i Europa produseres i stor grad av gass og kull. Vinteren 2020 var veldig kald, noe som gjorde at man måtte bruke opp mye av strømmen de hadde lagret. Dermed ble disse råvarene mye dyrere når etterspørselen økte voldsomt utover i året. Lite hjalp det også at land som Russland eksporterte mindre og at man møtte konkurranse fra Asia (Hovland, 2021).



Figur 7: De uavhengige variablene i hvert sitt diagram

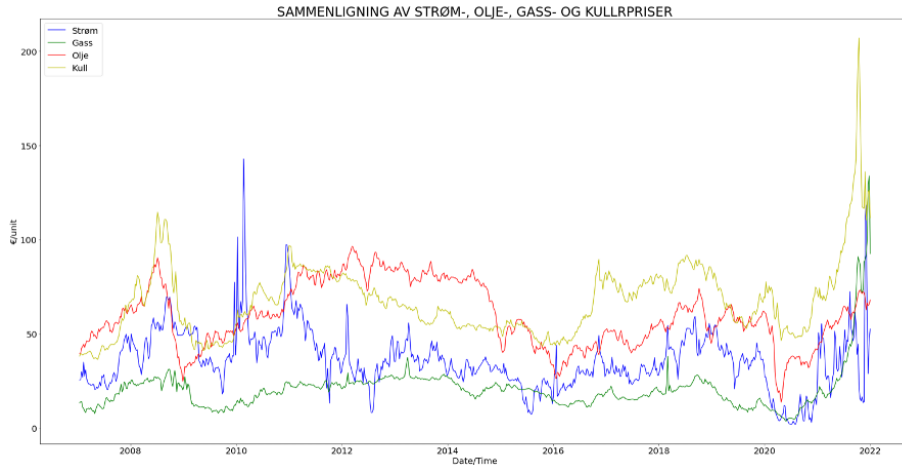
4.1.3 Forhold mellom strømpriser og uavhengige variabler

Prisene til strøm og råvarer varierer mye likt, noe som kan være positivt når man skal finne variabler til et nevralt nettverk. Ved å måle korrelasjonen til de ulike variablene kan man bedre fastsette hvilke som samhandler likt. Vi velger å benytte oss av de valgte variablene i våre modeller uansett hvilken korrelasjon de har.

Tabell 2: Korrelasjonsmatrise

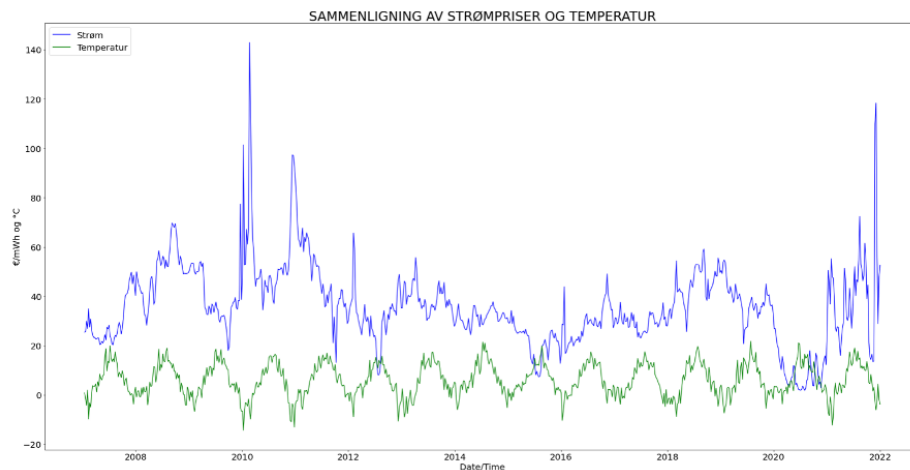
	Strøm	Gass	Olje	VIX	Kull	Uran	Temperatur	Nedbør
Strøm	1	0,21	0,24	0,07	0,3	0,04	-0,19	-0,04
Gass	0,21	1	0,39	-0,05	0,7	0,04	-0,07	0,04
Olje	0,24	0,39	1	-0,29	0,38	0,09	0,07	0,02
VIX	0,07	-0,05	-0,29	1	-0,05	0,15	-0,06	0,01
Kull	0,3	0,7	0,38	-0,05	1	-0,15	-0,01	0,05
Uran	0,04	0,04	0,09	0,15	-0,15	1	-0,002	0,02
Temperatur	-0,19	-0,07	0,07	-0,06	-0,01	-0,002	1	-0,01
Nedbør	-0,04	0,04	0,02	0,01	0,05	0,02	-0,01	1

Hovedsakelig er man opptatt av hvordan de valgte variablene korrelerer med strømprisen, men gjennom en korrelasjonsmatrise, vist i tabell 2, kan man lett se hvordan alle korrelerer med hverandre. Mønsteret til strømprisen korrelerer best med kull, olje, gass og temperatur. Det vil ikke si at de nødvendigvis påvirker strømprisen, men det kan antyde at om strømprisen øker kan de resterende også gjøre det. Ved å plote disse sammen i et diagram kan man lettere se hvordan de korrelerer, altså om svingningene til grafene er lik i noen grad. Dette er vist i figur 8.



Figur 8: Sammenligning av strøm-, olje-, gass- og kullpriser

De to variablene som korrelerer mest, er kull og gass. Disse er som nevnt de råvarene som brukes mest i Europas strømproduksjon, så det gir mening at de vil følge et likt mønster. Kull er dyrere enn gass, fordi kull er med å produsere 40 prosent av strømmen man bruker rundt om i verden, så det er den klart viktigste energibæreren for kraftproduksjon. Gass brukes til litt mer enn bare strøm, blant annet brukes det gass til å lage gjødsel, så prisen blir påvirket av etterspørsel fra flere fronter. Likevel fikk begge kjenne på en hard prisoppgang i slutten av 2021 slik som strømprisen. Strømmen er klart den prisen som varierer mest. I motsetning til råvarene så er strømmen styrt i stor grad av etterspørsel til forbrukerne, altså alle som bruker strøm. Strømbruken varierer ofte med tanke på sesong og tid på døgnet. Prisene til kull, gass og olje varierer ikke stort fra natt til dag, siden det handles over hele verden til alle tidspunkt.



Figur 9: Sammenligning av strømpriser og temperatur

Som tidligere nevnt korrelerer strømmen og temperatur slik at når temperaturen er lav, så vil strømprisen være høy, og motsatt. Dette gjenspeiles i figur 9. Noen unntak finnes, men de høyeste strømprisene finner man om vinteren når temperaturen er lav. For at strømprisen ikke skal bli så høy når det er kaldt, så er man avhengig av hvordan været har vært før

Disse variablene som korrelerer høyt med strømmen vil være til god nytte til de nevrale nettverkene, fordi de skjulte lagene vil være med å bestemme hvordan mønstrene til variablene skal vektlegges. Dette gjør at nettverket kan gjenkjenne mønster i nye data som sendes inn i nettverket og lager en bedre prediksjon.

4.2 Utforming av nevrale nettverk

4.2.1 Beskrivelse av nettverkene

For å kunne predikere verdier fremover i tid må vi benytte oss av et recurrent neural network (RNN eller LSTM). Likevel vil vi også ta i bruk et artificial neural network for å sammenligne forskjellene. Begge typer nettverk skal kunne gi oss en strømpris, men forskjellen er at ved å bruke gårsdagens data vil et recurrent neural network gi dagens strømpris, men et artificial neural network vil gi oss gårsdagens strømpris. Likt for alle modellene våre så splitter vi datasettet for de 12 første årene til treningsdata og lar de 3 siste årene få være test data. Dermed får man trent seg på store deler av dataene, i tillegg til de store spikene rundt starten av 2010, som kanskje kan hjelpe nettverket til å gjenkjenne mønster mot de høye prisene i 2021. Et nevral nettverk kan ha alt fra et til uendelig mange skjulte lag med forskjellige antall noder i hvert lag. Hvert lag har også en aktiverings funksjon som sier noe om hva slags data som sendes ut fra laget. Når vi har lagd modellene våre har vi forsøkt mange kombinasjoner, med et mål om å få en prediksjon så nærme de faktiske strømprisene.

4.2.2 ANN

Oppbyggingen av et ANN er simplere enn ved et RNN. Dette gjør derimot at vi ikke får predikert priser fremover i tid, men vi kan predikere prisene for en dag gitt prisene til de andre variablene for denne dagen. For å predikere en best mulig strømpris med ANN benytter vi datasettet med 168 laggede strømpriser. Dermed får man se innblikk i den siste ukens strømpriser som vil gjøre prediksjonen lettere for denne typen modell. I oppbyggingen av vår

ANN-modell endte vi opp med 3 skjulte lag med 30 noder i hver, og med 'tanh' aktiveringsfunksjon.

4.2.3 RNN og LSTM

Når man skal sende data inn til et RNN eller LSTM nettverk, så må man gå fra et todimensjonalt til et tredimensjonalt datasett. Noe som vil si at man må spesifisere antall observasjoner, antall tidsserier og antall variabler. For å gjøre veien videre lettere har vi videre tatt vekk de laggede verdiene for strømprisene. I tillegg må denne tredimensjonale dataen som spyttes inn i nettverket være formet slik at man skal kunne gi oss strømprisen for den neste dagen. Dermed må vi gjøre om dataen fra time series til et supervised learning problem, altså gjøre om så dagens input er gårsdagens verdier. Et slikt nettverk vil da ta vare på dataen man predikerer og benytte den til å predikere neste dag igjen. Videre vil vi teste ut dette datasettet på en LSTM modell med kun ett skjult lag, en SimpleRNN modell med flere skjulte lag og til slutt en Stacked LSTM modell. Når man i vårt tilfelle skal håndtere masse datapunkter, så har vektleggingen til de skjulte lagene og nodene mye å si for resultatet. Etter å ha trent modellen opp med forskjellige antall noder i lagene, fant vi til slutt et best mulig resultat når vi definerte modellene slik:

1. Ett skjult LSTM lag med 110 noder og trent den over 50 epoker.
2. Tre skjulte SimpleRNN lag med 30 noder i hver og trent den over 50 epoker.
3. Fire skjulte LSTM lag med henholdsvis 10, 50, 50 og 10 noder som blir trent over 50 epoker.

4.3 Predikerte resultater av nevralt nettverk

4.3.1 Beskrivelse

Datasettet ble delt opp i en treningsdel, som har blitt brukt til å trene opp modellene våre, og en testdel som modellen skal bruke for å predikere riktige strømpriser. I tillegg vil vi se hvordan nettverkene våre håndterer å predikere strømprisen for den første timen i 2022. For å gjøre det må man gi nettverkene et datasett som inkluderer timene fram til den siste timen av 2021, så denne timen vil bli brukt som den siste inputen. Her vil vi prøve oss fram med både de 24 siste

timene av 2021, men også hele datasettet fra 2007 og fram til slutten av 2021, for å se om det vil bli store forskjeller i prediksjonene.

Tidligere har vi sett på hvordan prisene ble usedvanlig store mot slutten av 2021, så vi vil også sammenligne hvordan modellene våre klarer å predikere de høye pristoppene. Det man burde se etter er om modellen er proaktive eller reaktive når man ser at prisen begynner å stige.

For å bedømme hvor gode modellene våre vil vi benytte oss av et par måleenheter. Den ene vi benytter er måleenheten RMSE (root mean squared error) eller på norsk, roten av den kvadratiske feilen (Zach, 2021).

$$RMSE = \sqrt{\sum \frac{(\hat{y}_i - y_i)^2}{n}}$$

Der \hat{y} er den predikert verdien og y er den faktiske verdien for periode i og n er antall observasjoner. Denne verdien sier oss hvor langt vår predikerte verdi i gjennomsnitt er fra den faktiske verdien, så om vi får en RMSE på 10, vil våre predikerte strømpris være i gjennomsnitt 10 € unna hver faktiske strømpris. Målet vårt er dermed å få så lav RMSE som mulig for hver enkelt modell.

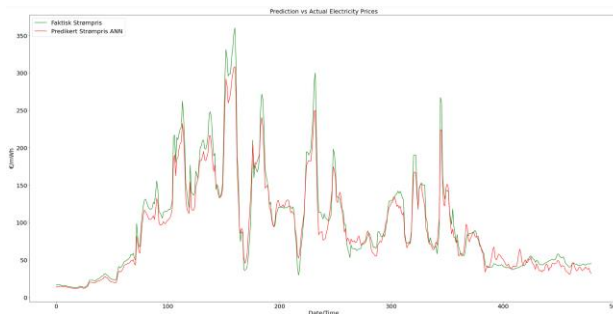
Den andre måleenheten vi benytter oss av er R-squared.

$$R^2 = 1 - \frac{\text{Uforklart varians}}{\text{Total varians}}$$

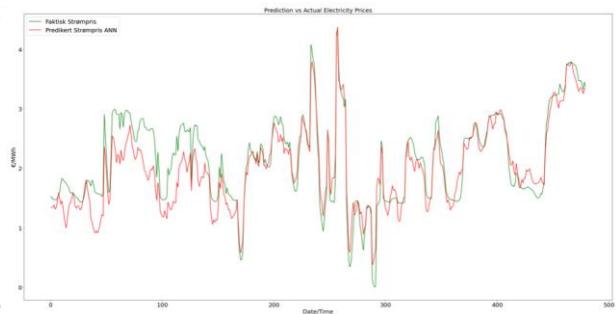
Denne verdien står for hvor mye av variansen til den avhengige variabelen som kan forklares av variansen til de uavhengige variablene (Fernando, 2021). Om vi oppnår en R-squared på 100% i våre modeller, så vil det si at all variansen til strømprisen kan forklares av variansen til olje-, kull og gasspriser osv. Målet er dermed å oppnå en R-squared så nærme 100% for testdataen vår i alle modellene.

4.3.2 ANN

ANN-nettverket er god til å følge strømprisen når det er lite variasjon. I figurene under sammenligner vi den predikerte strømprisen med den faktiske strømprisen for perioden fra 20. November 2021 til 10. Desember 2021, og fra 23. juni 2020 til 13. juli 2020. Under disse periodene har vi først en voldsom prisstigning, og i den andre hvor prisen faktisk så vidt ble negativ.



Figur 10: ANN resultater for høye strømpriser

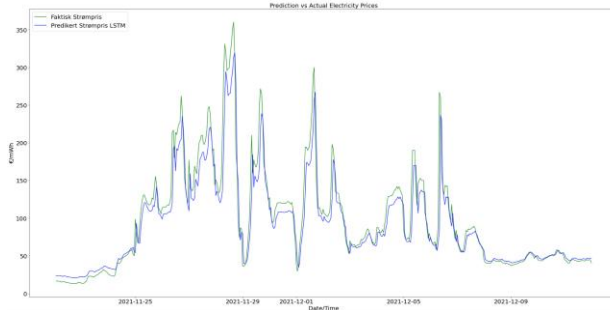


Figur 11: ANN resultater for lave strømpriser

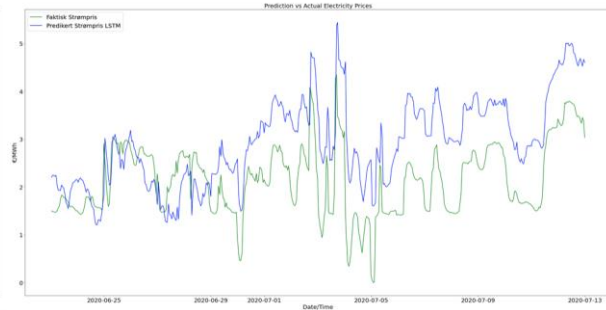
ANN modellen har en RMSE på 3,994 og en R-squared på 96,7%. I figur 10 og 11 sammenligner man verdiene til de predikerte strømprisene mot de faktiske. Nettverket klarer å følge mønsteret til den faktiske strømprisen i stor grad. Den klarer ikke helt å nå topp- og bunnpunktene, så det er nok der man får de største feilene.

4.3.3 LSTM

LSTM-nettverket har en annen utforming og tar inn et tredimensjonalt datasett, kontra et todimensjonalt datasett som ANN bruker. Dette gjør da at treningstiden øker i større grad, og at man må bruke rundt 6 minutter på å trene seg opp.



Figur 12: LSTM resultater for høye strømpriser

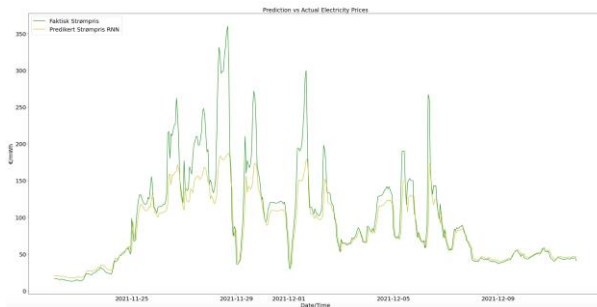


Figur 13: LSTM resultater for lave strømpriser

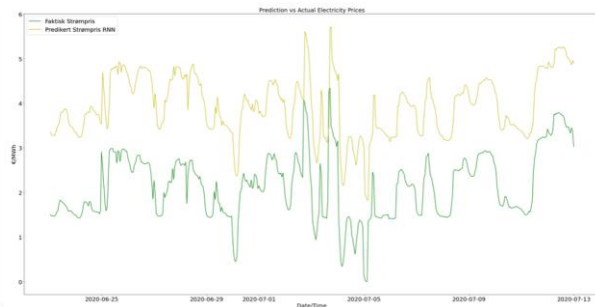
LSTM modellen har en RMSE på 4,408 og en R-squared på 96%. For den første timen i 2022 predikerer den at strømprisen vil være 32,481 €/MWh når vi benytter kun de 24 siste timene av 2021. Benytter vi hele datasettet vårt predikerer den at strømprisen vil være 35,879 €/MWh. I figur 12 og 13 sammenligner man LSTM-verdiene med de faktiske strømprisene i periodene med ekstreme høye- og lave strømpriser, så ser man at mønstrene samsvarer i stor grad. Når strømprisen ble negativ slet den med å tilpasse seg de lave verdiene, så den får større glipp enn ellers.

4.3.4 RNN

RNN-nettverket benytter også det tredimensjonale datasettet, og bruker omtrent samme tid som LSTM-nettverket på å trene seg opp.



Figur 14: RNN resultater for høye strømpriser

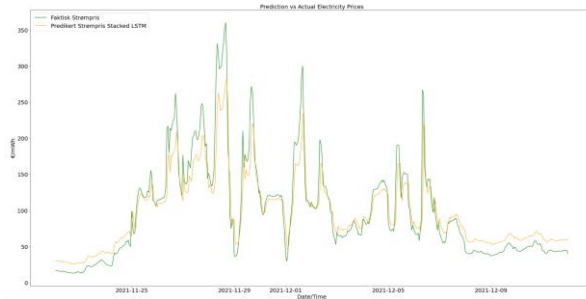


Figur 15: RNN resultater for lave strømpriser

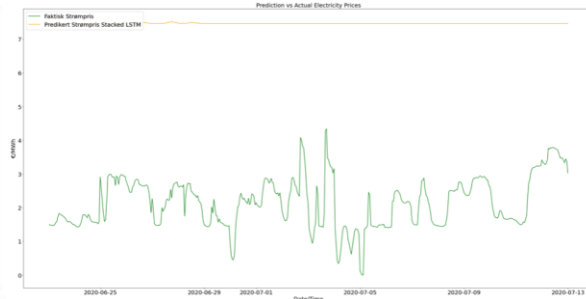
RNN modellen har en RMSE på 5,48 og en R-squared på 93,8%. For den første timen i 2022 predikerer den at strømprisen vil være 32,469 €/MWh når vi benytter kun de 24 siste timene av 2021. Benytter vi hele datasettet vårt predikerer den at strømprisen vil være 34,209 €/MWh. Nettverket klarer å danne et nesten eksakt mønster for de lave strømprisene, men den ligger konstant rundt to euro over for den faktiske strømprisen, som vist i figur 13. Heller ikke klarer den å nå de større toppene i figur 14 når strømprisen blir høy, men den er god på å følge strømprisen når den har lavere variasjoner.

4.3.5 Stacked LSTM

Stacked LSTM-nettverket er en utvidelse av LSTM-nettverket, men med forskjellige antall lag og noder. Dette nettverket er litt mer komplisert enn de tre andre, og bruker rundt 8 minutter på å trene seg opp.



Figur 16: Stacked LSTM resultater for høye strømpriser



Figur 17: Stacked LSTM resultater for lave strømpriser

Stacked LSTM-modellen har en RMSE på 6,198 og en R-squared på 92,1%. For den første timen i 2022 predikerer den at strømprisen vil være 32,53 €/MWh når vi benytter kun de 24 siste timene av 2021. Benytter vi hele datasettet vårt predikerer den at strømprisen vil være 46,35 €/MWh. Treffsikkerheten til Stacked LSTM i perioden med ekstremt lave priser er veldig dårlig. Figur 16 viser at den ikke har klart å danne seg et mønster, og gir en så og si konstant pris på litt over 7 €/MWh. Heldigvis predikerer den ikke konstant en rett linje hele tiden, og på figur 17 klarer man i større grad å følge svingningene til de høye prisene. Problemet er at man ikke når helt opp i toppene, og legger seg litt over prisen når den svinger mindre.

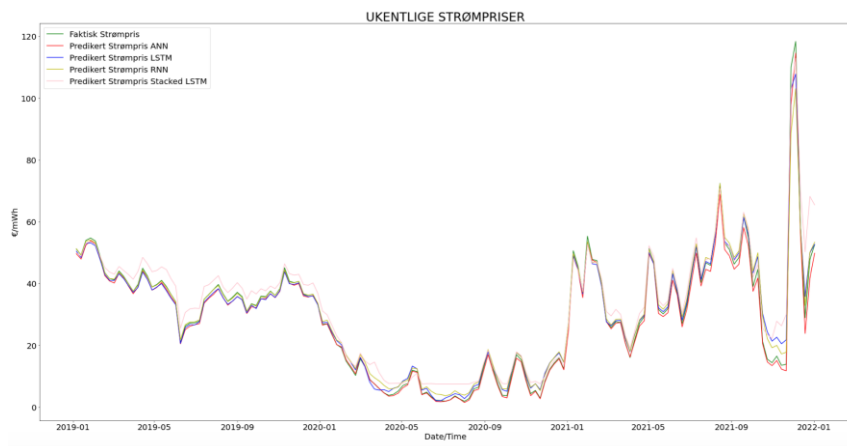
5. Diskusjon

5.1 Sammenligning og diskusjon av predikerte resultater

Nettverkene klarte i stor grad å tilpasse seg dataene våre. Resultatet varierte i stor grad da man skulle finne det mest optimale antall lag og noder. Av alle nettverkene var det ANN som hadde den laveste RMSE-scoren og den høyeste R-squared. Grunnen er nok at det fikk et mye større datasett å trene seg på, hvor hver time hadde kunnskap om den siste ukens strømpriser. Av nettverkene som kan bli brukt til framtidige prediksjoner, så fant vi at LSTM nettverket ga det beste resultatet. Når man sammenligner de forskjellige RMSE-scorene så skiller det ikke mer enn 2 mellom LSTM og Stacked LSTM som gjorde det dårligst.

Videre vil vi gjøre om €/MWh til øre/kWh for å bruke det mer i norske forbrukertilfeller. Vi benytter oss av valutakursen i årsskifte 2021/2022 som lå på omtrent 10 NOK/1 EUR (*Euro valutakurser - EUR til NOK, u.å.*). Ved kjapp omregning vil prisene bli gjort om eksempelvis slik: 40 €/MWh → 40 øre/kWh. Dermed kan det sies at alle nettverkene våre klarer å predikere strømpriser med en gjennomsnittlig feilmargin mellom omtrent 4 til 6 øre/kWh.

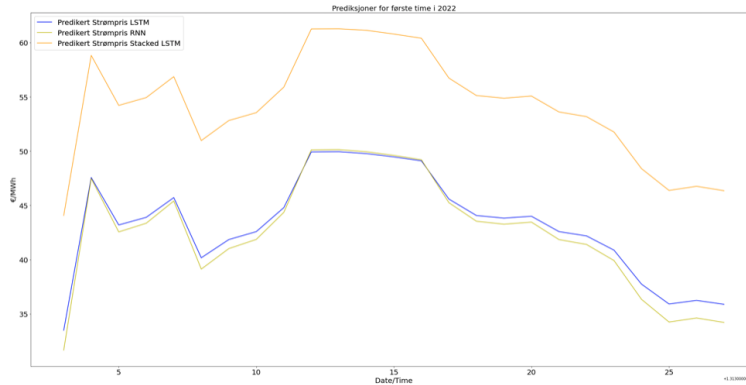
R-squared verdiene til nettverkene er alle over 90%, der de to beste er ANN og LSTM. Dette vil si at variablene vi har valgt å benytte oss av er med å forklare over 90% av variansen til strømprisene. Sammenligner man de predikerte prisene opp mot den faktiske prisen så vil de fleste prediksjonene følge svingningene til strømprisen til en viss grad. De henger som regel en time bak, og heller mer mot et reaktivt enn proaktivt mønster. Gjør man om timeprisene til ukespriser igjen vil man i figur 18 se de predikerte strømprisene følger den faktiske strømprisen veldig bra. Den eneste som skiller seg ut i større grad er Stacked LSTM som også hadde laveste R-squared.



Figur 18: Sammenligning av ukentlige prediksjoner

Det som mulig er det mest spennende med oppgaven er hvordan de nevrале nettverkene klarer å predikere framtidige strømpriser. Vi ønsket å predikere den første timen etter datasettet vårt. Dette har vi gjort på to forskjellige måter for hvert fremtidspredikerende nettverk. Først tok vi i bruk kun de siste 24 timene av 2021, så hele datasettet vårt. LSTM-nettverket ga oss priser på 32,481 og 35,879 €/MWh. RNN-nettverket predikerte priser på 32,469 og 34,209 €/MWh. Til slutt har vi Stacked LSTM-nettverket, som igjen skiller seg mest ut, den predikerer priser på 32,53 og 46,35 €/MWh. Ved å bruke få datapunkt gir de så og si den samme prediksjonen. Det skilte ikke mye mellom LSTM og RNN, når man benyttet seg av et stort datamateriale.

For å sammenligne hvilket som gjorde det best, så må man se opp mot den faktiske prisen denne timen. Faktisk strømpris for den første timen av 2022 var 46,60 €/MWh (Nord Pool, u.å.). Strømprisen fikk et hopp på årsskiftet, noe nettverkene kanskje ikke var forberedt på. Det er ikke lett å si hvorfor strømprisene i Midt-Norge hoppet opp litt ved årsskifte, når den gikk ned på Sør- og Østlandet (Høgseth og Svendsen og Hovland, 2022). Likevel er prisene for Midt- og Nord-Norge mye lavere enn resten av landet, siden disse som nevnt tidligere kontrolleres i mindre grad av det vest-europeiske kraftmarkedet. I tillegg, så var det et nettverk som klarte å predikere nesten korrekt pris. Stacked LSTM bommet med 0,25 € fra den faktiske strømprisen. Studerer man grafene av prediksjonene de nevrале nettverkene gjorde i figur 19, så ligger Stacked LSTM nettverket sin graf så og si parallellt over de to andre.



Figur 19: Prediksjon av den første timen av 2022

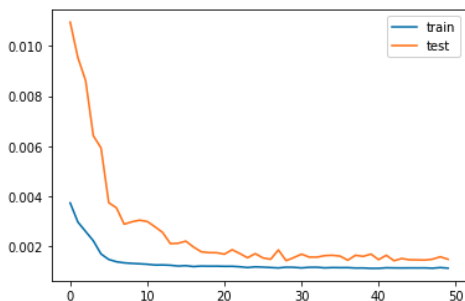
Sammenligner man de tidligere verdiene så har jo LSTM-, og RNN-nettverket prediksjoner som samsvarer i større grad med de faktiske strømprisene for det siste døgnet av 2021. Det er ikke mye som tilsier at prisen plutselig skulle gå opp, og tidligere har man sett at nettverkene tilpasser seg i større grad reaktivt, så det gir heller mindre indikasjon på at noen av nettverkene faktisk kunne klare å forutse prisendringen. Dermed kan det påstås at det var i større grad flaks at Stacked LSTM-nettverket klarte å predikere bedre enn de to andre.

5.2 Evaluering og forbedring av modellene

De nevrale nettverkene vi har formet er av den simple sorten når det kommer til å predikere fremtidige verdier. Her klarer vi bare å “se” en time fremover i tid, hvor i realiteten skulle det være mer ønskelig å klare muligens en dag eller en uke fremover i tid. Skulle dette være mulig måtte vi ha omformet det tredimensjonale datasettet slik at vi bruker 24 eller 168 timesteps, i stedet for kun 1 som vi brukte. Dette kunne bli gjort sånn som Carter Bouley har gjort i en lignende analyse han har gjort (Bouley, 2020). Det ble forsøkt også med datasettet vårt, men prosessen ble for tung for datamaskinen og det ga heller ingen gode resultater når vi prøvde å sette det inn i et nevralt nettverk. Treningstiden ble også mye lengre, i motsetning til nettverkene vi valgte som alle brukte under 10 minutter hver, så ville vi opplevd å bruke 30 minutter eller mer på hvert nettverk. Dette ble for voldsomt når man skal teste nettverket flere ganger i håp om å finne antallet lag og noder som gir de mest optimale verdiene.

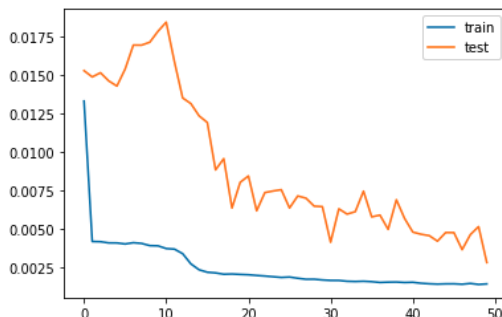
Noe som er viktig når man skal benytte nevrale nettverk, er at de er tilpasset godt både trenings- og testdatasettet. Eventuelle problemer som kan oppstå er at nettverkene er underfit eller overfit

som det kalles. Et underfit nettverk er et nettverk som tilpasser seg treningsdataen godt, men ikke testdataen, og et overfit nettverk det motsatte (Brownlee, 2017b). Ved testing av de nevrale nettverkene møtte man ofte på underfitting. Dette ble løst med å øke antallet noder, eller øke både noder og lag.



Figur 20: Trening av LSTM

Figur 20 viser hvordan trenings- og testdatasettet tilpasset seg LSTM-nettverket. Et nettverk som er godt tilpasset, skal i dette tilfellet ha en oransje linje som klarer å legge seg perfekt på den blå linjen. I tilfellet med LSTM-nettverket fikk vi det beste resultatet når vi benyttet 110 noder, og linjene la seg nesten sammen. Et nettverk som, etter mye trening, fortsatt lider av underfitting er Stacked LSTM.



Figur 21: Trening av Stacked LSTM

Dette nettverket slet med å tilpasse seg lenge, og det ble forsøkt med mange forskjellige kombinasjoner av noder, lag og aktiveringsfunksjoner. Her kunne det blitt benyttet flere epoker under treningen som vist i figur 21, så kunne man muligens sett at testlinjen hadde nærmet seg enda mer.

I vår oppgave ble datasettet delt opp i et sett for trening og et annet for testing. Alternativet kunne ha vært å ha et tredje sett og, et for validering. Dermed kunne man brukt de 9 første årene til trening, de 3 etter der for validering og til slutt de 3 siste årene for testing. Fordelen ved å

bruke et sett for validering, er at man kunne benyttet det når man trente opp nettverkene, i stedet for testsettene. Dermed hadde man fått et ferskt og ubenyttet sett for videre testing. Ulempen er at man ikke er sikret at det vil gi gode resultater på testsettet, siden man vil kun ta valideringssettet til betraktning når man skal finne den mest optimale utformingen av de nevrale nettverkene. Det er mulig et valideringssett hadde gitt mindre underfitting på nettverkene, men det kan også gi dårligere resultater når man skal predikere de ekstremt lave og høye strømprisene som kom de to siste årene.

Det viktigste for å klare å predikere gode nok strømpriser er å benytte seg av de beste mulige uavhengige variabler i nettverkene. I oppgaven har det kun blitt benyttet variabler som har en påvirkning til eller av strømprisen. Eventuelt kunne man valgt å ekskludere enkelte av variablene eller byttet de mot andre for å få bedre resultater. Riktig nok ble ikke dette forsøkt i denne oppgaven. I korrelasjonsmatrisen så man tidlig at en variabel som regn/nedbør ikke hadde stor korrelasjon til strømprisen, eller noen andre variabler. For nedbøren er med å fylle opp vannmagasinene rundt om i landet. Har man mye vann i magasinene, så har man muligheten til å produsere mye strøm. Vannet i disse magasinene slippes ikke jevnt ut hele tiden, riktig nok fordi man må balansere det med hvor mye nedbør som kommer og hvor mye is som smelter. Variabelen for nedbør kunne dermed blitt byttet med vannstanden i magasinene eller hvor mye vann som slippes ut av magasinene til de som produserer strøm for Midt-Norge. Det gir en større indikasjon på hvor mye strøm som produseres og om det hadde vært utslagsgivende på strømprisen.

Baserer man seg av korrelasjonen mellom strømprisen og de uavhengige variablene, så var også uran og VIX to som korrelerte lavt med strømprisen. At uran korrelerer lavt med norske strømpriser kan være forvirrende, spesielt siden vi importerer strøm fra blant annet Sverige som benytter seg av atomkraftverk. Riktig nok er det ikke alltid en kausal sammenheng mellom ting som korrelerer sterkt, så man skal ikke nødvendigvis utelukke variabler som ikke korrelerer. Skulle man ha funnet det mest optimale datasettet med de beste variablene, måtte man heller ha testet flere kombinasjoner både med og uten enkeltvariabler.

6. Konklusjon og videre arbeid

6.1 Konklusjon

I vår oppgave har vi ønsket å finne den mest optimale nettverkskonfigurasjonen for prediksjon av strømpriser, gjennom å analysere de nevrale nettverkens prestasjoner. Nettverket med det beste resultatet for vårt formål var LSTM med 4,408 RMSE og 96% R-squared, som gir en veldig høy nøyaktighet. Faktisk klarer nettverket å predikere verdier med en gjennomsnittlig feilmargin på kun 4 €/MWh eller 4 øre/kWh. Likevel er nettverket trolig ikke optimalt slik som det er nå hvis det skal benyttes til å predikere strømpriser for lengre tidsperioder.

Strømpriser kan være veldig vanskelige å forutse og særlig hvis strømprisen endrer seg drastisk over en kort periode. Det er mange faktorer som slår inn når strømpriser endrer seg, og for at nettverkene skal kunne predikere endringene er de avhengige av å få tilgang til de relevante variablene som påvirker prisen. Det å utelate viktige variabler kan medføre at nettverket ikke presterer godt. Måten nettverkene predikerer prisene på er ved å benytte tidligere mønster og erfaringer i datasettet. Dette kan medføre at nettverket får større avvik ved store svingninger i prisene, slik som vi har observert i våre nevrale nettverk.

6.2 Videre arbeid

Det tredimensjonale datasettet kan omformes slik at det kan predikere flere timer fremover. Eksempelvis kan vår modell omformes slik at de nevrale nettverkene predikerer en time frem, for å så kunne bruke denne prediksjonen til å predikere den neste timen igjen. Dette kan repeteres og strekkes ut til så mange timer som er ønskelig, men flere timer vil redusere treffsikkerheten til nettverket. Dette vil kunne gi flere og mer relevante applikasjoner av de nevrale nettverkene, slik som å analysere neste døgns strømpriser til et kostnadsminimerende formål.

For videre automatisering av prediksjoner kan det settes opp en dataauthentings kode som oppdaterer seg selv kontinuerlig ved ønskede tidspunkt. Ved å innhente de nyeste prisene på strøm og andre relevante variabler ved gitte intervaller kan nettverkene tilpasses til det oppdaterte datamaterialet og gi løpende prediksjoner.

Det er mange kombinasjoner av lag, noder og aktiveringsfunksjoner som ikke er prøvd ut i vårt prosjekt. Dette impliserer at det fortsatt er mange muligheter for å forbedre nettverkene, og dette er et område hvor videre forskning kan prøve å finne metoder for å optimalisere.

7. Referanseliste

Artificial neural network (2022) Tilgjengelig fra:

https://en.wikipedia.org/wiki/Artificial_neural_network#History (Hentet: 19.april 2022).

Aurdal, B.H. (2021) Nordea Markets: Overdreven frykt for gassprisene, *finansavisen.no*.

Tilgjengelig fra: <https://finansavisen.no/nyheter/energi/2021/10/02/7746743/firedobling-av-gassprisene-bekymrer-ikke-nordea-strategene> (Hentet: 01.april 2022).

Bach, D. og Høgseth, M.H. (2021) Strømprisene falt kraftig: I natt kunne du få betalt for å bruke strøm, *e24.no*. Tilgjengelig fra: <https://e24.no/olje-og-energi/i/ALpWAM/stroemprisene-falt-kraftig-i-natt-kunne-du-faa-betalt-for-aa-bruke-stroem> (Hentet: 03.april 2022).

Blaker, M. (2022) Strømkrisen er ikke over: Varsler ekstrempriser på strøm i lang tid, *Nettavisen.no*. Tilgjengelig fra: <https://www.nettavisen.no/okonomi/stromkrisen-er-ikke-over-varsler-ekstrempriser-pa-strom-i-lang-tid/s/12-95-3424240056> (Hentet: 01.april 2022).

Bouley, C. (2020) *Recurrent Neural Networks for Electricity Price Prediction*. Tilgjengelig fra: <https://towardsdatascience.com/recurrent-neural-networks-for-electricity-price-prediction-a26f8411ea44> (Hentet: 12.april 2022).

Brownlee, J. (2017a) *A Tour of Recurrent Neural Network Algorithms for Deep Learning*. Tilgjengelig fra: <https://machinelearningmastery.com/recurrent-neural-network-algorithms-for-deep-learning/> (Hentet 06.april.2022).

Brownlee, J. (2017b) How to Diagnose Overfitting and Underfitting of LSTM Models. Tilgjengelig fra: <https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/> (Hentet: 04.april 2022)

Deepak, S. og K.S, S. (2011) *Electricity price forecasting using artificial neural networks*. Tilgjengelig fra: <https://doi.org/10.1016/j.ijepes.2010.12.009> (Hentet 06.april.2022).

Dvergsdal, H. (2019) nevralt nettverk, *Store norske leksikon*. Tilgjengelig fra: https://snl.no/nevralt_netverk (Hentet 06.april 2022).

Energi Norge (2022a) *Bare fire prosent har fastpris på strøm*. Tilgjengelig fra: <https://www.energinorge.no/nyheter/2022/flere-bor-vurdere-fastpris-pa-strom/> (Hentet: 10.april 2022).

Energi Norge (2022b) *Derfor er strømprisen uvanlig høy*. Tilgjengelig fra:

- <https://www.energinorge.no/fagomrader/strommarked/derfor-er-stromprisen-hoyere-i-ar-enn-i-fjor/> (Hentet: 04.april 2022).
- Euro valutakurser - EUR til NOK* (u.å) Tilgjengelig fra: <https://www.valuta-kurser.no/euro-valutakurser> (Hentet: 01.april 2022).
- Hasan, S. (2020) *Recurrent Neural Network and it's variants....* Tilgjengelig fra: <https://medium.com/analytics-vidhya/recurrent-neural-network-and-its-variants-de75f9ee063> (Hentet: 06.april 2022).
- Hovland, K.M. (2021) Rådyr gass gir rekordinntekter: – Tror de vil kunne gå høyere, *e24.no*. Tilgjengelig fra: <https://e24.no/olje-og-energi/i/QyPq5q/raadyr-gass-gir-rekordinntekter-tror-de-vil-kunne-gaa-hoyere> (Hentet: 03.april 2022).
- Hovland, K.M. (2020) I natt fikk du penger for å bruke strøm: – Andre gang i historien, *e24.no*. Tilgjengelig fra: <https://e24.no/olje-og-energi/i/VqqaJr/i-natt-fikk-du-penger-for-aa-bruke-stroem-andre-gang-i-historien> (Hentet: 03.april 2022).
- Høgseth, M.H. og Svendsen, M. og Hovland, K.M. (2022) Laveste strømpris på én måned, *e24.no*. Tilgjengelig fra: <https://e24.no/olje-og-energi/i/bGaV85/laveste-stroempris-paa-en-maaned> (Hentet: 06.april 2022).
- Kroepelien, K. (2022) Flere bør vurdere fastpris på strøm, *Finansavisen.no*. Tilgjengelig fra: <https://finansavisen.no/nyheter/debattinnlegg/2022/02/01/7811235/flere-bor-vurdere-fastpris-pa-strom> (Hentet: 04.april 2022).
- Lang-Ree, A. og Sæter, S.O. (2021) Kullprisen mer enn tredoblet i år: – Vi må komme med støttepakker, *e24.no*. Tilgjengelig fra: <https://e24.no/olje-og-energi/i/bGkVEA/kullprisen-mer-enn-tredoblet-i-aar-vi-maa-komme-med-stoettepakker> (Hentet: 04. april 2022).
- Nord Pool (u.å) *Day-ahead Prices*. Tilgjengelig fra: <https://www.nordpoolgroup.com/en/Market-data1/Dayahead/Area-Prices/ALL1/Hourly/?view=table> (Hentet: 04. april 2022).
- NRK (2010) *Tidenes høyeste strømpris*. Tilgjengelig fra: <https://www.nrk.no/okonomi/-tidenes-hoyeste-strompris-1.7005858> (Hentet: 02.april 2022).
- Rustad, M.E. (2021) Ny elbil-rekord: Over 110.000 biler solgt i 2021, *e24.no*. Tilgjengelig fra: <https://e24.no/det-groenne-skiftet/i/34X0Wv/ny-elbil-rekord-over-110000-biler-solgt-i-2021> (Hentet: 02.april 2022).
- Saeed, M. (2021) *An Introduction To Recurrent Neural Networks And The Math That Powers Them*. Tilgjengelig fra: <https://machinelearningmastery.com/an-introduction-to-recurrent->

[neural-networks-and-the-math-that-powers-them/](#) (Hentet: 03. april 2022).

Saxena, S. (2021) *Introduction to Long Short Term Memory (LSTM)*. Tilgjengelig fra:

<https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> (Hentet: 03. april 2022).

Solli, M. (2021) Regningen for desember-strømmen er klar: - Som å betale 100 kroner for et

brød, *Nettavisen.no*. Tilgjengelig fra: <https://www.nettavisen.no/okonomi/regningen-for-desember-strommen-er-klar-som-a-betale-100-kroner-for-et-brod/s/12-95-3424225168>

(Hentet: 24. mars 2022).

Tidemann, A. og Elster, A.C. (2022) maskinlæring, *Store Norske Leksikon*. Tilgjengelig fra:

<https://snl.no/maskinl%C3%A6ring> (Hentet: 06. april 2022).

Zach (2021) *MSE vs. RMSE: Which Metric Should You Use?*. Tilgjengelig fra:

<https://www.statology.org/mse-vs-rmse/> (Hentet: 08. april 2022).

8. Vedlegg

Vedlegg A: Kodefiler

