

Henrik Bøhler Evensen
Ole Jakob Dahl
Magnus Greiff Bjordal

En prediktiv analyse av strømprisen ved bruk av maskinlæring

Bacheloroppgave i Økonomi og administrasjon - Business Analytics
Veileder: Denis M. Becker
April 2023

Henrik Bøhler Evensen
Ole Jakob Dahl
Magnus Greiff Bjordal

En prediktiv analyse av strømprisen ved bruk av maskinlæring

Bacheloroppgave i Økonomi og administrasjon - Business Analytics
Veileder: Denis M. Becker
April 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen



Kunnskap for en bedre verden

FORORD

Dette er vår avsluttende oppgave i bachelorstudiet i økonomi og administrasjon ved NTNU Handelshøyskolen i Trondheim. Oppgaven er knyttet til spesialiseringen innenfor *Business Analytics* og ble skrevet våren 2023.

Med stort mediafokus og samfunnsinteresse knyttet til strømpriser har det vært av stor interesse å ta et dypdykk i dette svært dagsaktuelle temaet. Vi har tatt i bruk kunnskap fra hele studieløpet under arbeidet med denne oppgaven. Særlig har ferdigheter fra faget *Essentials of Business Analytics* kommet godt med under konstruksjonen av de ulike modellene. Hele prosessen har vært svært lærerik og interessant, og vi håper at oppgaven vår kan hjelpe til bedre forståelse rundt strømpriser, samt fremtidige prediksjonsmodeller.

Vi ønsker å rette en takk til vår veileder, Denis Becker, som har gitt oss god hjelp ved ulike hindringer under arbeidet med oppgaven. I tillegg har han gjort fagene i *Business Analytics* interessante og lærerike, noe som ga oss et sterkt grunnlag når vi skulle skrive denne oppgaven.

Innholdet i denne oppgaven står for forfatterens regning.

God lesing!

SAMMENDRAG

Dereguleringen av kraftmarkedet har ført til et økende behov for risikostyring i energimarkedet. Prediksjon av strømpriser er dermed en viktig problemstilling og har stor betydning for både produsenter og forbrukere. Strømprisene er ene og alene bestemt av markedet, det vil si tilbud og etterspørsel. Resultatet av dette er i økonomisk sammenheng en effektiv løsning, men byr på problemer knyttet til høy usikkerhet og volatilitet for markedsaktører som produsenter, selgere og konsumenter. Det er flere faktorer som forklarer den høye volatiliteten man ser i strømprisene. Den viktigste forklaringsvariabelen er trolig det faktum at elektrisitet ikke kan lagres på en effektiv måte. Dette fører til at produksjon og konsum må balanseres kontinuerlig. Som følge av dette vil tilbud- og etterspørselssjokk ikke kunne utjevnes enkelt, og påvirker likevektsprisene direkte.

I denne studien har vi gjort et dypdykk i elektrisitetsmarkedet for å forsøke å avdekke hvilke variabler som er med på å forklare strømprisen for prisområde NO1, Østlandet. Ved hjelp av maskinlæring har vi forsøkt å avdekke trender i historisk data for å kunne si noe om fremtidige strømpriser. Maskinlæringsmodellene som er tatt i bruk er både lineær og logistisk regresjon, nevralt nettverk, samt en bayesiansk modell. Studien har tatt utgangspunkt i en periode på 10 år - fra 2013 til 2023, hvor vi har brukt data fra både åpne og private kilder.

Modellene preges i stor grad av omfattende strukturelle endringer i strømprisen fra og med året 2022, hvor den åpenbare årsaken til denne strukturelle endringen er krigen mellom Russland og Ukraina. Resultatet viser derfor til modeller som kun til en viss grad er i stand til å forutsi endringer i strømprisen. Likevel har funnene gitt oss mer innsikt og forståelse i hvilke faktorer som er med på å fastsette strømprisen.

ABSTRACT

The deregulation of the power market has led to an increasing need for risk management in the energy market. Prediction of electricity prices is thus an important issue and has great significance for both producers and consumers. Electricity prices are solely determined by the market, i.e. supply and demand. The result of this is, in an economic context, an effective solution, but may lead to certain problems related to high uncertainty and volatility for market participants; not only for producers and sellers, but also consumers. There are several factors that explain the high volatility seen in electricity prices. The most important explanatory variable is probably the fact that electricity cannot be stored efficiently. This means that production and consumption must be continually balanced. As a result, supply and demand shocks cannot be smoothed out easily, and will affect equilibrium prices directly.

In this study, we have dug deeper into the functionalities of the electricity market to try to uncover which variables help explain the electricity price for price area NO1, Eastern Norway. With the help of machine learning, we have tried to exhibit trends in historical data in order to be able to say something about future electricity prices. We have used machine learning models like linear and logistic regression, neural networks and a Bayesian model. The study is based on a ten year period - from 2013 to 2023, where we have used both open and private sources.

To a large extent, the models are characterized by extensive structural changes in the electricity price from and including the year of 2022. The obvious reason for this structural change is the war between Russia and Ukraine. For that reason, the result of our study visualizes models which are just some-what capable of predicting the electricity price. Nevertheless, our discoveries have given us more insight and comprehension regarding what drives the electricity price.

INNHOLDSFORTEGNELSE

1. INNLEDNING.....	1
1.1 Aktualisering av tema.....	1
1.2 Presentasjon av problemstilling.....	2
1.3 Struktur.....	2
2. TEORI.....	3
2.1 Om kraftmarkedet og det norske kraftsystemet.....	3
2.2 Strømforsyning og strømnettet.....	6
2.3 Hva påvirker strømprisen?.....	7
2.4 Maskinlæring.....	10
2.4.1 Kunstig nevralt nettverk.....	10
2.4.2 Lineær regresjon.....	12
2.4.3 Logistisk regresjon.....	14
2.4.4 Naive Bayes Classification.....	15
2.5 Korrelasjon.....	15
3. DATA OG METODE.....	17
3.1 Datasettet.....	17
3.1.1 Avhengig variabel.....	18
3.1.2 Uavhengige variabler.....	18
3.2 Databehandling.....	19
3.2.1 Manglende verdier.....	19
3.2.2 Omgjøring av variabler.....	21
3.3 Konstruksjon av modeller.....	21
3.3.1 Lineær regresjon.....	21
3.3.2 Naive Bayes Classification.....	23
3.3.3 Logistisk regresjon.....	24
3.3.4 ANN.....	25
4. ANALYSE OG RESULTAT.....	28
4.1 Visualisering og deskriptiv statistikk.....	28
4.1.1 Avhengig variabel.....	28
4.1.2 Uavhengige variabler.....	30
4.2 Resultater.....	35
4.2.1 Lineær regresjon.....	35
4.2.2 Naive Bayes Classification.....	37
4.2.2 Logistisk regresjon.....	39
4.2.3 ANN.....	39
5. DISKUSJON.....	41
5.1 Generell diskusjon.....	41
5.2 Forbedring og evaluering av modellene.....	42
6. KONKLUSJON.....	44
6.1 Videre arbeid.....	45
7. LITTERATURLISTE.....	46
8. APPENDIX.....	50
8.1 Korrelasjonsmatrise.....	50
8.2 Resterende uavhengige variabler.....	51
8.3 Zip-fil.....	56

FIGURLISTE

Figur 2.1:	Illustrasjon av kraftmarkedet, 2022.....	6
Figur 2.2:	Price calculation, 2023.....	7
Figur 2.3:	Loss functionality, 2023.....	9
Figur 2.4:	Neural Network Architecture, 2019.....	11
Figur 3.1:	Manglende verdier.....	20
Figur 3.2:	Oppdeling av dataen.....	22
Figur 3.3:	Funksjonen $\tanh(x)$	27
Figur 4.1:	Spotpris elektrisitet NO1 i perioden 2013 - 2023.....	28
Figur 4.2:	Gjennomsnittlig strømpris og standardavvik i perioden 2013-2022.....	29
Figur 4.3:	Korrelasjonsmatrise.....	30
Figur 4.4:	Absolutt korrelasjon med spotpris i synkende rekkefølge.....	30
Figur 4.5:	Gasspris i perioden 2013-2023.....	31
Figur 4.6:	Sammenligning spotpris, elektrisitet og gasspris i perioden 2013-2022.....	32
Figur 4.7:	Kvotepris på CO ₂ i perioden 2013-2023.....	33
Figur 4.8:	Kullpris i perioden 2013-2023.....	33
Figur 4.9:	Importert elektrisitet i perioden 2013-2023.....	34
Figur 4.10:	Visualisering av testresultat for multippel lineær regresjonsanalyse.....	37
Figur 4.11:	Observasjoner i hver priskategori.....	38
Figur 4.12:	Fargekart av testsettet for Naive Bayes.....	38
Figur 4.13:	Fargekart av testsettet for logistisk regresjon.....	39
Figur 4.14:	Sammenligning mellom predikert og faktisk strømpris.....	40

TABELLOVERSIKT

Tabell 1:	Oversikt over variabler og tilhørende datakilde.....	17
Tabell 2:	Oversikt over gjennomsnittsverdier, maksimums- og minimumsverdier, og standardavvik for strømpris.....	29
Tabell 3:	Resultater fra multippel lineær regresjonsanalyse.....	35

APPENDIX - Figurer

Figur 8.1: Korrelasjonsmatrise	Figur 8.7: NOK til EUR 2013-2023
Figur 8.2: Oljepris 2013-2023	Figur 8.8: NOK til USD 2013-2023
Figur 8.3: Eksport av elektrisitet 2013-2023	Figur 8.9: Fyllingsgrad NO1 2013-2023
Figur 8.4: Middelterperatur 2013-2023	Figur 8.10: Fyllingsgrad NORGE 2013-2023
Figur 8.5: Nedbør 2013-2023	Figur 8.11: Norges energiforbruk 2013-2023
Figur 8.6 USD til EUR 2013-2023	

1. INNLEDNING

1.1 Aktualisering av tema

Å kunne si noe om fremtidige strømpriser er en viktig problemstilling i energimarkedet, og har stor betydning for både aktører og forbrukere. Strømprisene kan til tider utvise høy volatilitet og vil avhenge av en rekke ulike faktorer. Å ha en underliggende forståelse av strømprisene og kunne predikere fremtidige priser kan være avgjørende for å fatte rasjonelle beslutninger i energibransjen og i forbrukermarkedet. I dagens samfunn er energibruk og klimautslipp stadig økende. Som en kilde til elektrisitet er strøm en nødvendighet for å drive et moderne samfunn. Samtidig er strømproduksjon- og distribusjon ofte en av de største kildene til CO₂-utslipp og annen forurensning. I lys av dagens økende fokus på reduksjon av karbonutslipp og overgang til fornybare energikilder, blir det stadig viktigere å kunne forstå og analysere strømpriser for å kunne planlegge og gjennomføre effektive tiltak både i energibransjen og i forbrukermarkedet.

I dag er Norge en del av et felles kraftmarked i Norden, som inkluderer Sverige, Danmark og Finland. Det nordiske kraftmarkedet er igjen integrert til det europeiske kraftmarkedet som følge av forbindelser til Tyskland, Nederland, Baltikum og Polen. Samtidig har utbyggingen av North Sea Link-kabelen sørget for forbindelse mellom Norge og Storbritannia etter at den ble tatt i bruk i oktober 2021. Siden det europeiske markedet i stor grad har vært preget av regionale initiativ og frivillige samarbeid, jobbes det fortsatt med å utarbeide det indre energimarkedet for å koble sammen de europeiske markedene. Markedet i Europa består i dag av 25 land, og omfatter omtrent 95 prosent av Europas kraftforbruk (Energifakta Norge, 2022).

Siden 2021 har prisen på strøm skutt i været, og blitt et særdeles omtalt tema i mediene. Strømanalytiker Tor Reier Lilleholt i selskapet Volue Insight anslår at prisen i 2022 har blitt rundt tre ganger høyere enn rekordåret 2021, og omtrent 20 ganger høyere enn i 2020. Lilleholt viser til at situasjonen i Europa, med krigen i Ukraina, har hatt stor påvirkningskraft, og at det foreligger en ubalanse i det europeiske strømmarkedet. Samtidig påpeker han at strømprisene allerede var høye ved inngangen til 2022, som følge av at Putin kuttet eksporten av gass til Europa høsten 2021 (E24, 2022). De rekordhøye prisene har ført til at regjeringen

har innført midlertidige støtteordninger i håp om å bidra til mer stabilitet og forutsigbarhet. Likevel antyder eksperter og prognoser at prisen fortsatt kommer til å være høy i periodene fremover, noe som vil påvirke både husholdninger og næringslivet (Olje- og energidepartementet, 2023).

1.2 Presentasjon av problemstilling

På bakgrunn av ovennevnte observasjoner vil denne studien foreta et dypdykk i strømmarkedet, hvor vi ønsker å utforske hvilke faktorer som påvirker strømprisen. Ved hjelp av maskinlæring ønsker vi å lage modeller som kan fange opp relevante faktorer og predikere strømprisen for påfølgende dag. Vi har valgt å snevre inn oppgaven ved å fokusere på prisområdet NO1 Østlandet. Samtidig har vi valgt å fokusere på data som strekker seg over en tiårsperiode med håp om å kunne avdekke hvilke faktorer som har vært relevante for prisutviklingen. Gjennom vår forskning håper vi å bidra til å forbedre forståelsen av strømprismekanismen, samt gi verdifull innsikt til aktører i energisektoren for å kunne ta informerte beslutninger og optimalisere sin virksomhet. Ved å forstå de faktorene som påvirker strømprisene og utvikle prediksjonsmodeller, kan vi bidra til økt stabilitet og forutsigbarhet i energimarkedet, samt økt effektivitet og inntjening for aktører i energibransjen.

Problemstillingen kan følgelig oppsummeres slik:

- 1) *Hvordan påvirker ulike fundamentalvariabler strømprisen på strømbørsen NordPool?*
- 2) *Hvilke modeller er i stand til å predikere day-ahead-prisen på Nord Pool?*

1.3 Struktur

Videre i oppgaven ønsker vi å ta for oss og presentere relevant teori i forhold til modellene vi ønsker å konstruere, samt et mer utfyllende innblikk i kraftmarkedet, strømforsyningen i Norge, og faktorer som er med på å påvirke strømprisen. Etter presentasjon av relevant teori, ønsker vi å presentere datasettet i oppgaven, hvor vi viser til hvilke modeller vi har tatt i bruk og eventuelle utfordringer knyttet til denne prosessen. Resultatene vi oppnår vil deretter presenteres i analysedelen av oppgaven. Her vil modellenes resultater forklares, samt visualiseres, før vi diskuterer våre observasjoner knyttet opp mot problemstillingen. Til slutt ønsker vi å trekke frem de mest nevneverdige resultatene, samt forslag til eventuell videre forskning innenfor fagfeltet.

2. TEORI

2.1 Om kraftmarkedet og det norske kraftsystemet

En teoretisk definisjon av elektrisk strøm kan beskrives som elektrisk ladning i bevegelse. Ved hjelp av en leder kan man frakte energi fra ett sted til et annet. Mengden av elektrisitet som strømmer gjennom et punkt for en gitt periode blir et produkt av tid og effekt. Energien som overføres måles vanligvis i kilowattimer (kWh) eller megawattimer (MWh).

Overføringen er både enkel og strømmen kan produseres ved hjelp av flere forskjellige energikilder. Som den mest anvendelige energiformen utgjør elektrisk strøm en betydelig rolle i et moderne samfunn, da den på mange måter inngår i alle samfunnsaktiviteter. Som følge av den elektriske energiens betydning, er man svært sårbar for svingninger i strømforsyningen (Grøn, 2021).

I økonomisk forstand er strøm å anse som en relativt ny handelsvare. Dessuten skiller strøm seg fra materielle handelsvarer som for eksempel olje eller naturgass. Elektrisk strøm kan beskrives som utskiftbart, i den forstand at én MWh strøm produsert av kull eller naturgass tilsvarer den samme mengden med energi. Samtidig må produksjon og forbruk av strøm foregå simultant, ettersom lagring av strøm er svært lite effektivt og veldig dyrt (CME Group, 2023).

Det norske kraftmarkedet og kraftsystemet har gjennom historien vært preget av vannkraftverk med strømforsyning til sitt lokale område. De lokale energiverkene hadde monopol, samt plikt til å levere strøm til sitt nærområde (NVE, 2021). Statens store satsing på vannkraftutbygging på tidlig 1900-tallet gjorde at det ble nødvendig å etablere Norges vassdrags- og elektrisitetvesen (NVE) for å forvalte Statens kraftverk (Statkraft, 2020). I 1986 ble statskraftverkene skilt ut fra NVE, og Statkraft ble opprettet som forvaltningsbedrift med ansvar for bygging og drift av statens kraftverk (Hofstad, 2020). I 1990 skjedde det store endringer i det norske kraftmarkedet, hvor en ny energilov åpnet for et fritt kraftmarked. Loven skulle føre til at prisfastsettelsen styres av tilbud og etterspørsel, slik at kraftsystemet blir mer markedsstyrt, og mindre påvirket av politiske vedtak (Askheim, 2023). I dag er kraftmarkedet i Norge et konkurranseutsatt marked der forbrukerne kan velge mellom flere strømleverandører. Siden forbrukeren kan velge leverandør på grunnlag av pris og andre faktorer, oppstår det konkurranse mellom de ulike kraftleverandørene (Olje- og

energidepartementet, 2021). Dette har resultert i et økende behov for risikostyring i energimarkedet.

Som følge av den nye energiloven i 1990 ble det lagt fokus på å bygge et skille mellom produsenter og distributører av kraft. Kraftsektoren ble organisert i en konkurransedel og en monopoldel, hvor selve transporten av kraft fortsatt skulle være monopolbasert, mens produksjonen og omsetningen av kraft skulle være åpen for konkurranse (NVE, 2021). Dette førte til at Statkraft ble delt i to statsforetak; Statkraft og Statnett. Statkraft fikk ansvaret for produksjon og omsetning, mens Statnett fikk ansvar for Norges sentralnett for kraftoverføring (Statkraft, 2020). Statkraft er i dag Norges største produsent av elektrisk kraft, og står for nesten halvparten av Norges samlede kraftproduksjon. I likhet med Statkraft er de fleste store kraftprodusenter i Norge offentlig eide, men det finnes også noen store private aktører, som Norsk Hydro og Elkem (Rosvold, 2023).

Selv om Norge har hatt krafthandel med utlandet siden 1960-tallet, ble det først i 1996 opprettet en felles nordisk kraftbørs. Dette skjedde etter at de andre nordiske landene også gjennomførte reformer i sine energilover. Denne nordiske kraftbørsen fikk navnet NordPool, og var verdens første kraftbørs der man kunne handle kraft over landegrensene. (Olje- og energidepartementet, 2021). Det nordiske kraftmarkedet, som Norge er en del av, er igjen integrert med det europeiske kraftmarkedet gjennom forbindelser i kraftnettet. Hensikten bak denne europeiske markedskoblingen er at kraften skal flyte i henhold til prisene. Mer kraft vil overføres fra områder med lavere priser til områder med høyere priser. Dette vil gi en bedre utnyttelse av eksisterende nett- og produksjonsressurser. (Energifakta Norge, 2022).

I kraftmarkedet er det nettselskapene som holder styr på hvor mye kraft produsentene tilfører kraftnettverket, og hvor mye sluttforbruker tar ut. Når kraft blir tilført nettverket flytter det minste motstands vei, og det er dermed ikke mulig å skille kraftleveranser fra hverandre. Når det kommer til kraftmarkedet deler man hovedsakelig inn i engrosmarkedet og sluttforbrukermarkedet. Det er i sluttforbrukermarkedet at forbrukere som husholdninger, industri, butikker og hoteller handler kraft fra ulike kraftleverandører. Her er det fri konkurranse mellom kraftleverandørene, og sluttforbruker står fritt til å velge kraftleverandør (Energifakta Norge, 2022). I Norge er det mange aktører innenfor kraftleveranse til sluttforbrukermarkedet. Over 140 strømleverandører opererer i landet, men mange er små,

lokale selskap. Blant de større selskapene finner man blant annet Fjordkraft, Fortum og Hafslund (Lorvik, 2021).

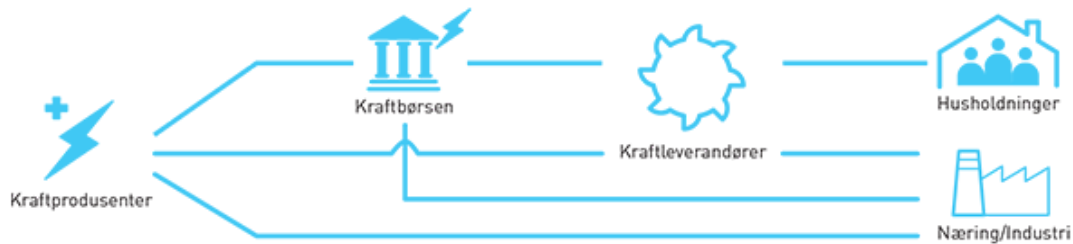
I engrosmarkedet selger og kjøper aktører, som kraftleverandører, produsenter, store industrikunder og meglere, store kraftvolum. Det er her kraftleverandørene handler kraft som de så selger videre til sluttforbrukerne. Siden kraftmarkedet er avhengig av total balanse mellom produksjon og forbruk, eksisterer det flere organiserte markeder innenfor engrosmarkedet for å oppnå markedslikevekt (Energifakta Norge, 2022). Disse markedene er:

- Day-aheadmarkedet
- Kontinuerlig intradagmarked
- Balansemarkeder

Day-aheadmarkedet er hovedmarkedet innen krafthandelen på engrosmarkedet. Her sikres i hovedsak balansen mellom tilbud og etterspørsel. Handelen skjer på organiserte kraftbørser, og i Norge har Nord Pool og EPEX Spot konsesjon for å drive markeds plass for salg av elektrisk energi. I day-aheadmarkedet leverer aktører inn salgs- og kjøpsbud til kraftbørsen for neste døgn. I Norge gir Statnett informasjon om transmisjonskapasiteten for hvert budområde, og budene må legges inn før auksjonen stenges kl 12. Dermed beregnes prisene for kommende døgn på bakgrunn av de innkomne salgs- og kjøpsbudene, samt transmisjonskapasiteten i strømmettet. Siden det europeiske strømmettet er tett integrert, er det nordiske day-aheadmarkedet koblet til store deler av det europeiske day-aheadmarkedet (Energifakta Norge, 2022). Denne koblingen skjer gjennom såkalt implisitt auksjon, nemlig en auksjon der det bys på energi og nødvendig overføringskapasitet i strømmettet i samme operasjon (El-ordboken, 2016).

Det kan oppstå hendelser som forstyrrer balansen mellom produksjon og forbruk avtalt i day-aheadmarkedet. Avbrudd i kraftproduksjon, endrede vær-prognoser eller uregelmessig høyt forbruk hos sluttforbruker kan være årsaker til at aktører i engrosmarkedet trenger å justere handelen for å oppnå balanse mellom tilbud og etterspørsel. For å unngå dette problemet kan aktørene handle seg i balanse på intradagmarkedet. I intradagmarkedet kan det handles kontinuerlig i tidsrommet mellom day-ahead markedet stenger og inntil en time før driftstimen (Energifakta, 2022).

På tross av day-ahead markedet og intradagmarkedet vil det stadig være hendelser som skaper ubalanse i selve driftstimen. For å unngå dette benyttes balansemarkedet for å regulere forbruk eller produksjon opp eller ned. Statnett handler inn primær- og sekundærreserver i døgn- og ukesmarkeder, og ved ubalanse i strømmettet aktiveres disse reservene. Ved ytterligere behov aktiveres tertiærreservene, som er anskaffet i regulerkraftmarkedet, et felles balansemarked for det nordiske kraftsystemet (Energifakta, 2022).



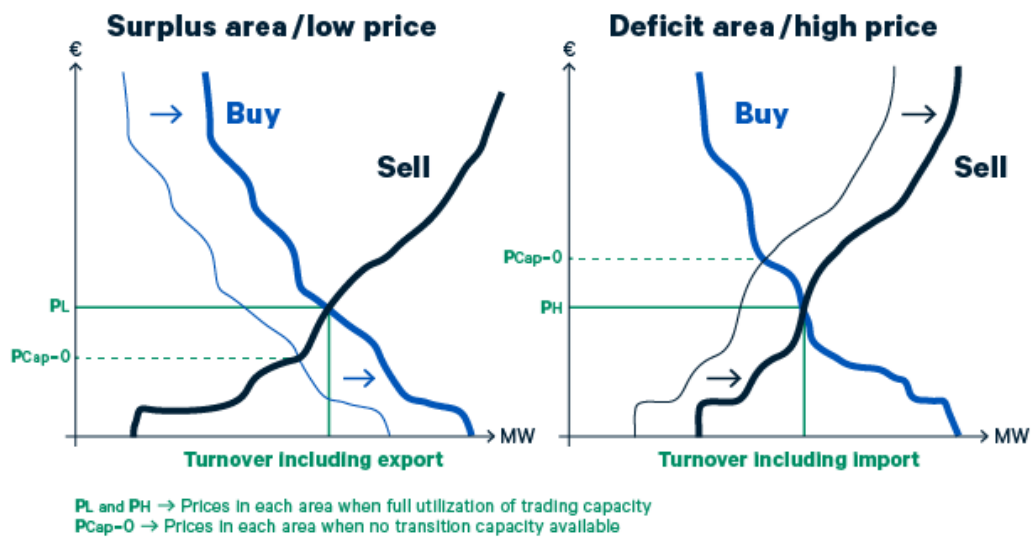
Figur 2.1: Illustrasjon av kraftmarkedet, 2022

2.2 Strømforsyning og strømmettet

Siden det er store avstander mellom produksjon og forbruk av kraft i Norge, er et velutbygd strømmnettverk elementært. Strømmettet må være kapabelt til å håndtere de største toppene i kraftforbruket, samt eksportere strøm ut når produksjonen er større enn forbruket. Strømmettet kan deles inn i transmisjonsnett, regionalnett og distribusjonsnett. Transmisjonsnett er et landsdekkende system og omfatter blant annet utenlandsforbindelsene. Dette strømmettet binder sammen store produsenter og forbrukere. Statnett, som er systemansvarlig i det norske kraftmarkedet, opererer transmisjonsnett. Regionalnett binder sammen transmisjonsnett og distribusjonsnett, mens distribusjonsnett sørger for kraft til mindre sluttforbrukere (Energifakta Norge, 2019).

Det norske strømmnettverket er delt inn i fem prisområder. Disse sonene er landsdelene i landet vårt; Østlandet, Sørlandet, Midt-Norge, Nord-Norge og Vestlandet. Bakgrunnen for ulike prissoner kommer av at mesteparten av det norske kraftmarkedet er værbasert, og kraftsituasjonen vil variere mellom landsdelene. Det er ikke alltid balanse mellom kraftforbruket og produksjonen i de ulike områdene, og det kan dermed oppstå behov for å overføre strøm fra et område til et annet. Her kan det oppstå såkalte flaskehals, nemlig begrensninger i hvor mye kraft som kan overføres i strømmettet. I tillegg til fysiske begrensninger i overføringen av strøm i strømmettet kan det også være nødvendig å begrense overføringer av hensyn til driftssikkerheten. Siden kraftoverføringen er begrenset, er det

nødvendig med flere prisområder for å oppnå markedslikevekt. Skulle Norge hatt en felles pris ville det vært nødvendig med mye mer nett (Statnett, 2022).



Figur 2.2: Price calculation, 2023

2.3 Hva påvirker strømprisen?

Enkelt sagt bestemmes strømprisen av tilbud og etterspørsel; med andre ord hvor mye som produseres, og hvor mye som forbrukes (NVE, 2018). Prisen påvirkes av hvor mye kraft som er tilgjengelig og hvor mye som forventes brukt (Helgelandkraft, 2022). Selv om denne prosessen tilsynelatende kan virke enkel, er det en rekke faktorer som påvirker både tilbudet (tilgjengelig kraft) og etterspørselen (forventet forbruk av kraft).

Nedbørmengde og fyllingsgrad i vannmagasiner

Norges kraftproduksjon skjer hovedsakelig gjennom vannkraft. Vannkraft utgjør omtrent 90% av den norske kraftforsyningen, og avhenger naturligvis av den årlige nedbørmengden (Energifakta Norge, 2022). Med andre ord vil nedbørmengden i stor grad påvirke strømprisen. I perioder med lite nedbør og tilsig i vannmagasinene, vil vannet i magasinene få en høyere verdi noe som isolert sett resulterer i en høyere strømpris. Tilsvarende vil perioder med mye nedbør resultere i en lavere pris.

Temperaturforhold

Ikke overraskende forbrukes mer strøm når det er kaldere. I vinterhalvåret vil flere konsumenter ta i bruk varmepumper eller elektriske ovner. Da vil etterspørselen etter strøm

være høy, og strømprisen går opp. Motsatt vil de fleste av oss forbruke mindre strøm når det er varmt ute, og følgelig vil strømprisen være lav.

Import og eksport

Med tanke på at Norge er en del av et globalt kraftmarked og at kraft flyter på tvers av landegrensene, vil strømprisene på resten av kontinentet også påvirke de norske strømprisene (NTE, 2022).

Kull- og gasspris

Prisen på kull og gass er andre faktorer som kan forklare hvordan den norske strømprisen avhenger av vår tilknytning til et større kraftnettverk. Fastlands-Norge har ingen gass- eller kullkraftverk, men gass- og kullkraftverk i Europa bidrar i vesentlig grad med å danne det totale kraftbildet. I Norge benyttes naturlig vannforsyning til å produsere mesteparten av kraften, mens gass- og kullkraftsprodusenter må betale for råvarene som medgår til produksjonen av kraft. På den måten vil kostnaden ved produksjon av strøm øke dersom prisen på kull og/eller gass øker (Helgelandkraft, 2022). Verdt å merke seg er at klimaverstingen kull er i ferd med å fases ut, og vil i fremtiden sannsynligvis ikke lengre ha innvirkning på strømprisen. Naturgass ses på som en sannsynlig overgangsressurs for mange land.

Valutakurs

Strømprisen fastsettes i euro på det europeiske kraftmarkedet, mens kull blir priset i dollar. For norske forbrukere vil strømprisen derfor øke dersom den norske kronen svekker seg mot euro. På lik linje vil prisen på kull i norske kroner øke når dollarkursen øker, som naturligvis resulterer i høyere råvarekostnader. Produsentene må dermed senke produksjonen, noe som minker tilgjengelig kraft på kraftbørsen. Dette fører til økte strømpriser.

Prisen på CO₂

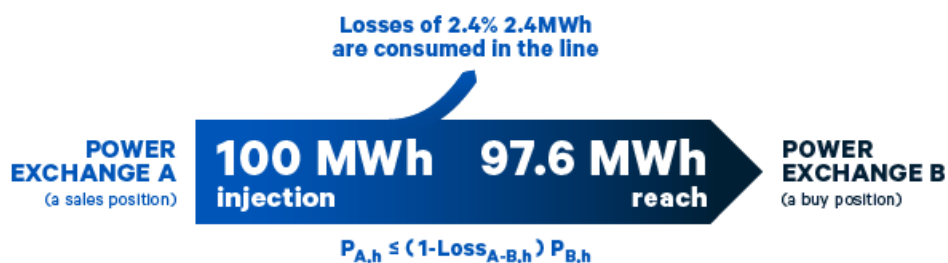
Kvotesystemet til EU på CO₂ setter tak på hvor mange CO₂-ekvivalenter som maksimalt kan slippes ut fra år til år. Én klimakvote gir tillatelse til å slippe ut et tonn med CO₂-ekvivalenter. Aktører i industrien kan kjøpe og selge klimavoter for å gjøre opp for sitt utslipp. Siden taket på CO₂-ekvivalenter reduseres over tid, vil prisen på CO₂-kvoter gå opp med mindre aktører reduserer utslipp fremfor å kjøpe kvoter (Miljødirektoratet, 2019). Høye priser på

CO₂-kvoter resulterer i høyere produksjonskostnad ved gass- og kullkraftverk, og dermed høyere strømpriser.

Andre faktorer

I tillegg til ovennevnte faktorer, er det også andre faktorer som er med å påvirke spotprisen. Felles for disse faktorene er at de kan klassifiseres som ikke-kvantifiserbare variabler. En faktor er for eksempel overføringskapasitet på strømmettet. Dersom det eksempelvis er vedlikehold på kabler og følgelig begrenset kapasitet på strømmettet, vil det oppstå flaskehalser mellom tilbud og etterspørsel (Helgelandkraft, 2022). Om de nordiske landene ikke får overført nok kraft til Sentral-Europa, vil prisene falle i Norden som følge av et overskudd på tilbudssiden. Likeledes vil prisen på strøm i Sentral-Europa øke som følge av tilbudsunderskudd. I likhet med overføringskapasitet er forbruksvaner en annen påvirkningsfaktor. Dette handler i bunn og grunn om betalingsvilligheten til hver enkelt forbruker. Når vi ikke ser oss villige til å betale for dyr strøm, velger vi å forbruke mindre. Lavere etterspørsel resulterer i en lavere pris. Med andre ord vil pris påvirke pris.

En annen forklaringsvariabel verdt å nevne er hvor mye strøm som går tapt i overføringsprosessen på strømmettet. Dette kan vi i matematisk optimalisering og beslutningsteori definere som en tapsfunksjon (fra engelsk: *Loss functionality*). Strømmettet er altså utsatt for tap. Med andre ord kan deler av den energien som strømlinjen mottar i én ende gå tapt, noe som tilsier at mottatt strøm i den andre enden vil være mindre enn hva som i utgangspunktet er sendt ut. Tapsfaktoren på det europeiske strømmettet ligger vanligvis et sted mellom 2,4% - 3,4%. Hvis vi forutsetter at mengden strøm som går tapt er 2,4%, vil overføringsprosessen se slik ut:



Figur 2.3: Loss functionality, 2023

Følgelig vil energien som til slutt importeres tilsvare mengden energi som blir eksportert minus mengden energi som går tapt under overføringsprosessen (Nordpoolgroup, 2023).

2.4 Maskinlæring

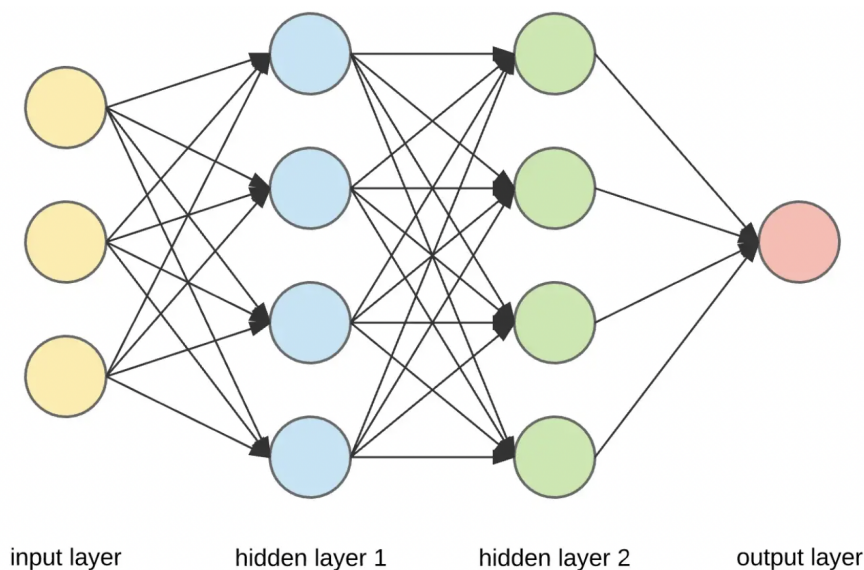
Maskinlæring er en gren av kunstig intelligens og kan forstås som vitenskapen bak opplæring av maskiner til å analysere og lære av data på lik linje med det vi mennesker gjør (IBM, 2016). Målet med maskinlæring er derfor å “trene” maskiner slik at de selv kan identifisere mønstre uten videre programmering og menneskelig intervensjon. Man benytter statistiske metoder for å la maskinen finne slike mønstre i store datamengder. Maskinlæring er dermed nært knyttet til *prediksjon*, eller *prediktiv analyse*, som handler om å uttale seg om fremtiden på bakgrunn av komplekse mønstre i historisk data (Ramsøy, 2021). Dette gjøres ved å dele opp datamengden i et treningssett og et testsett. Den aktuelle modellen anvendes på treningssettet hvor hensikten er at maskinen selv trener seg opp for å kunne predikere mønstre i nye data (Elster & Tidemann, 2022).

Det finnes en rekke former for maskinlæring. Noen vanlige metoder, som for øvrig vil bli brukt i denne oppgaven, er nevrale nettverk, regresjon og Naive Bayes klassifisering.

2.4.1 Kunstig nevrale nettverk

Kunstige nevrale nettverk (ANN) (fra engelsk: *Artificial Neural Network*) er basen i “dyp læring” (fra engelsk: *deep learning*), en undergren av maskinlæring, hvor algoritmene er inspirert av strukturene i menneskehjernen (IBM, 2021). Nevrale nettverk tar inn data (*input*), trenes opp til å gjenkjenne mønstre i dataene, for så å predikere produktet (*output*) på ny, usett data. Samtidig er nevrale nettverk avhengig av å kjøre på treningsdata for å kunne forbedre treffsikkerheten over tid. Dessuten, når nettverket er godt trent med en høy treffsikkerhet, kan de anses som svært kraftfulle verktøy innenfor datateknologi og kunstig intelligens. Et godt trent nevralt nettverk vil kunne behandle og klassifisere data under høy hastighet.

ANN består av lag av nevroner (eller *noder*); et lag hvor data kommer inn (*input layer*) og et lag som predikerer det endelige produktet (*output layer*). Mellom disse endepunktene ligger ett eller flere *skjulte* lag (fra engelsk: *hidden layers*) som foretar de fleste kalkulasjonene.



Figur 2.4: Neural Network Architecture, 2019

Det nevrale nettverket er, som figuren viser, bygget opp av et større antall enkle nevroner. Hvert nevron er tilknyttet et annet, og disse både mottar og sender signaler seg imellom. Isolert sett har nevronene begrensede funksjoner, men i et større nettverk kan samspillet mellom dem utføre store, komplekse oppgaver. Hvert kunstig nevron har sin egen koeffisient, eller “vekt”. Vektene, betegnet w_n , fungerer som multiplikatorer for nevronets input-verdi, og i et komplett nettverk vil disse bli summert. Her betegnes input-verdiene som $x_1, x_2, x_3 \dots x_n$ (Elementsofai, 2023). Et nettverk med kun én input og én output kan derfor betegnes ved formelen $y = x \cdot w$. Med andre ord beregnes output-verdien som vektoren multiplisert med vekten. Vektene er normalt sett ikke oppgitt, da riktig vekting oppnås gjennom trening av nettverket.

Mange av oppgavene man ønsker å gjennomføre ved hjelp av nevrale nettverk vil nødvendigvis ikke la seg gjennomføre ved lineære sammenhenger. Derfor legger man gjerne til en aktiveringsfunksjon. Aktiveringsfunksjonen tillater nettverket å utvikle komplekse representasjoner og funksjoner basert på input-verdiene, som ikke ville vært mulig ved hjelp av en enkel lineær regresjonsmodell. De tre mest brukte funksjonene er *sigmoid*, *tanh* og *ReLU* (Chng, 2022).

2.4.2 Lineær regresjon

Regresjon kan forklares som statistiske metoder der man utforsker hvordan sammenhengen mellom en rekke uavhengige variabler forklarer variasjonen i en avhengig variabel, betegnet y . De uavhengige variablene kan betegnes som x_1, x_2, \dots, x_k (Heldal, 2006).

Dersom man forutsetter linearitet; en konstant sammenheng mellom variabler, kan man benytte lineær regresjon til å si noe om hvordan de aktuelle variablene påvirker strømprisen. Den lineære sammenhengen beregnes ved hjelp av OLS (fra engelsk: *Ordinary Least Squares*); en teknikk innen lineær regresjon som har til hensikt å finne den lineære sammenhengen som best representerer et sett av datapunkter. I likhet med nevroner i et kunstig nevralt nettverk, har hver input-variabel i en lineær regresjon også hver sin koeffisient, eller vekt. Koeffisientene benyttes til å minimere summen av de kvadrerte avstandene mellom faktiske og predikerte verdier. En enkel lineær regresjonsmodell kan betegnes på følgende måte (Heldal, 2006):

$$y = \alpha + \beta x + e$$

y : avhengig variabel

x : uavhengig variabel

β : stigningstall

α : skjæringspunkt med y -akse

e : feilledd

β kalles helningskoeffisienten, og utgjør stigningstallet for modellen. Den angir hvor mange enheters økning som forventes per enhets økning i den uavhengige variabelen (forklaringsvariabelen). Feilleddet, e , utgjør resten av variasjonen i den avhengige variabelen som ikke fanges opp av modellen.

I denne oppgaven vil vi imidlertid benytte flere forklaringsvariabler. Dette betegnes ofte som multipel lineær regresjon (MLR):

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e$$

$$y = \alpha + \sum_{j=1}^k \beta_j x_j + e$$

y: avhengig variabel

x: uavhengige variabler

α : skjæringspunkt med y-akse

β_k : stigningstall til parameter x_k

e: feilledd

Når vi estimerer regresjonslinjen ved bruk av OLS, er målet å minimere modellens varians. Variansen utgjør det kvadrerte avviket mellom observert og predikert verdi i den konstruerte modellen. Jo større variansen blir, desto mer sårbar vil modellen være når det kommer til å predikere basert på nye observasjoner. Med andre ord vil modellen produsere mindre nøyaktige resultater. Modellens totale varians består av to typer varians, kalt SSE (fra engelsk: *Sum of Squares Errors*) og SSR (fra engelsk: *Sum of Squared Residuals*). SSR viser til det kvadrerte avviket mellom regresjonslinjen og gjennomsnittet for den avhengige variabelen Y, mens SSE kan betegnes som uforklart varians og utgjør avstanden mellom observerte og predikerte verdier. Den uforklarte variansen oppstår som følge av at andre variabler, som ikke er inkludert i modellen, har påvirkning på den avhengige variabelen. Summen av SSE og SSR kalles for SST (fra engelsk: *Sum of Squares Total*), og utgjør modellens totale varians. Ved hjelp av OLS ønsker man å finne den linjen som har minst gjennomsnittlig kvadrerte avstander mellom observerte verdier og predikerte verdier (NTNU, 2019). Med andre ord er målet å minimere den uforklarte variansen, SSE:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

For å avgjøre modellens reliabilitet, brukes gjerne den forklarte variansen. Denne verdien varierer mellom 0 og 1, omtales som determinasjonskoeffisienten, og regnes ut slik:

$$R^2 = \frac{SSR}{SST}$$

Tallet viser til hvor mye av variansen vi kan forklare ved hjelp av regresjonslikningen. En høy verdi av forklart varians indikerer at de uavhengige variablene i modellen i stor grad forklarer variasjonen i den avhengige variabelen. Likevel gir den forklarte variansen kun en indikasjon på samvariasjonen. Med andre ord gir det ingen forklaring i kausal forstand, men viser kun i hvilken grad de uavhengige variablene og den avhengige variabelen samvarierer.

2.4.3 Logistisk regresjon

I tilfeller hvor Y er en *kategorivariabel*; altså når verdimengden består av et endelig antall mulige verdier, foretar vi en logistisk regresjon (Heldal, 2006). Ofte står vi overfor variabler som kun har to svarkategorier, også kalt “dikotome variabler”. Det vil da være hensiktsmessig å kode svarkategoriene med $Y = 0$ og $Y = 1$. En logistisk modell vil beregne sannsynligheten for at den avhengige variabelen er 1 basert på de aktuelle forklaringsvariablene. Sannsynligheten kan formuleres slik:

$$E(Y|x) = 1 \cdot P(Y = 1|x) + 0 \cdot P(Y = 0|x) = P(Y = 1|x) = \pi(x)$$

For å sikre at $\pi(x)$ holder seg innenfor intervallet $[0,1]$, kan vi ikke formulere modellen som en lineær funksjon slik som for den lineære regresjonsmodellen. En logistisk regresjon svarer til at vi benytter følgende funksjonsform:

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$

$$\pi(x) = \frac{e^{\alpha+\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\alpha+\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

Det foreligger en rekke forutsetninger i en slik modell. For det første må, som nevnt, Y være dikotom, altså enten 0 eller 1. De uavhengige variablene må være kontinuerlige eller kategoriske, samt uavhengige av hverandre.

2.4.4 Naive Bayes Classification

Naive Bayes Classification er en maskinlæringsalgoritme som brukes til klassifiseringsoppgaver. Metoden kan beskrives som en sannsynlighetsrelatert klassifiseringsmetode med tanke på at den er basert på Bayes' teorem. Teoremet, også kalt Bayes regel, gir oss muligheten til å beregne betingede sannsynligheter. Med andre ord kan man regne ut sannsynligheten for at en hendelse finner sted, gitt at en annen hendelse allerede har funnet sted. Dette gjelder også i motsatt tilfelle (IBM, 2023). Beregningen er gitt ved formelen:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Naive Bayes skiller seg fra andre maskinlæringsalgoritmer i den forstand at den opererer med et par antagelser. For det første tar modellen høyde for at variabler ikke relaterer eller avhenger av hverandre. Med andre ord er teorien bak modellen at alle input-variabler er uavhengige av hverandre. Den andre antagelsen handler om at alle variabler bidrar like mye til resultatet. Disse antagelsene vil derimot ikke fungere like godt i praksis, da variabler sjeldent er utelukkende uavhengige eller er like store bidragsyttere. Av den grunn beskrives modellen gjerne som naiv. Likevel er modellen i stand til å forenkle klassifiseringsoppgaver ved å gjøre oppgaven mer beregningsmessig håndterbar (IBM, 2023). Med tanke på at modellen er basert på en svært enkel algoritme, er den i stand til å gjøre raske beregninger på større datasett, hvor resultatene som fremkommer ofte er av høy treffsikkerhet.

2.5 Korrelasjon

Et nyttig statistisk mål for samvariasjon mellom variabler er korrelasjon. Korrelasjon viser hvordan to variabler beveger seg i forhold til hverandre, og måler altså graden av lineær sammenheng mellom en variabel X og en variabel Y . Det poengteres at korrelasjon ikke er det samme som kausalitet, og vil således ikke avdekke en årsakssammenheng (Finanssenteret, 2023). Sammenhengen mellom to variabler måles av korrelasjonskoeffisienten, og denne koeffisienten varierer mellom -1 og 1. Formelen betegnes slik:

$$Corr[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X] \cdot Var[Y]}}$$

Dersom korrelasjonskoeffisienten er lik -1 eller 1, angir dette en perfekt lineær sammenheng mellom X og Y:

$$\text{Corr}[X, Y] = \{-1, 1\}$$

Dersom korrelasjonskoeffisienten er lik 0, forteller dette at det er ingen lineær sammenheng mellom X og Y:

$$\text{Corr}[X, Y] = 0$$

Tallverdien til korrelasjonen i hvor stor grad X og Y går i takt eller utakt. Dersom $\text{Corr}[X, Y] > 0$, tenderer X og Y til å gå i takt. Dersom $\text{Corr}[X, Y] < 0$, tenderer X og Y til å gå i utakt (NTNU, 2016).

3. DATA OG METODE

3.1 Datasettet

Datasettet består av de variabler som er forutsatt å ha påvirkning på strømprisen (se seksjon 2.3). Vi har for enkelhets skyld valgt en tidsramme på ti år, fra 01.01.2013 til 31.12.2022. Data er innhentet fra Eikons databaser, Noregs Vassdrags- og energidirektorat, og Norsk Klimaservicesenter. Tabellen under viser de valgte variablene og deres respektive datakilder.

<i>Variabel</i>	<i>Beskrivelse</i>	<i>Kilde</i>
Spotpris elektrisitet NO1	€/MWh	Nord Pool
Oljepris - Brent oil	USD/BBL	EIA
Gasspris - TTF	€/MWh	Refinitiv Eikon
Valutakurs - USD til EUR	USD/EUR	Refinitiv Eikon
Valutakurs - NOK til EUR	NOK/EUR	Refinitiv Eikon
Valutakurs - NOK til USD	NOK/USD	Refinitiv Eikon
CO2 kvotepris - EEX - EU	€/Metric Tonne	EEX
Kullpris - ICE gC Newc	\$/Metric Tonne	ICE
Norges Energiforbruk	GWh - monthly	Eurostat
Import elektrisitet	MNOK - quarterly	SSB
Eksport elektrisitet	MNOK - quarterly	SSB
Middeltemperatur	Celsius	Klimaservicesenter
Nedbør	Millimeter	Klimaservicesenter
Fyllingsgrad - NO1	$\frac{\sum E_{M i, uke j}}{\sum E_{M i, maks}}$	NVE
Fyllingsgrad - NORGE	$\frac{\sum E_{M i, uke j}}{\sum E_{M i, maks}}$	NVE

Tabell 1: Oversikt over variabler og tilhørende datakilde

NVE: Norges vassdrag- og energidirektorat

EIA: Energy Information Administration, United States

ICE: Intercontinental exchange

EEX: European energy exchange

E_M: Energiinnhold, GWh

3.1.1 Avhengig variabel

Spotpris elektrisitet NOI

Spotprisen er prisen på et gitt tidspunkt som strømmen blir solgt eller kjøpt for. Dette er den faktiske prisen på strømbørsen Nord Pool. Spotprisen kjøpes og selges i euro, og måles opp mot megawattimer.

3.1.2 Uavhengige variabler

Oljepris - Brent oil

Oljepris av typen *Brent crude*. Dette er oljen vi refererer til som Nordsjøoljen, og benyttes som referanseindeks for Europa, Afrika og Midtøsten. Måles i dollar per tønne.

Gasspris - TTF

Prisen på europeisk gass, målt ved den nederlandske referanseindeksen Dutch TTF Natural Gas. Gassen måles i euro per megawattime.

Valutakurser

Forholdet mellom amerikanske dollar og euro, norske kroner og euro, og norske kroner og amerikanske dollar.

CO2 kvotepris - EEX - EU

Prisen på klimakvoter. Måles i euro per tonn med CO₂-ekvivalenter.

Kullpris - ICE gc Newc

Prisen på kull, målt i dollar per tonn, ved referanseindeksen i Newcastle.

Norges Energiforbruk

Energiforbruket på landsbasis målt i gigawattimer. Tallene er på månedlig basis.

Import elektrisitet

Import av elektrisitet til Norge, målt i millioner NOK. Tall på kvartalsvis basis.

Eksport elektrisitet

Eksport av elektrisitet fra Norge, målt i millioner NOK. Tall på kvartalsvis basis.

Middeltemperatur

Middeltemperatur målt ved Blindern Værstasjon i Oslo. Et mål på gjennomsnittstemperaturen i løpet av et døgn. Måles i grader Celsius.

Nedbør

Nedbørsmengde Blindern Værstasjon i Oslo. Målt i millimeter.

Fyllingsgrad NO1/NORGE

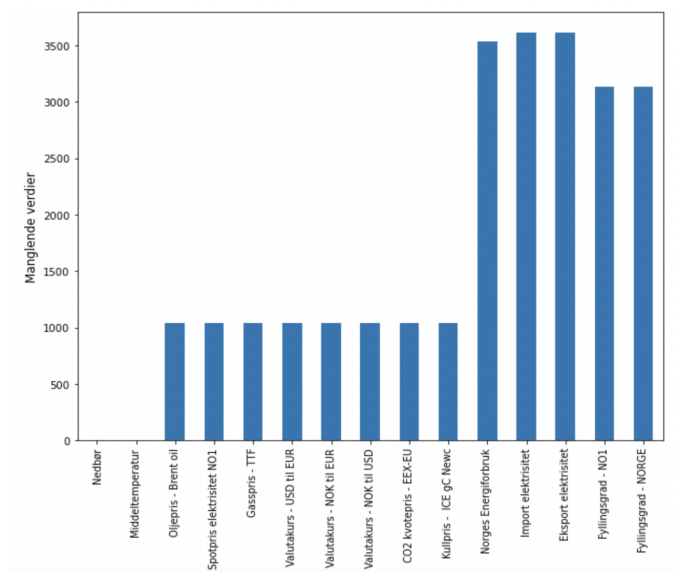
Vannvolum i magasinene omregnes til energiinnhold. Fyllingsgraden er energiinnholdet i magasinene delt på energiinnholdet ved fulle magasiner. Energiinnholdet er i terawattimer. *Fyllingsgrad NO1* referer til fyllingsgraden på østlandet, mens *Fyllingsgrad NORGE* er fyllingsgraden til samtlige magasiner i landet.

3.2 Databehandling

3.2.1 Manglende verdier

Et ideelt datasett ville hatt verdier for alle dager i den valgte sample-perioden. En utfordring med behandlingen av datasettet er imidlertid hvordan vi forholder oss til manglende verdier (fra engelsk: *missing values*). Manglende verdier i datasettet kan føre til mindre presise prediksjoner og en lavere forklaringskraft i modellen. Flere av variablene har store mengder manglende verdier. Siden vi opererer på dagsbasis i sample-perioden er vi avhengig av at variablene oppdateres daglig. Dette er imidlertid kun tilfelle for variablene *Nedbør* og *Middeltemperatur*. Sistnevnte har også én manglende verdi, da det mangler data for 01.05.2015.

Norges Energiforbruk har månedlige verdier, som oppdateres den 15. i hver måned. Det vil si at for resterende dager i datasettet vil variabelen ha manglende verdier. Det må også nevnes at *Norges Energiforbruk* mangler data for de to siste månedene i 2022. Tilsvarende har *Import elektrisitet* og *eksport elektrisitet* data for hvert kvartal, og oppdateres den 15. hver februar, mai, august og november. Disse variablene mangler data for november 2022. NVE utfører vannstandsmåling hver mandag, så når det kommer til *Fyllingsgrad - NO1* og *Fyllingsgrad NORGE*, har vi kun verdier for mandager, altså ukentlig. De resterende variablene har kun data for hverdager, og har således manglende verdier for helgedager.



Figur 3.1: Manglende verdier

Spørsmålet er hvordan vi tar hensyn til manglende verdier i datasettet. En utbredt praksis er å simpelthen fjerne disse observasjonene fra datasettet. Dette vil være svært problematisk, da det i vesentlig grad reduserer antall observasjoner vi har å jobbe med. Det er totalt 3652 dager/observasjoner i sample-perioden vår. Skulle vi fjernet alle dager med manglende verdier vil datasettet reduseres til kun 40 dager/observasjoner. Dette er en betydelig reduksjon, og et datasett med kun 40 observasjoner vil kunne påvirke modellens forklaringskraft og reliabilitet negativt. Alternativt kan variablene med store mengder manglende verdier fjernes totalt fra datasettet, slik at vi kun behøver å ta hensyn til manglende verdier for variablene med manglende data i helgedager. Dette er heller ikke ønskelig, da vi mister potensielt viktige forklaringsvariabler. I tillegg vil det å fjerne manglende verdier for variabler uten verdier for helgedager ha en uheldig effekt, da det fører til et datasett med totalt 1043 færre observasjoner.

For å oppnå et datasett med flest mulig observasjoner, samt inkludere alle utvalgte variabler, har vi benyttet oss av metoden *Forward Fill*. I praksis løses dette med kommandoen `fillna(method='ffill')` i Python. I denne metoden byttes manglende verdier ut med den siste observerte verdien. For eksempel vil variabler som har manglende verdier for helgedager få disse byttet ut med siste observerte verdi, som er fredag. Selv om dette fører til et komplett datasett uten manglende verdier, er det viktig å reflektere rundt hvordan denne operasjonen kan påvirke modellene. Særlig vil variablene som oppdateres kvartalsvis og månedlig ha mange like verdier, noe som kan ha stor påvirkningskraft på modellen.

3.2.2 Omgjøring av variabler

En utfordring knyttet til dataen er at flere av variablene har hatt en omfattende strukturell endring de siste par årene, noe som kan gjøre det mer problematisk å predikere. Spesielt var 2022 et spesielt år sett i historisk sammenheng, med blant annet svært høye strømpriser. Faren er at man ikke har nok historie på denne dataen med ekstreme strømpriser til at modellen finner et representativt mønster i tidsserien. Vi tenker dette potensielt kan løses på to måter. Én måte er å fjerne dataen for 2022, hvor strømprisene er på sitt mest ekstreme. Dette gjør dog at viktige momenter ved problemstillingen faller bort. En annen, mer respektabel tilnærming, er å teste dataen på relative endringer i stedet for absolutte prisendringer. Dette er imidlertid heller ikke problemfritt, da grafen over spotprisen tyder på at volatiliteten også har økt de siste par årene. Etter gjentatte prøvelser viser det seg at ingen av tilnærmingene gir særlig gode tester. Det er i hovedsak de geopolitiske hendelsene og ringvirkningene av disse som gjør det vanskelig å prognostisere dataen. Vi velger på bakgrunn av dette å beholde variablene slik de opprinnelig var, fremfor å enten fjerne data eller omgjøre variabler til relative endringer.

3.3 Konstruksjon av modeller

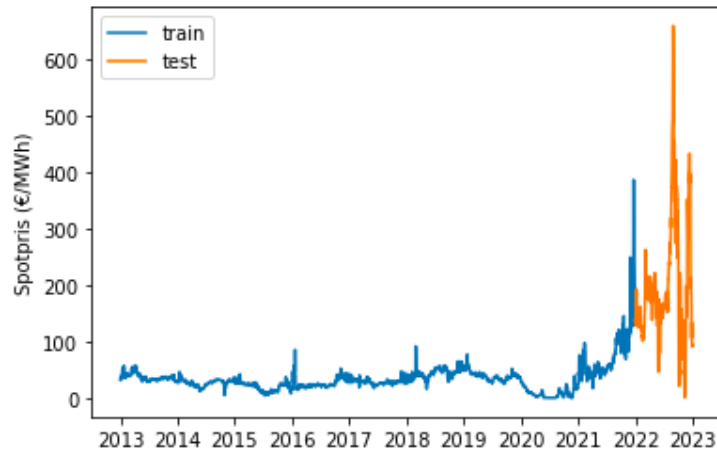
3.3.1 Lineær regresjon

Oppbygging

I oppgaven har vi konstruert en lineær regresjonsmodell ved hjelp av biblioteket *statsmodels*. Dette biblioteket gir klasser og funksjoner for estimering av mange ulike statistiske modeller, blant dem lineær regresjon. Ved hjelp av denne modulen kan vi utforske statistiske data på en svært effektiv måte. Metoden som benyttes for å estimere regresjonslinjen er OLS (se seksjon 2.4.2).

Inndeling

Det totale datasettet på 10 år deles opp i et treningssett og testsett, slik at vi kan trene en modell basert på dataen vår. I vårt tilfelle har vi valgt å dele opp dataen slik at treningsdelen består av de 9 første årene med data, mens det siste året (2022) utgjør testdataen. Hensikten bak inndelingen er å inkludere en del av de store prisendringene fra 2020 og 2021 i modellen vår. Tanken med dette er at dersom modellen trenes på mer variert data, vil den forhåpentligvis også fungere bedre på nye observasjoner. Ulempen med denne inndelingen er imidlertid at testdelen blir liten i forhold til treningsdelen. Dette illustreres i figur 3.2.



Figur 3.2: Oppdeling av dataen

Vurdering av modellen

Regresjonsmodellen vurderes utelukkende basert på R^2 -verdien (fra engelsk: *R-squared*). Ved å bruke *statsmodels* kan vi ta i bruk kommandoen *model.summary()*, hvor vi får en oppsummerende statistikk over modellen. Samtidig vil det bli relevant å se på p-verdiene for de uavhengige variablene for å evaluere om disse er signifikante. Likevel er kanskje det vanligste målet for modellens prestasjon å se på det kvadratiske gjennomsnittet av feilene, MSE (fra engelsk: *mean squared errors*). Denne verdien vil gi oss en pekepinn på hvor feilaktig modellen predikerer sammenlignet med faktiske observerte verdier. Dersom verdien er høy vil det indikere en svak modell, mens en lav verdi på MSE vil indikere en bedre modell.

Forutsetninger for bruk av lineær regresjon

Lineær regresjon er en kraftig analyseteknikk, som i mange tilfeller også er robust. Det finnes likevel noen forutsetninger for at regresjonsanalysen skal være pålitelig. Forutsetningene kan oppsummeres i fem punkter (Jacobsen, 2015):

1. Lineær sammenheng

Det skal være lineær og additiv sammenheng mellom den avhengige og de uavhengige variablene i populasjonen. En lineær og additiv sammenheng vil i dette tilfellet bety at endringer i de uavhengige variablene vil gjenspeiles i den avhengige variabelen. På denne måten vil man kunne produsere forskjellige resultater med en lineær sammenheng for nye observasjoner. Sagt på en enkel måte vil den avhengige variabelen økes/reduceres med tilsvarende verdi av endringen i en uavhengig variabel.

2. *Ingen multikollinearitet*

I en multippel regresjonsanalyse forutsettes det at det ikke er en perfekt eller tilnærmet perfekt sammenheng mellom to eller flere av de uavhengige variablene. Dersom dette er tilfelle, oppstår det multikollinearitet. En tommelfingerregel viser til at dersom man kan måle en korrelasjonsverdi på over 0,7 (pearsons r) mellom to uavhengige variabler, er det fare for multikollinearitet. Grunnen til at man ikke ønsker for høy samvariasjon mellom to uavhengige variabler er fordi dette vil føre til at standardfeilen til parameterestimatene øker. Resultatet av dette vil også føre til at T-verdien endres.

3. *Uavhengighet mellom variablene*

Modellen forutsetter at observasjonene er uavhengige av hverandre. Med andre ord forutsetter modellen at verdiene for residualene er uavhengige.

4. *Homoskedastisitet*

I regresjonsanalysen ønsker vi at feilleddene har konstant variasjon rundt regresjonslinjen. Observasjonene bør altså være spredt jevnt rundt regresjonslinjen i populasjonen, og variasjonen skal ikke variere med ulike verdier på de uavhengige variablene. Dette omtales som homoskedastisitet. Det motsatte tilfellet, heteroskedastisitet, kan bety at det for eksempel er liten variasjon i spredningen rundt lavere verdier, mens variasjonen i spredningen øker med høyere verdier.

5. *Feiltermene i regresjonen er normalfordelte*

Dersom vi har normalfordelte feiltermene i regresjonen, får vi en normalfordeling hvis vi lager et histogram over avvikene fra regresjonslinjen i populasjonen. Dersom denne forutsetningen foreligger, burde normalt også residualene være normalfordelte for tilfeldige utvalg.

3.3.2 Naive Bayes Classification

Oppbygging og inndeling

For å kunne konstruere en Naive Bayes modell er vi nødt til å importere funksjonen *GaussianNB* fra biblioteket *scikit-learn*. Med tanke på at denne modellen er basert på en klassifiseringsalgoritme (Bayes teorem), deler vi opp strømprisen i ulike klasser, eller kategorier. For en sluttforbruker vil det kunne være av like stor interesse å vite i hvilket sjikt strømprisen vil ligge enn å vite den nøyaktige prisen. Vi kategoriserer spotprisen (avhengig

variabel) ut ifra disse nivåene: lav pris, middels pris, høy pris og ekstremt høy pris. Vi tok utgangspunkt i medianen på spotprisen når vi kategoriserte prisen. Middels kategori satte vi i intervallet +/- 10 €/MWh av medianen. Alt under dette kategoriseres til lav pris. Siden intervallet over +10 €/MWh av medianen blir såpass stort, så vi det nødvendig å dele dette inn i to kategorier. Høy pris er alt over middels pris opptil det dobbelte av medianen. Ekstrem høy pris er resterende verdier. Vi fikk dermed følgende kategorier (målt i €/MWh):

- Lav spotpris: 0 - 23
- Middels spotpris: 23 - 43
- Høy spotpris: 43 - 86
- Ekstremt høy spotpris: 86 - 661

Denne nye variabelen med priskategoriene har vi gitt navnet “Price_catagory”. Målet med modellen er at man ved hjelp av de uavhengige variablene klarer å avdekke på hvilket nivå man kan kategorisere spotprisen. For å kunne gjøre dette må den kategoriske variabelen gjøres om til en numerisk variabel. Dette løses ved å ta i bruk *label encoding* i Python. Denne teknikken gjør om kategoriene til tall fra 0-3. “Ekstremt høy” representeres med tallet 0, “Høy” med 1, “Middels” med 3 og “Lav” med 2.

På samme måte som ved lineær regresjon har vi delt opp datasettet i en treningsdel og en testdel. Treningssettet består av de 9 første årene, mens testsettet består av året 2022.

Vurdering av modellen

For å vurdere modellen bruker vi kommandoen *model.score()*, som representerer treffsikkerheten knyttet til å kategorisere til riktig kategori. Denne kommandoen brukes i hovedsak på testsettet, men kan også være relevant å bruke på treningssettet. For å gi et bedre bilde av modellens treffsikkerhet er det vanlig å visualisere modellens ytelse i et fargekart (fra engelsk: *heatmap*).

3.3.3 Logistisk regresjon

Oppbygging og inndeling

Vi bygger en logistisk regresjonsmodell for å predikere om strømprisen går opp eller ned. For både forbrukere og produsenter vil det være av interesse å potensielt kunne vite hvilken retning strømprisen vil bevege seg i fremtiden, uansett om dette ikke forteller hvor mye

prisen endres. Det første vi gjør er å lage en ny variabel fra “Spotpris elektrisitet NO1” som angir den prosentvise endringen dag til dag. Denne variabelen kaller vi “Spotpris_pct”. Deretter definerer vi en ny dikotom variabel, “Direction”, som angir retningen på den prosentvise endringen; 0 for negativ prosentvis endring, og 1 for positiv endring.

Y-verdien vår blir dermed “Direction”, mens X-verdiene er de uavhengige variablene. Inndelingen i et treningssett og testsett skjer på samme måte som ved de foregående modellene.

Vurdering av modellen

Vi vurderer modellen på samme måte som Naive Bayes-modellen; med kommandoen `model.score()`. Samtidig har vi valgt å visualisere modellens resultater i et fargekart.

3.3.4 ANN

Vi modellerer det nevralt nettverket ved hjelp av Keras-biblioteket i Python. Keras tillater interaksjon med dype nevralt nettverk og maskinlæringsalgoritmer. Vi kjører Keras ved hjelp av *Tensorflow*, et open-source bibliotek, for å trene maskinlæringsmodeller. Keras er API-en som forenkler denne prosessen (Team, 2023).

Dataprosessering

Ved ANN velger vi også å dele inn i et treningssett bestående av de 9 første årene i datasettet vårt, mens testsettet består av året 2022. Dataen som vi har samlet inn har en rekke ulike måleenheter, som for eksempel valutakurser, euro per gigawattimer, og nedbørsmengder. De forskjellige måleenhetene gjør at vi får variabler som inneholder veldig høye verdier, mens andre variabler har relativt mindre verdier. Av den grunn ser vi det hensiktsmessig å skalere dataen ved hjelp av funksjonen *MinMaxScaler*. På denne måten vil vi kunne endre høye og lave verdier til forholdstall som befinner seg mellom 0 og 1.

Input og layers

Modellens input-dimensjon spesifiseres først i modellen, og representerer alle de 14 uavhengige variablene. Med andre ord bygges modellen opp med en vektor med størrelsen 1x14. Deretter er modellen bygget opp av to skjulte lag, som består av hhv. 25 og 20 noder. Disse skjulte lagene genererer en output-verdi gjennom hver enkelt vektning sammen med

aktiveringsfunksjonen *tanh*. Output-verdien er spesifisert til å bestå av kun én node, med hensikt at output-verdien skal kunne vise den predikerte spotprisen.

Epochs og early stoppage

I modellen implementerer vi funksjonen *EarlyStoppage* med hensikt å forhindre at modellen blir tilpasset treningsdataen for godt (fra engelsk: *overfitted model*). Samtidig er målet å minimere det totale tapet for å skape en velfungerende modell. Dersom modellen blir for godt tilpasset treningsdataen, vil den predikere svakere på nye observasjoner i form av større avvik. Denne funksjonen vil derfor bidra til at modellens treningsprosess vil stoppes når minimeringen av tap avtar. Treningsprosessen fungerer ved at modellen itererer over hele dataen, hvor det totale tapet registreres etter iterasjonen. I funksjonen spesifiserer vi at dersom tapet ikke bedres i løpet av to iterasjoner, så vil treningsprosessen stoppe.

Loss function

Minimering av tapet måles ved hjelp av en tapsfunksjon. Denne funksjonen kvantifiserer forskjellen mellom modellens output-verdi, og den faktiske observasjonen. Vi benytter *mean squared error* som tapsfunksjon:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\widehat{Y}_i - Y_i)^2}{n}}$$

Y_i : faktisk observasjon

\widehat{Y}_i : predikert verdi

n: antall observasjoner

En fare ved utarbeidelsen av nevralt nettverk er at modellen blir *underfit* eller *overfit*. Et overfit datasett vil gjenkjenne trenden i dataen den er trent på for godt, og vil således ikke prestere godt på ny, usett data. Underfit vil si at modellen ikke evner å prestere godt på ny data. Sistnevnte oppstår ofte når modellen er for enkel og ikke evner å skape en klar relasjon

mellom input- og outputverdien, noe som bidrar til større feilverdier på testsettet enn treningsettet. (Baheti, 2023).

Optimizer

En *optimizer* kan beskrives som en algoritme som bidrar til å redusere modellens tap ved å endre enkelte attributter som modellens vektning og læringsrate. Vi har valgt å bruke typen *adam*, som er pålitelig og godt egnet til å operere på større datasett.

Aktiveringsfunksjon

Aktiveringsfunksjonen som brukes i nettverket er *tanh* (fra engelsk: *hyperbolic tangent function*), og er gitt ved:

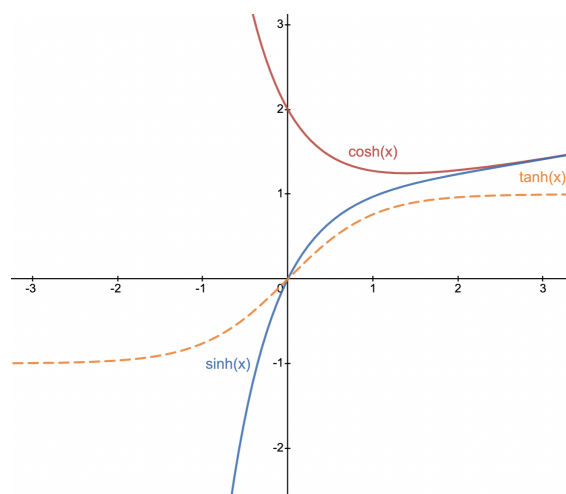
$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

, hvor *sinh* og *cosh* er henholdsvis hyperbolske funksjoner av *sinus* og *cosinus* (Wikipedia, 2023).

$$\lim_{x \rightarrow \infty} \tanh(x) \rightarrow \tanh(x) = \frac{\infty - 0}{\infty + 0} = \frac{\infty}{\infty} = 1$$

$$\lim_{x \rightarrow -\infty} \tanh(x) \rightarrow \tanh(x) = \frac{0 - \infty}{0 + \infty} = \frac{-\infty}{\infty} = -1$$

Dette viser at $\tanh(x)$ har et mulighetsområde tilsvarende $[-1, 1]$.



Figur 3.3: Funksjonen $\tanh(x)$

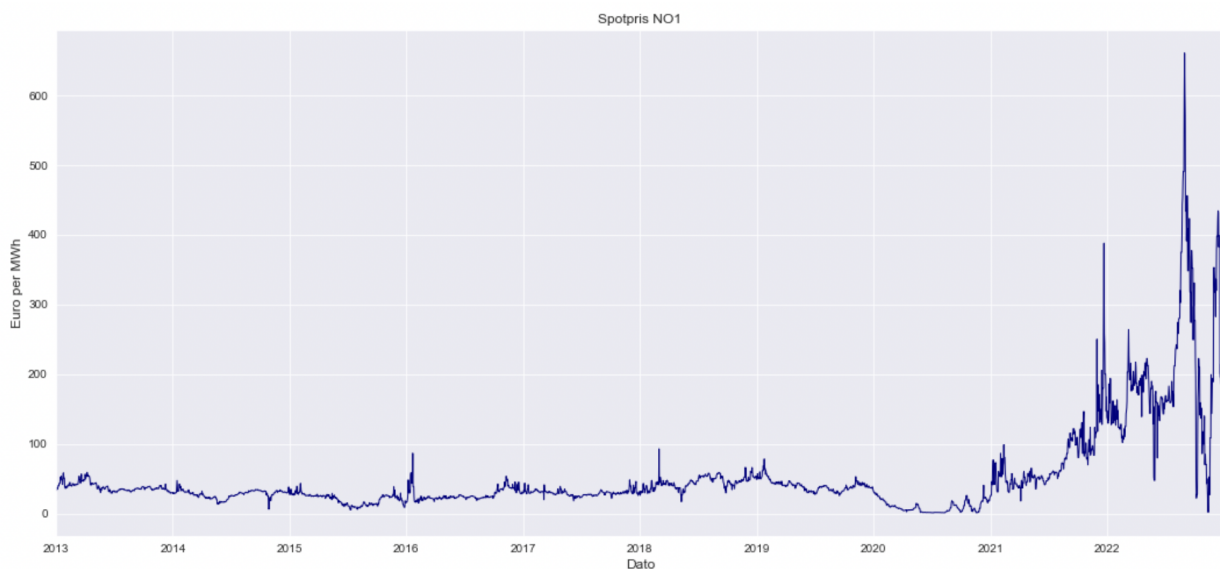
En fordel ved bruk av denne algoritmen er at output-verdien av tanh-funksjonen er sentrert mot null som tilsvarer at den både omfavner negative og positive verdier. Som følge av dette kan vi enkelt klassifisere output-verdien som sterk negativ, nøytral, eller sterk positiv (Baheti, 2022).

4. ANALYSE OG RESULTAT

4.1 Visualisering og deskriptiv statistikk

4.1.1 Avhengig variabel

Spotpris elektrisitet NO1



Figur 4.1: Spotpris elektrisitet NO1 i perioden 2013 - 2023

Av grafen kan vi se at strømprisen har holdt seg forholdsvis stabil i perioden 2013-2019. I 2020 kan det virke som om prisen lå på et svært lavt nivå, før den gjorde et massivt byks fra 2021. I 2022 kan det se ut som om verdien på et tidspunkt var seksdoblet sammenlignet med nivåer fra 2021. Dette kommer også tydelig fram av tabellen under (tabell 2). Der ser vi at gjennomsnittsprisen har ligget på et relativt stabilt nivå fra årene 2013-2019, mens den i 2020 var sammenlignbart svært lav, før den så stiger kraftig i årene 2021 og 2022. Som grafen også antyder, er det store svingninger i prisene i 2021 og 2022. Gapet mellom høyeste og laveste pris er stort sammenlignet med de andre årene. Store svingninger i prisen gjør at standardavviket blir forholdsvis stort de to siste årene.

<i>ÅRSTALL</i>	<i>MEAN</i>	<i>MAX</i>	<i>MIN</i>	<i>STD</i>
2013	38,16	58,84	28,93	6,17
2014	27,69	47,40	6,11	5,62
2015	20,38	43,11	5,42	7,45
2016	26,92	86,39	14,9	8,45
2017	29,49	48,10	18,05	3,83
2018	44,41	92,70	16,96	8,47
2019	39,87	78,13	25,67	7,73
2020	9,58	40,46	0,91	8,03
2021	76,25	387,45	17,89	43,88
2022	199,49	660,06	1,93	105,87

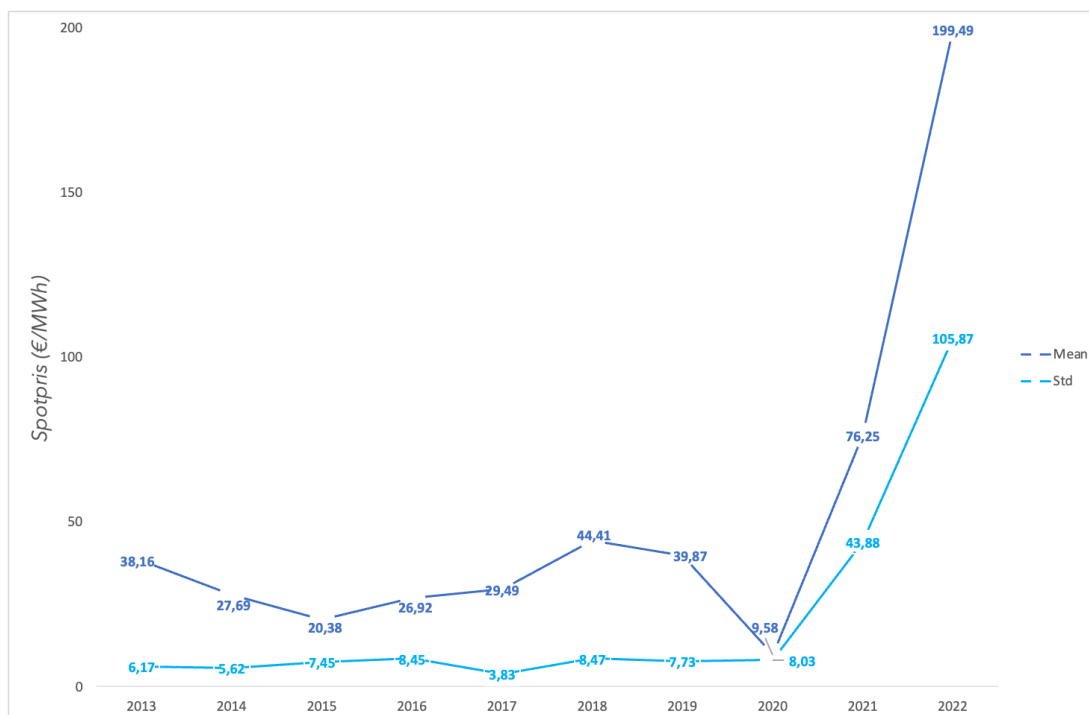
Tabell 2: Oversikt over gjennomsnittsverdier, maksimums- og minimumsverdier, og standardavvik for strømprisen

MEAN: Gjennomsnittlig strømpris

MAX: Maksverdien, høyeste strømprisen

MIN: Minimumsverdien, laveste strømprisen

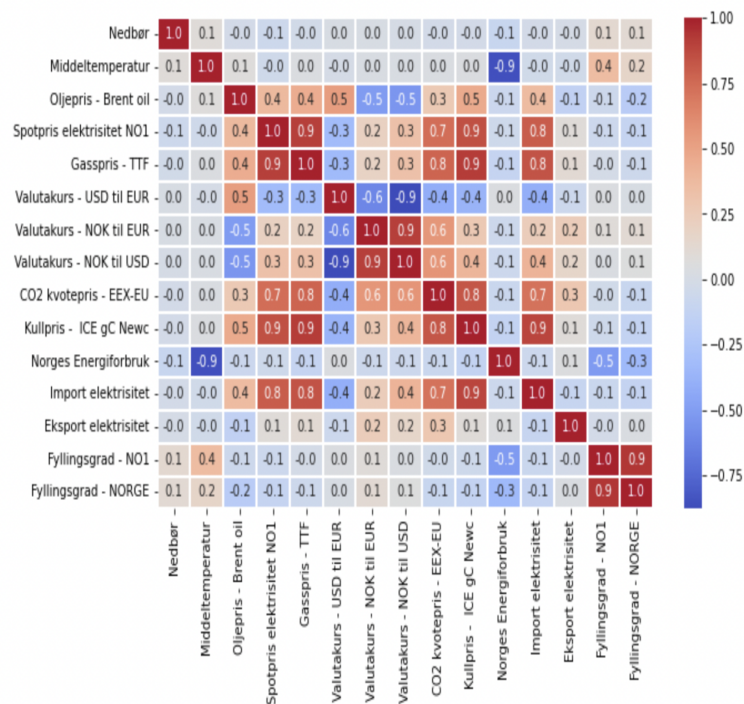
STD: Standardavviket



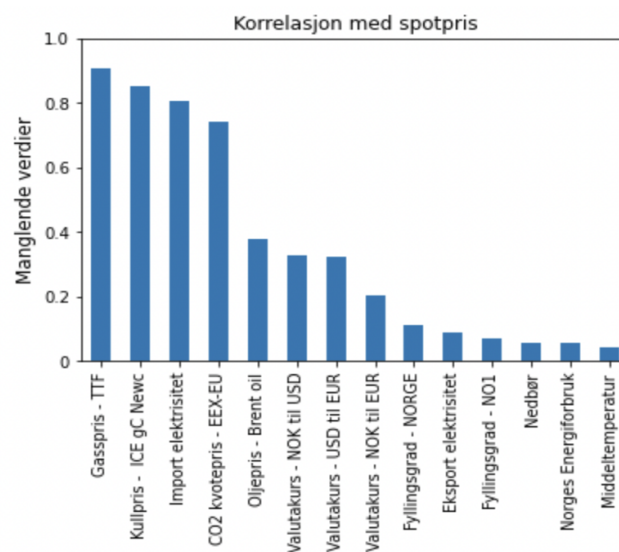
Figur 4.2: Gjennomsnittlig strømpris og standardavvik i perioden 2013-2022

4.1.2 Uavhengige variabler

Vi ønsker å undersøke, hver for seg, forholdet mellom de uavhengige variablene og spotprisen. Dette har vi illustrert grafisk ved hjelp av en korrelasjonsmatrise, samt et stolpediagram som viser de uavhengige variablenes korrelasjon med spotprisen i synkende rekkefølge.



Figur 4.3: Korrelasjonsmatrise

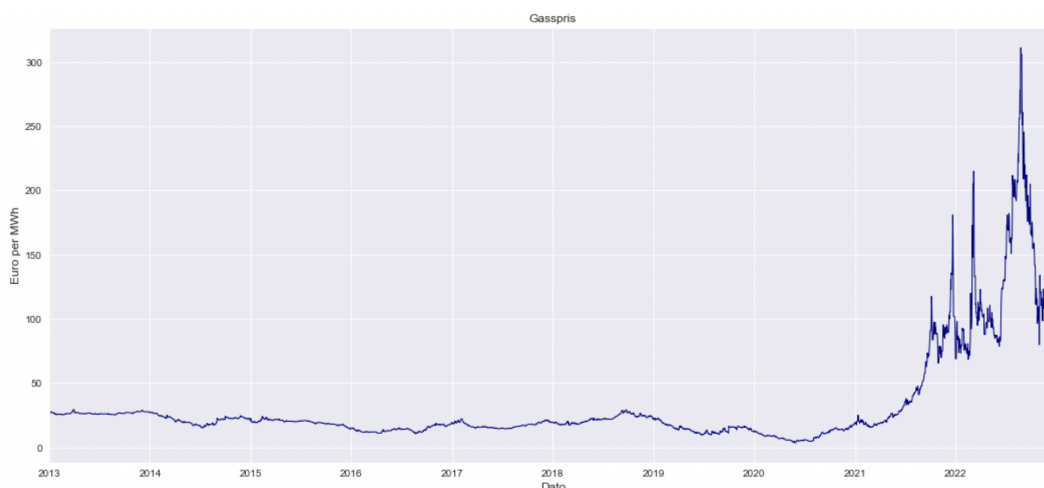


Figur 4.4: Absolutt korrelasjon med spotpris i synkende rekkefølge

Selv om korrelasjon ikke måler årsakssammenheng, kan tallene likevel gi et nyttig innblikk i de ulike variablene. Av figur 4.4 ser man at *gasspris*, *kullpris*, *import av elektrisitet* og *CO₂ kvotepris* er de fire variablene som korrelerer mest med spotprisen for strøm. Siden gass og kull brukes til produksjon av elektrisitet, eller som en substitusjon for elektrisk kraft, er det ikke overraskende at man finner en sterk sammenheng mellom de to variablene og spotprisen for elektrisitet. Siden høye priser i et område fører til import av elektrisitet fra områder med lavere priser, og omvendt, er det heller ikke oppsiktsvekkende at man finner en sterk korrelasjon her. Når det kommer til kvotepris for CO₂ vet vi at når prisen på kvotene går opp vil strømproduksjonen ved kull- og gasskraftverk bli dyrere, og dermed påvirke strømprisen i samme retning. CO₂-kvoter kan derfor ses på som en skattlegging av gassutslipp, og på den måten vil prisene på disse redusere etterspørselen etter gass.

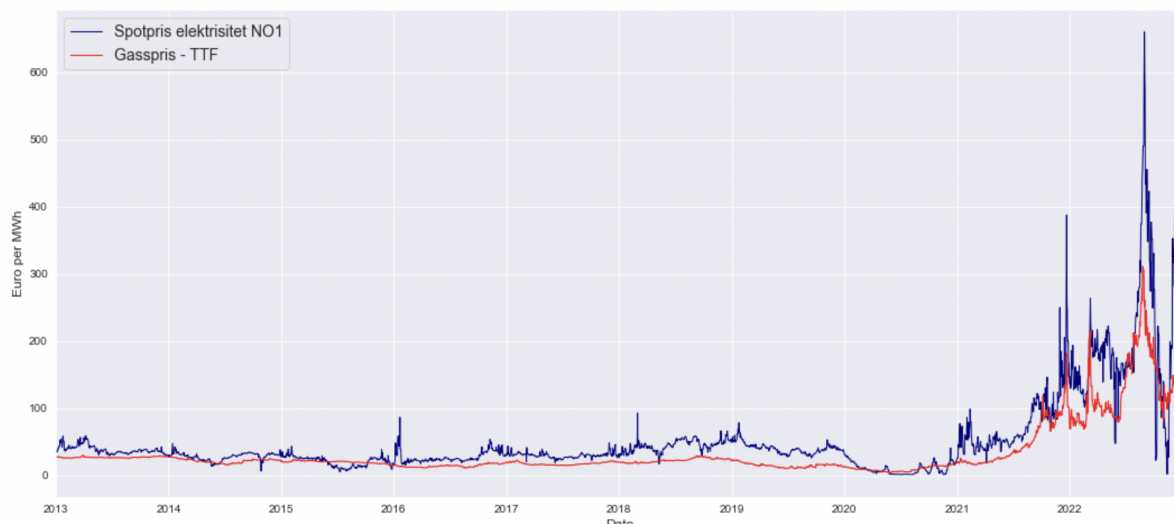
Verdt å merke seg er også at variablene *Middeltemperatur* og *Norges Energiforbruk* har en utpreget stor korrelasjon mellom seg. Som nevnt tidligere vil lave temperaturer føre til større energiforbruk på grunn av økt nødvendighet for oppvarming, og korrelasjonen er dermed logisk. Når det kommer til de fire variablene som korrelerer mest med strømprisen, ser man også en høy korrelasjon mellom disse. På motsatt side av skalaen er det mest overraskende resultatet at man finner en merkverdig liten korrelasjon mellom nedbør og fyllingsgrad. Det ville vært nærliggende å anta at økt nedbør ville medføre økt fyllingsgrad i magasinene.

Gasspris - TTF



Figur 4.5: Gasspris i perioden 2013 - 2023

Slik som spotprisen, holdt gassprisen seg på et stabilt nivå i årene 2013-2019 og nådde et lavere prisnivå i midten av 2020. I løpet av 2021 steg gassprisen betraktelig og 2022 var preget av store svingninger i prisen, med et historisk høyt toppnivå i midten av året. Mot siste halvdel av 2022 falt prisen igjen nærmere mot 2021-prisnivået. En del av årsaken til særlig høye gasspriser i 2022 kan forklares av Russlands invasjon av Ukraina. Europa er nemlig storforbruker av russisk gass. I slutten av 2021 stod russisk gass for rundt 40% av Europas gassimport. I utgangen av 2022 var importen av russisk gass betraktelig redusert, og stod kun for litt over 10% av samlet europeisk gassimport (European Council, 2023). Stopp i leveranser fra Russland gjennom Nord Stream-ledningene førte til at tilbudet på gass sank. Etterspørselen etter gass var fortsatt den samme, og således fikk man en økning i prisen (Delebekk, 2023). Det store fallet i gassprisen mot slutten av 2022 kom av ekstraordinære milde temperaturer i Europa. I tillegg observerte man en rekordhøy import av den flytende gassformen LNG, som senket etterspørselen etter gass til oppvarming (Nilsen, 2022).

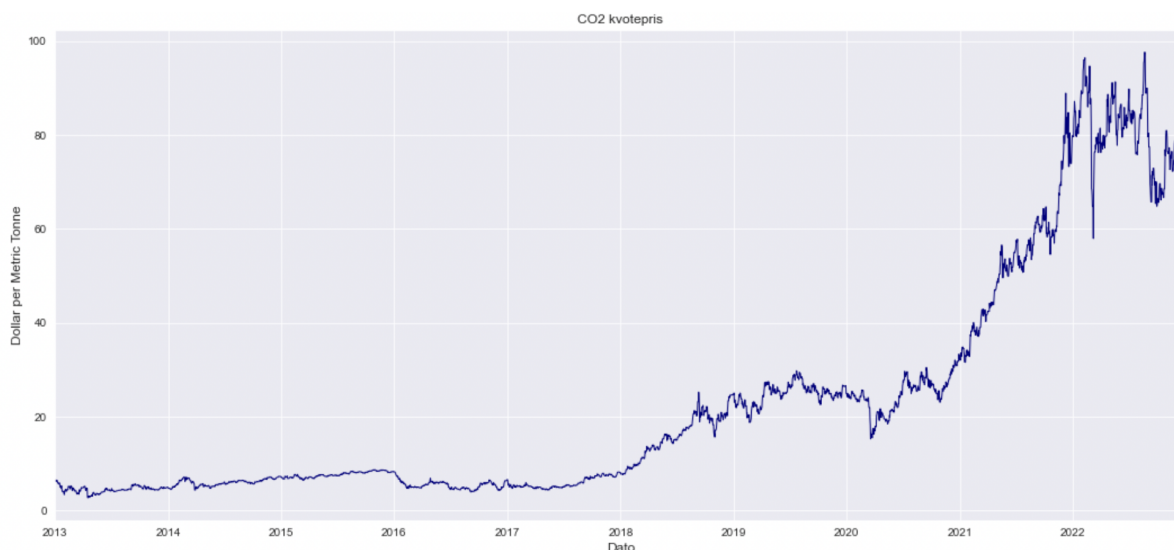


Figur 4.6: Sammenligning spotpris, elektrisitet og gasspris i perioden 2013-2022

Gassprisen var den variabelen som korrelerte mest med spotprisen for elektrisitet.

Sammenligner man grafene til de to prisene ser man store likheter. Prisen på både gass og elektrisitet holdt seg på et stabilt nivå frem til 2021, og de to grafene følger hverandre tett. I løpet av 2021 og 2022 er det store svingninger i prisene, men man ser tydelig at prisen på elektrisitet når høyere toppnivåer. Selv om man fortsatt ser en viss sammenheng mellom de to prisene, blir det for enkelt å kun skylde på økte gasspriser som en årsak til økte strømpriser.

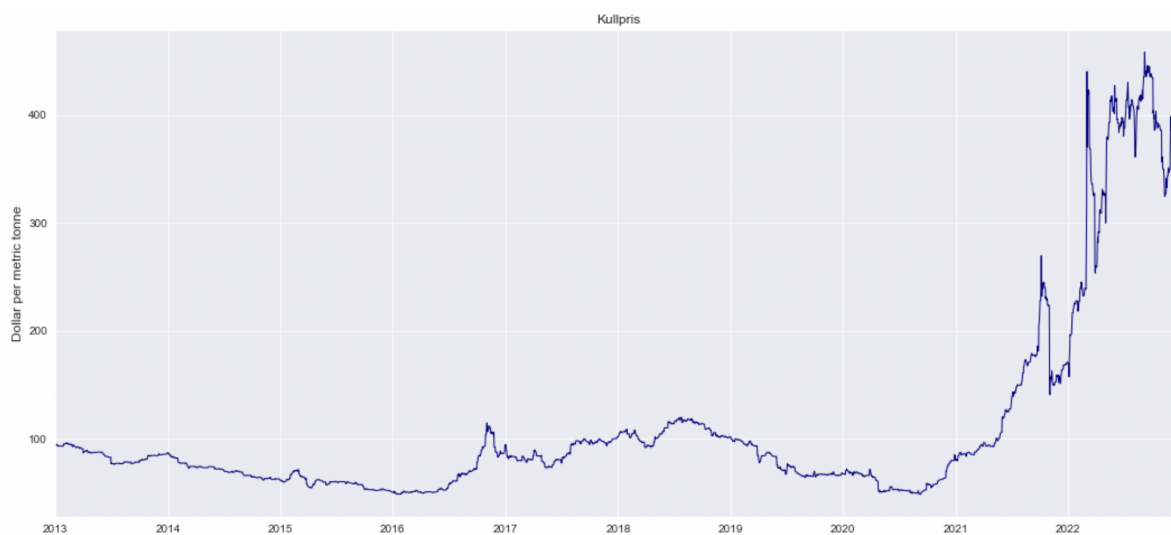
CO2 kvotepris - EEX - EU



Figur 4.7: Kvotepris på CO₂ i perioden 2013 - 2023

Prisen på CO₂-kvoter har fulgt en relativt stigende kurve siden 2018. Det er imidlertid først i 2021 man ser en markant stigning i prisen. Dette kommer rundt samme tidspunkt man ser en økning i prisen på elektrisitet. Volatiliteten ser også ut til å øke i denne perioden.

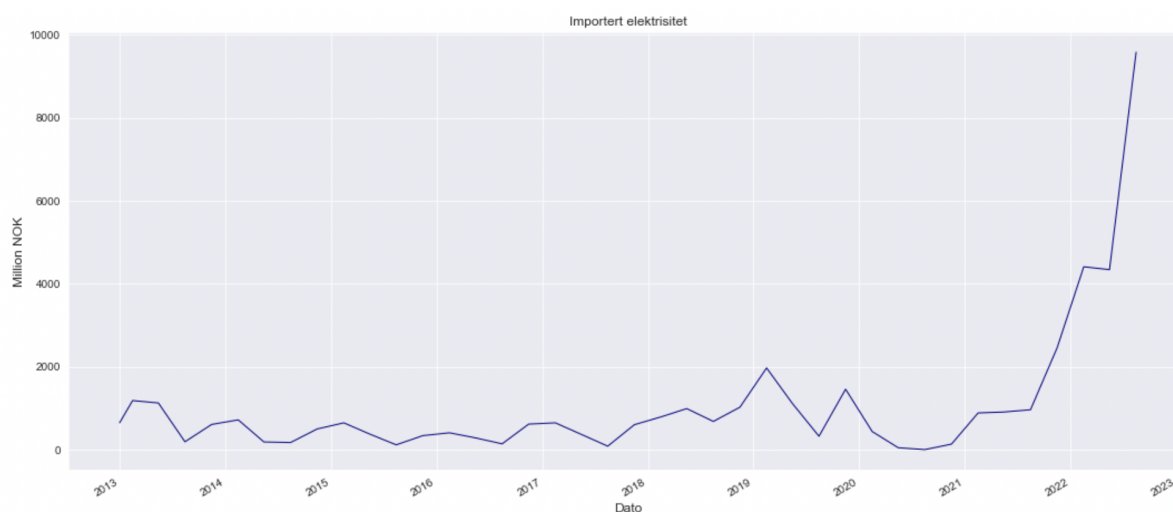
Kullpris - ICE gc Newc



Figur 4.8: Kullpris i perioden 2013 - 2023

Akkurat som for prisen på elektrisitet, gass og CO₂-kvoter, ser man en tydelig økning i prisen på kull i 2021, med store fluktuasjoner i prisen i 2022. Det kommer tydelig frem at prisen på de fire handelsvarene følger hverandre tett over tidsrammen.

Import elektrisitet



Figur 4.9: Importert elektrisitet i perioden 2013 - 2023

Det er av interesse at importen av elektrisitet øker kraftig når prisen på elektrisitet øker. Dette er i samhold med prinsippet om fri flyt av elektrisitet over landegrensene. Ved høye priser i et område vil importen øke fra andre områder med lavere priser. Man ser en stabil, relativt lav import av elektrisitet de samme årene spotprisen på elektrisitet holdt seg lav. Året 2020, da spotprisen var på sitt laveste, ser man også en dal i grafen for import. Når spotprisen stiger i 2021 begynner man også å se en stigning i importen. Året 2022 er preget av eksepsjonell høy import.

Resterende uavhengige variabler

Av grafene for de resterende variablene (se Appendix) ser man blant annet at eksporten av strøm var veldig høy i 2020, da strømprisen var lav. Eksporten avtok i 2022 da strømprisen nådde toppnivået. Nedbørsnivået har holdt seg relativt stabilt, men 2022 ser ut til å ha hatt noe mindre nedbør enn tidligere år. Det er ingen spesielle utslag i temperaturen fra år til år. Samtidig kan man observere i grafene for valutakurser at NOK har svekket seg i forhold til euro og dollar. Fyllingsgraden har også sett rimelig lik ut hvert år, men man ser et bunnivå i 2022. Det er en plausibel hypotese at siden prisen er så høy vil magasinene tappes mer. Til slutt, kan man se at energiforbruket ikke har økt de siste årene, men har hatt en relativt lik syklus som tidligere.

4.2 Resultater

4.2.1 Lineær regresjon

Modell	OLS
Ant. Observasjoner	3285
R-squared	0,812
Prob > F	0,00
MSE	6906,7
Root MSE	83,1
MAE	59,1

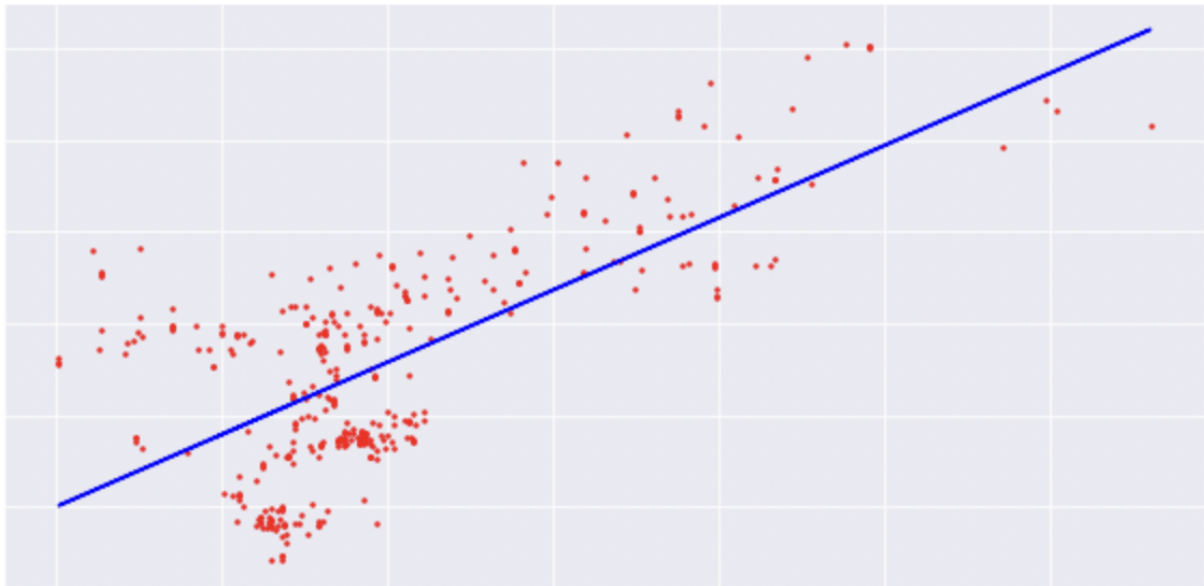
<i>Spotpris elektrisitet NO1</i>	<i>Koeffisient</i>	<i>Standardavvik</i>	<i>t-verdi</i>	<i>p-verdi</i>
Konstant	168,056	25,855	6,500	0,000
Nedbør	-0,165	0,035	-4,746	0,000
Middeltemperatur	-0,613	0,049	-12,478	0,000
Oljepris	0,023	0,031	-0,714	0,475
Gasspris	0,753	0,029	26,114	0,000
USD/EUR	-116,701	21,855	-5,340	0,000
NOK/EUR	11,285	3,024	3,732	0,000
NOK/USD	-16,407	3,384	-4,867	0,000
CO ₂ -kvoter	0,242	0,031	7,859	0,000
Kullpris	0,157	0,013	12,242	0,000
Norges Energiforbruk	-0,001	0,000	-4,023	0,000
Import	0,010	0,001	14,314	0,000
Eksport	0,010	0,002	3,729	0,000
Fyllingsgrad - NO1	7,845	2,742	2,861	0,004
Fyllingsgrad - Norge	-9,538	3,807	-2,458	0,014

Tabell 3: Resultater fra multippel lineær regresjonsanalyse

Modellen gir en R^2 -verdi på 0,812. Dette indikerer at de uavhengige variablene forklarer 81,2% av variansen i den avhengige variabelen, spotprisen.

- *Konstant*: Når alle de uavhengige variablene er lik null, vil den forventede strømprisen være 168,056 €/MWh.
- *Nedbør*: Hvis nedbørmengden øker med 1 millimeter, vil den forventede strømprisen reduseres med 0,165 €/MWh.
- *Middeltemperatur*: Hvis middeltemperaturen øker med 1°C, vil den forventede strømprisen reduseres med 0,613 €/MWh.
- *Oljepris*: Hvis oljeprisen øker med 1\$ per tønne, økes den forventede strømprisen med 0,023 €/MWh.
- *Gasspris*: Hvis gassprisen øker med 1€, vil den forventede strømprisen øke med 0,753 €/MWh.
- *USD/EUR*: Hvis kursen på euro målt mot amerikansk dollar økes med én enhet, vil den forventede strømprisen reduseres med 116,701 €/MWh.
- *NOK/EUR*: Hvis eurokursen målt mot den norske kronen økes med én enhet, vil den forventede strømprisen øke med 11,285 €/MWh.
- *NOK/USD*: Hvis kursen på amerikansk dollar økes med én enhet målt mot den norske kronen, vil den forventede strømprisen reduseres med 16,407 €/MWh.
- *CO₂-kvoter*: Hvis prisen på Co2-kvoter økes med 1€, vil den forventede strømprisen øke med 0,242 €/MWh.
- *Kullpris*: Hvis prisen på kull øker med 1\$, vil den forventede strømprisen øke med 0,157 €/MWh.
- *Norges Energiforbruk*: Hvis Norges energiforbruk øker med 1 GWh, vil den forventede strømprisen reduseres med 0,001 €/MWh.
- *Import*: Hvis Norges import øker med 1 MNOK, vil den forventede strømprisen øke med 0,01 €/MWh.
- *Eksport*: Hvis Norges eksport øker med 1 MNOK, vil den forventede strømprisen øke med 0,01 €/MWh.
- *Fyllingsgrad - NOI*: Hvis fyllingsgraden for elspotområde 1 (NO1) øker med 1 TWh, vil den forventede strømprisen øke med 7,845 €/MWh.
- *Fyllingsgrad - Norge*: Hvis fyllingsgraden for hele Norge øker med 1 TWh, vil den forventede strømprisen reduseres med 9,538 €/MWh.

Signifikansnivået α ble satt til 5%. Dette medfører at p-verdier $\leq 0,05$ vurderes som statistisk signifikante. I vår modell er alle variabler signifikante, unntatt én. Modellen viser til at variabelen *Oljepris* har en p-verdi på 0,475. Med andre ord er ikke denne variabelen gjeldende for modellen. Variabelen *Fyllingsgrad - Norge* følger etter med en p-verdi på 0,014. Denne er imidlertid å regne som signifikant da verdien ikke overstiger grensen på 0,05.

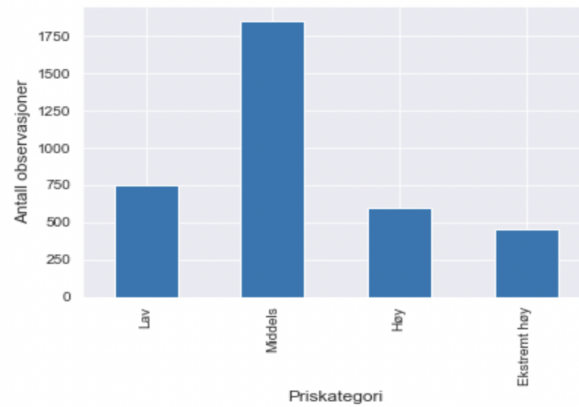


Figur 4.10: Visualisering av testresultat for multippel lineær regresjonsanalyse

Feilledene i modellen finner vi ved å observere minste kvadraters feil (MSE), og kvadratroten av minste kvadraters feil (Root MSE). Minste kvadraters feil viser til en verdi på 6906,7, mens kvadratroten av de kvadrerte feilene har en verdi på 83,1. Det kan også være verdt å nevne at absoluttverdien av de gjennomsnittlige feilene (MAE) er på 59,1.

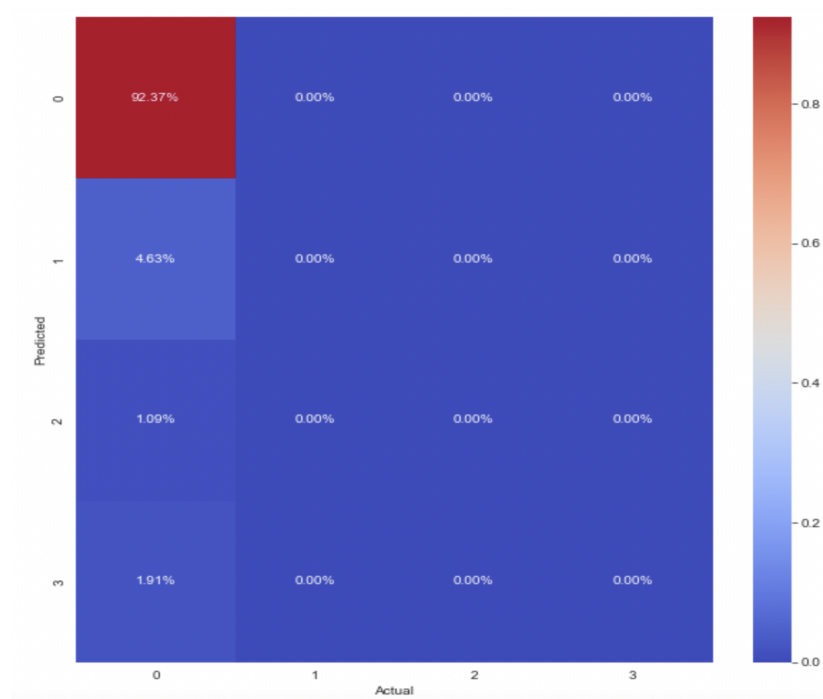
4.2.2 Naive Bayes Classification

Figur 4.11 viser fordelingen av observasjoner i de ulike prisklassene; lav, medium, høy og ekstrem høy spotpris.



Figur 4.11: Observasjoner i hver priskategori

Figuren viser til at antall observasjoner som er å anse som middels spotpris utgjør den største delen av datasettet, med 1856 observasjoner. I lav, høy og ekstrem høy prisklasse finner vi hhv. 750, 593 og 453 observasjoner. Hvor godt modellen klarer å plassere prediksjoner i riktig prisklasse vises i figuren under.



Figur 4.12: Fargekart av testsettet for Naive Bayes

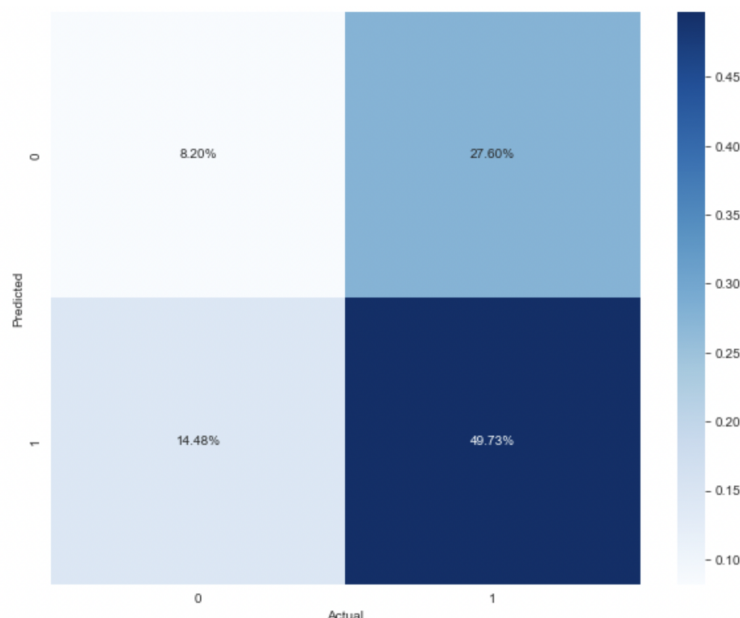
Fargekartet viser til faktiske verdier langs x-aksen, mens de predikerte verdiene utgjør y-aksen av matrisen. Aksene har verdiene 0, 1, 2 og 3, og representerer hhv. pris-kategoriene ekstremt høy, høy, lav, og middels. Testsettet har en treffsikkerhet på 92,37%. Av figuren kan man se at samtlige observasjoner i testsettet faller under kategorien ekstrem høy spotpris.

Dette kommer av at testsettet består av året 2022 som er preget av ekstraordinære høye priser. Dermed vil scoren på testsettet kun reflektere hvor nøyaktig modellen klarer å predikere verdier i kategorien ekstremt høy. Treningssettet inneholder et mer balansert forhold mellom kategoriene, utenom kategorien ekstrem høy da disse observasjonene stort sett havner i testsettet. Treningssettet har en treffsikkerhet på 72,39%.

4.2.2 Logistisk regresjon

Modellen gir en treffsikkerhet på treningssettet på 64,66% og 57,92% på testsettet.

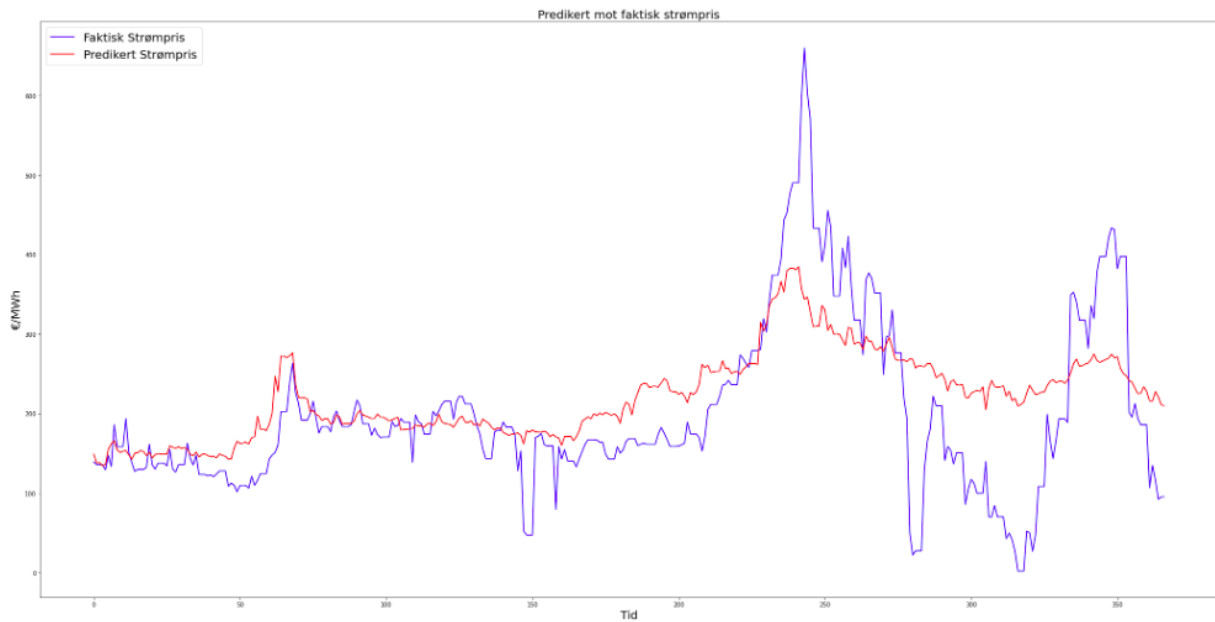
Resultatene viser at presisjonen på testsettet er 36% for 0-verdiene, det vil si når strømprisen går ned. For 1-verdier er presisjonen 64%. Det er i alt 131 0-verdier, og 235 1-verdier i testsettet. Dette visualiserer vi ved hjelp av et fargekart. X-aksen viser faktiske verdier, mens y-aksen består av de predikerte verdiene.



Figur 4.13: Fargekart av testsettet for logistisk regresjon

4.2.3 ANN

Figur 4.14 viser ANN-modellens evne til å predikere på testdataen (året 2022). Den røde linjen illustrerer modellens prediksjon, mens den blå linjen representerer den faktiske spotprisen. Man kan observere at modellen sliter med å følge de faktiske svingningene i løpet av perioden. Modellen fanger imidlertid opp nyansene av svingninger, hvor vi kan se at modellen følger den faktiske trenden. Grunnen til at modellen ikke klarer å fange opp de ekstreme svingningene kan være på grunn av store strukturelle endringer i løpet av 2022. Dette gjenspeiles i grafen ved at spotprisen har variert stort fra start til slutt.



Figur 4.14: Sammenligning mellom predikert og faktisk strømpris

Resultatet av modellen viser til en R^2 -score 0,745 på treningssettet, mens den gir 0,429 i R^2 -score på testsettet. Forskjellen mellom trenings- og testsett gjenspeiles også i feilverdiene i modellen. Minste kvadratiske feil (fra engelsk: *mean squared errors*) viser til 144,3 på treningssettet, og 6358,8 på testsettet. Med andre ord gir modellen et betydelig bedre resultat på treningssettet i form av høyere forklaringskraft og vesentlig mindre feilledd.

5. DISKUSJON

5.1 Generell diskusjon

Det er relativt store forskjeller i hvordan de ulike modellene presterer når det kommer til prediksjon. Det er viktig å gjennomføre en grundig refleksjon rundt modellenes resultater og ulike feilkilder som kan påvirke modellene. En høy treffsikkerhet på testsettet er ikke ensbetydende med en modell med god forklaringskraft. Det er derfor elementært å ta et dypere dykk inn i modellenes resultater.

ANN-nettverket evner til en viss grad å tilpasse seg datasettet vårt. Resultatet var i stor grad avhengig av valg av antall noder og lag i modellen. Etter omfattende prøving og feiling evnet vi å komme frem til en modell vi kan si oss tilstrekkelig fornøyde med. Dette betyr dog ikke at modellen er særlig godt egnet til prediksjon. Vi ser at modellen til dels fanger opp trenden i utviklingen, men avviker i stor grad ved ekstreme verdier, samt når volatiliteten øker.

Feilverdiene på testsettet utgjør en MSE-verdi på 6358.8, noe som er svært høyt. Samtidig gir modellen en R^2 -verdi på 0,429, noe som kan tyde på at variablene ikke egner seg så godt til å predikere strømprisen.

Ser vi derimot på den lineære regresjonsmodellen, kan vi ved første øyekast observere gode resultater. Modellen har svært høy forklaringskraft. Av den grunn kan man argumentere for at de utvalgte uavhengige variablene bidrar godt til å forklare endringene i strømprisen. Likevel har modellen fortsatt relativt store feilverdier. Videre foreligger det et sett av forutsetninger for regresjonsmodellen som må tas i betraktning. For det første er variabelen oljepris ikke signifikant, og således ikke gyldig i modellen. Videre kan man observere faresignaler når det kommer til multikollinearitet - graden av lineær sammenheng mellom forklaringsvariabler. Ut fra korrelasjonsmatrisen kan vi observere at blant annet gasspris har svært høy korrelasjon med både CO_2 kvotepris og kullpris hvor korrelasjonsverdier er på hhv. 0,8 og 0,9. Dette er med på å påvirke modellens validitet og kan føre til falske konklusjoner. De høye korrelasjonsverdiene kan muligens forklares ved at strømmarkedet er et kompleks system hvor flere faktorer er med på å påvirke hverandre. Kompleksiteten gjør at det blir vanskeligere å danne en lineær sammenheng, noe man kan observere i visualiseringen av regresjonslinjen (se figur 4.10).

Naive Bayes-modellen har den høyeste treffsikkerheten på testsettet av alle modellene. Likevel blir det for lett å konkludere med at dette er en modell som klarer å predikere strømprisen på en god måte. Modellen predikerer kun hvilken prisklasse verdien tilhører. Det er altså ingen presis prediksjon av strømprisen; kun en klassifisering. I tillegg inneholder testsettet kun verdier som tilhører priskategorien ekstremt høy. Dermed får man ikke et bilde av hvordan modellen klarer å predikere verdier som ligger innenfor andre priskategorier. Samtidig tar modellen utgangspunkt i at alle uavhengige variabler bidrar like mye og er uavhengige av hverandre, noe som nødvendigvis ikke er realiteten i dette tilfellet.

Den logistiske regresjonsmodellen predikerer om prisen vil stige eller synke sammenlignet med foregående dag. Slik som ved Naive Bayes-modellen er det her heller ikke en presis prediksjon av strømprisen, men hvilken retning prisen vil bevege seg. Modellen har store problemer med å klare å predikere korrekt når prisen synker, og er dermed ikke en særlig hensiktsmessig modell.

I vurderingen må det også tas hensyn til ikke-kvantifiserbare faktorer som er med på å direkte og/eller indirekte påvirke strømprisen. Den viktigste forklaringsvariabelen i vår studie er trolig krigen mellom Ukraina og Russland som i stor grad har påvirket energimarkedet. Dette er en viktig geopolitisk faktor som ikke direkte lar seg tallfestes. Ringvirkningene av krigen kan imidlertid ses i strømprisens utvikling det siste året hvor det har vært store fluktuasjoner i prisen. I tillegg til krigen er det også andre ikke-kvantifiserbare faktorer, som politiske vedtak, overføringskapasitet i strømmettet og tap i strømmettet, som spiller inn på strømprisen. Som følge av slike uforutsette faktorer, blir det vanskelig for modellene å gjenkjenne tydelige trender når det kommer til prediksjon.

5.2 Forbedring og evaluering av modellene

Et mer komplett datasett med færre manglende verdier ville potensielt kunne ført til modeller med større prediksjonskapabilitet. Vi har blant annet variabler som kun har data på kvartalsvis-, månedlig- og ukentlig basis, og siden strømprisen er noe som endres dag til dag ville et mer oppdatert datasett kunne ført til mer presise prediksjoner. Siden manglende verdier erstattes med siste observerte verdi, vil særlig variabler med kun kvartalsvis og månedlig data få mange like verdier, noe som kan påvirke modellen. Et mer oppdatert datasett med flere observasjoner ville vært ønskelig for å oppnå en bedre modell.

Året 2022 har også vært særlig problematisk for utarbeidelsen av modellene. For å kunne predikere de ekstreme prisene har vi måttet inkludere 2021 i treningssettet, som også var et ekstraordinært år sett i historisk sammenheng. Dette gir et testsett på bare 10% av det totale datasettet, som også i stor grad inneholder uvanlige priser. Et lite testsett med abnorme priser gjør prediksjon krevende. Man kan se at historiske priser kan komme til kort når man skal predikere priser i år preget av eksepsjonelle hendelser, slik som konflikten i Ukraina.

En annen faktor som kunne forbedret modellene ville vært å inkludere flere variabler i modellen. Siden det norske kraftmarkedet er tett knyttet opp mot det europeiske kunne variabler slik som strømproduksjon- og forbruk i ulike europeiske land gitt en modell med bedre forklaringskraft. Samtidig er det verdt å påpeke at relevante variabler i stor grad vil påvirke hverandre, slik som for eksempel gasspris og kullpris påvirker hverandre. Av den grunn kan det være vanskelig å generere gode modeller som tar utgangspunkt i at variablene er uavhengige av hverandre.

Det nevralt nettverket kunne vært forbedret ved å endre antall noder og/eller lag i modellen, samtidig som man eventuelt kunne tatt i bruk andre aktiveringsfunksjoner. Med tanke på at det konstruerte nettverket er nokså simpelt, kunne man også utformet andre typer nettverk som for eksempel et LSTM- nettverk (fra engelsk: *Long short-term memory*). Et slikt nettverk er bedre egnet til å behandle sekvenser av data, samtidig som det i større grad klarer å memorere data. Det vil uansett være viktig å utforme et nettverk som ikke er for godt tilpasset treningsdataen (overfit) eller for dårlig tilpasset treningsdataen (underfit). For et godt tilpasset nettverk som ikke bærer preg av over- eller underfit, vil trenings- og testsettet nærme seg hverandre.

6. KONKLUSJON

Hensikten med studien har vært å avdekke hvilke modeller som evner å predikere strømprisen på en god måte, samt hvilke faktorer som viser seg å være de viktigste forklaringsvariablene. Å kunne predikere strømprisen er viktig med tanke på risikostyring i energimarkedet, spesielt i uvanlige år som 2022.

Resultatene viser til en høy korrelasjon mellom strømprisen, prisen på gass, kullprisen, Norges import av elektrisitet, samt prisen på CO₂-kvoter. Ut ifra modellenes resultater kan vi vise til en god treffsikkerhet ved Naive Bayes klassifiseringsmodell. Da modellen ikke er i stand til å predikere den gitte strømprisen, er den imidlertid i stand til å kategorisere prisen i klasser. Samtidig består testsettet kun av verdier innenfor én kategori, noe som gjør at modellen ikke får testet på andre prisnivåer. Av den grunn blir ikke modellen særlig pålitelig. Videre kan man argumentere for at regresjonsmodellene ikke er like godt egnet til å forutsi strømprisene. Den lineære modellen viser til gode resultater i form av god forklaringskraft, men sliter imidlertid med å oppfylle forutsetningene for modellen. På lik linje evner ikke den logistiske regresjonsmodellen å predikere oppgang eller nedgang i strømprisen i særlig grad. Disse modellene kan vise seg å være for simple med tanke på å forutsi strømprisen i et komplekst strømmarked hvor flere uavhengige variabler i stor grad påvirkes av hverandre. Til slutt kan man argumentere for at det nevralt nettverket er å anse som den best egnede modellen - spesielt dersom nettverket utbedres eller trenes mer. Til tross for lite overbevisende resultater, har denne modellen bedre forutsetninger for å trenes opp til å håndtere strømprisens store svingninger. Dette vises til en viss grad i resultatene.

Når det kommer til hvilke faktorer som er med på å påvirke strømprisen, kan man vise til at strømmarkedet er et sammensatt system. De uavhengige variablene som er tatt i bruk i modellene er klart faktorer som er med på bidra til endringer i strømprisen, hvor korrelasjonsmatrisen kan vise til at gasspris, kullpris, import og CO₂ kvotepris i størst grad samvarierer med strømprisen. Likevel oppnår ikke modellene mer solide resultater, noe som kan komme av at flere variabler mangler observasjoner, til en viss grad er avhengige av hverandre, og at ikke-kvantifiserbare faktorer ikke er inkludert. Til tross for dette, bidrar de uavhengige variablene til å fange opp nyanser av fluktuasjonen i strømprisen som vises ved nokså god forklaringskraft. Summen av dette tyder på at variablene som er inkludert i

modellen absolutt er relevante, men at det er nærliggende å tro at både flere og mer komplekse faktorer også er med på å drive strømprisen.

6.1 Videre arbeid

Videre forskning innenfor feltet burde ta utgangspunkt i gode og solide modeller som egner å prosessere større datasett. Vår forskning har basert seg på 14 uavhengige variabler, og det kan være hensiktsmessig med både flere observasjoner og flere variabler. Samtidig har denne studien tatt utgangspunkt i data fra de siste ti årene, hvor de fleste variablene har vært på et stabilt lavere nivå enn de siste 2-3 årene. Det vil være av interesse for senere forskning å inkludere de siste årene med abnorme priser i treningssett for å så kjøre prediksjon på et testsett med mer normale priser sett i historisk sammenheng.

Det ville også vært interessant å undersøke om strømpriser er gjenstand for tilfeldig vandring (fra engelsk: *random walk*). Denne teorien forsøker å fastsette om endringer i priser på ulike aktiva er tilfeldige (Smith, 2023). Dersom dette er tilfelle, vil ikke historiske strømpriser kunne anvendes til å predikere fremtidige strømpriser på en god måte.

7. LITTERATURLISTE

- Askheim (2023). *Energiloven*. Tilgjengelig fra: <https://snl.no/energiloven>, Hentet 02.02.2023
- Baheti, P (2023). *Overfitting vs Underfitting in Machine Learning [Differences]*. Tilgjengelig fra: <https://www.v7labs.com/blog/overfitting-vs-underfitting>. Hentet 20.04.2023
- Baheti, P (2022). *Activation Functions in Neural Networks*. Tilgjengelig fra: <https://www.v7labs.com/blog/neural-networks-activation-functions>. Hentet 23.03.2023
- CME Group (2023). *Understanding Basics of the Power Market*. Tilgjengelig fra: <https://www.cmegroup.com/education/courses/introduction-to-energy/introduction-to-power/understanding-basics-of-the-power-market.html>. Hentet 01.02.2023
- Chng, Z. M. (2022). *Using Activation Functions in Neural Networks*. Tilgjengelig fra: <https://machinelearningmastery.com/using-activation-functions-in-neural-networks/>. Hentet 15.03.2023.
- Delebekk, N. (2023). *Norge tjener store penger på Putins krig*. Tilgjengelig fra: <https://www.faktisk.no/artikler/z41ov/norge-tjener-store-penger-pa-putins-krig?embed=1>. Hentet 28.03.2023
- Elementsofai (2023). *Neural network basics*. Tilgjengelig fra: <https://course.elementsofai.com/5/1>. Hentet 01.02.2023
- El-ordboken (2016). Tilgjengelig fra: <https://kraftnytt.no/el-ordboken/>. Hentet 16.02.2023
- Elster, A. C., Tidemann, A. (2022). *maskinl ring*, Tilgjengelig fra: snl.no/maskinl ring. Hentet 24.01.2023
- Energifakta Norge. (2022). *Kraftmarkedet*. Tilgjengelig fra: <https://energifaktanorge.no/norsk-energiforsyning/kraftmarkedet/>. Hentet 02.02.2023
- Energifakta Norge. (2022). *Kraftproduksjon - Energifakta Norge*. Tilgjengelig fra: <https://energifaktanorge.no/norsk-energiforsyning/kraftforsyningen/>. Hentet 01.02.2023
- Energifakta Norge. (2019). *Str mnett*. Tilgjengelig fra: <https://energifaktanorge.no/norsk-energiforsyning/kraftnett/>. Hentet 18.04.2023
- European Council. (2023). *Infographic - Where does the EU's gas come from?* Tilgjengelig fra: <https://www.consilium.europa.eu/en/infographics/eu-gas-supply/>. Hentet 27.03.2023

- E24 (2022). *Strømprisrekord i Sør-Norge i 2022*. Tilgjengelig fra: <https://e24.no/naeringsliv/i/3EqQdA/stroemprisrekord-i-soer-norge-i-2022>. Hentet 03.02.2023
- Finanssenteret (2023). *Hva Viser En Korrelasjonsmatrise?* Tilgjengelig fra: <https://www.finanssenteret.as/emne/1763/hva-viser-en-korrelasjonsmatrise>. Hentet 24.03.2023
- Funksjonen tanh(x)* (2023). [Digitalt bilde]. Tilgjengelig fra: <https://www.desmos.com/calculator/mv0ymvkjze>. Hentet 09.02.2023.
- Grøn, Ø., (2021). *elektrisk strøm*. Tilgjengelig fra: https://snl.no/elektrisk_str%C3%B8m. Hentet 01.02.2023
- Heldal (2006). Logistisk regresjon - kurskompendium i byråskolens kurs SM507. (n.d.). Tilgjengelig fra: https://www.ssb.no/a/publikasjoner/pdf/notat_200654/notat_200654.pdf. Hentet 26.02.2023
- Helgelandkraft (2022). *Hva påvirker spotprisen?* Tilgjengelig fra: <https://www.helgelandkraft.no/strom/privat/aktuelt/hva-pavirker-spotprisen/>. Hentet 01.02.2023
- Hofstad, K. (2020). *Norges vassdrags- og energidirektorat*. Tilgjengelig fra: https://snl.no/Norges_vassdrags-_og_energidirektorat. Hentet 01.02.2023
- IBM (2016). *What is Machine Learning?*. Tilgjengelig fra: <https://www.ibm.com/topics/machine-learning>. Hentet 24.01.2023
- IBM (2021). *What are Neural Networks?*. Tilgjengelig fra: <https://www.ibm.com/topics/neural-networks>. Hentet 27.01.2023
- IBM (2023). *What is Naïve Bayes*, Tilgjengelig fra: <https://www.ibm.com/topics/naive-bayes>. Hentet 24.03.2023
- Illustrasjon av kraftmarkedet*. (2022). [Digitalt bilde]. Tilgjengelig fra: https://energifaktanorge.no/?attachment_id=16796. Hentet 02.02.2023
- Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser? Innføring i samfunnsvitenskapelig metode*. 3. utgave. Cappelen Damm akademisk
- Lorvik, N. (2021). *Energi Norge avslører hvilke strømselskap som er med i sertifiseringsordningen "Trygg strømhandel"*. Tilgjengelig fra: <https://www.nettavisen.no/okonomi/sjekk-oversikten-stromselskapene-til-en-million-nordmenn-har-ikke-fatt-godkjent-stempel/s/5-95-284895>. Hentet 30.03.2023

- Loss functionality*. (2023). [Digitalt bilde]. Tilgjengelig fra:
<https://www.nordpoolgroup.com/en/trading/Day-ahead-trading/loss-functionality/>.
Hentet 23.03.2023
- Miljødirektoratet (2019). *EUs system for klimakvoter*. Tilgjengelig fra:
<https://www.miljodirektoratet.no/ansvarsomrader/klima/klimakvoter/eus-klimakvotesystem/>. Hentet 27.03.2023
- Neural Network Architecture*. (2019). [Digitalt bilde]. Tilgjengelig fra:
<https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a>. Hentet 01.02.2023
- Nilsen, A. (2022). *Kraftig fall i gassprisen:- Kan snu veldig fort*. Tilgjengelig fra:
<https://e24.no/energi-og-klima/i/jl3oX9/kraftig-fall-i-gassprisen-kan-snu-veldig-fort>.
Hentet 28.03.2023
- Nordpoolgroup (2023). *Loss functionality* (2023). Tilgjengelig fra:
<https://www.nordpoolgroup.com/en/trading/Day-ahead-trading/loss-functionality/>.
Hentet 23.03.2023
- NTE (2022). *Hva påvirker strømprisen?*. Tilgjengelig fra:
<https://nte.no/blogg/hva-pavirker-stromprisen/>. Hentet 01.02.2023
- NTNU (2016). *Korrelasjon*. Tilgjengelig fra:
<https://tma4245.math.ntnu.no/forventing-og-varians/korrelasjon/>. Hentet 24.03.2023
- NTNU (2019). *Minste kvadraters metode*. Tilgjengelig fra:
<https://tma4245.math.ntnu.no/enkel-line%C3%A6r-regresjon/minste-kvadraters-metode/>. Hentet 29.03.2023
- NVE (2021). *1991: Den nye energiloven - fra forvaltning til forretning*. Tilgjengelig fra:
<https://www.nve.no/om-nve/vassdrags-og-energihistorie/nves-historie/1991-den-nye-energiloven-fra-forvaltning-til-forretning/>. Hentet 01.02.2023
- NVE (2018). *Dette bør du vite om strømprisen og kraftsituasjon - NVE*. Tilgjengelig fra:
<https://www.nve.no/reguleringsmyndigheten/kunde/stroem/stromkunde/dette-bor-du-vite-om-stromprisen-og-kraftsituasjon/>. Hentet 01.02.2023
- NVE (2021). *Strømprisen så langt i 2021*. Tilgjengelig fra:
<https://www.nve.no/om-nve/spoer-nve/om-stroempriser/stromprisen-sa-langt-i-2021/>.
Hentet 01.02.2023
- Olje- og energidepartementet (2023). *Regjeringens strømtiltak*. Tilgjengelig fra:
<https://www.regjeringen.no/no/tema/energi/regjeringens-stromtiltak/id2900232/>,
Hentet 03.02.2023

Olje- og energidepartementet. (2021). *Kraftmarkedet og strømpris*. Tilgjengelig fra: <https://www.regjeringen.no/no/tema/energi/stromnett/kraftmarkedet-og-strompris/id2076000/>. Hentet 01.02.2023

Price calculation (2023). [Digitalt bilde]. Tilgjengelig fra: <https://www.nordpoolgroup.com/en/trading/Day-ahead-trading/Price-calculation/>. Hentet 03.02.2023

Ramsøy, C. (2021). *Predictive analytics: Dette bør du tenke på før du setter i gang*. Visma Blogg - om teknologi, regnskap, skatt, lønn, innkjøp, HR. Tilgjengelig fra: <https://www.visma.no/blogg/predictive-analytics-tenke-pa-for-du-starter/>. Hentet 01.02.2023

Rosvold, K. (2023). *Kraftselskap*. Tilgjengelig fra: <https://snl.no/kraftselskap>. Hentet 02.02.2023

Smith, T. (2023). *Random Walk Theory: Definition, How It's Used, and Example*. Tilgjengelig fra: <https://www.investopedia.com/terms/r/randomwalktheory.asp>. Hentet 24.03.2023

Statkraft (2020). *Historien vår*. Tilgjengelig fra: <https://www.statkraft.no/om-statkraft/historien-var/#1992-2000/1992>. Hentet 01.02.2023

Statnett (2022). *Derfor har vi prisområder*, Tilgjengelig fra: <https://www.statnett.no/om-statnett/bli-bedre-kjent-med-statnett/om-strompriser/fakta-om-prisomrader/>. Hentet 01.02.2023

Team, K. (2023). *Keras documentation*, Tilgjengelig fra: <https://keras.io/about/>. Hentet 29.03.2023

Wikipedia (2023). *Hyperbolic functions*, Tilgjengelig fra: https://en.wikipedia.org/wiki/Hyperbolic_functions. Hentet 27.03.2023

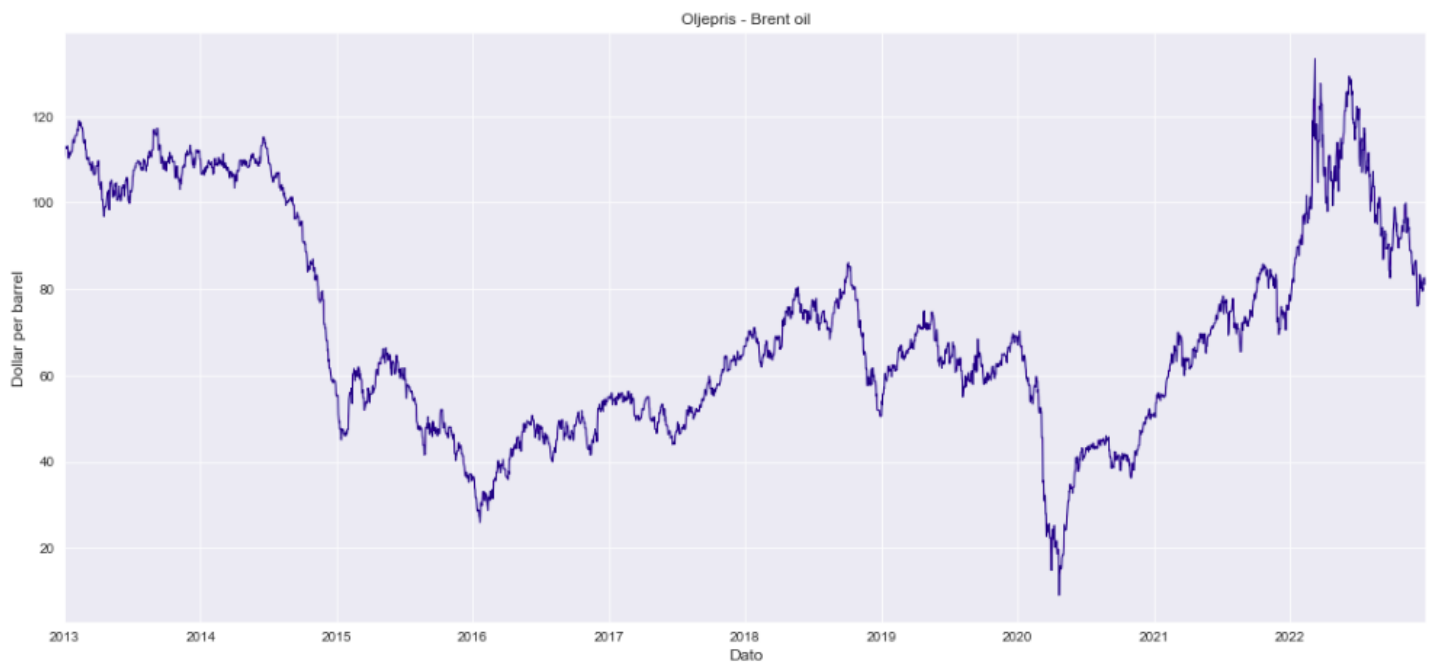
8. APPENDIX

8.1 Korrelasjonsmatrise

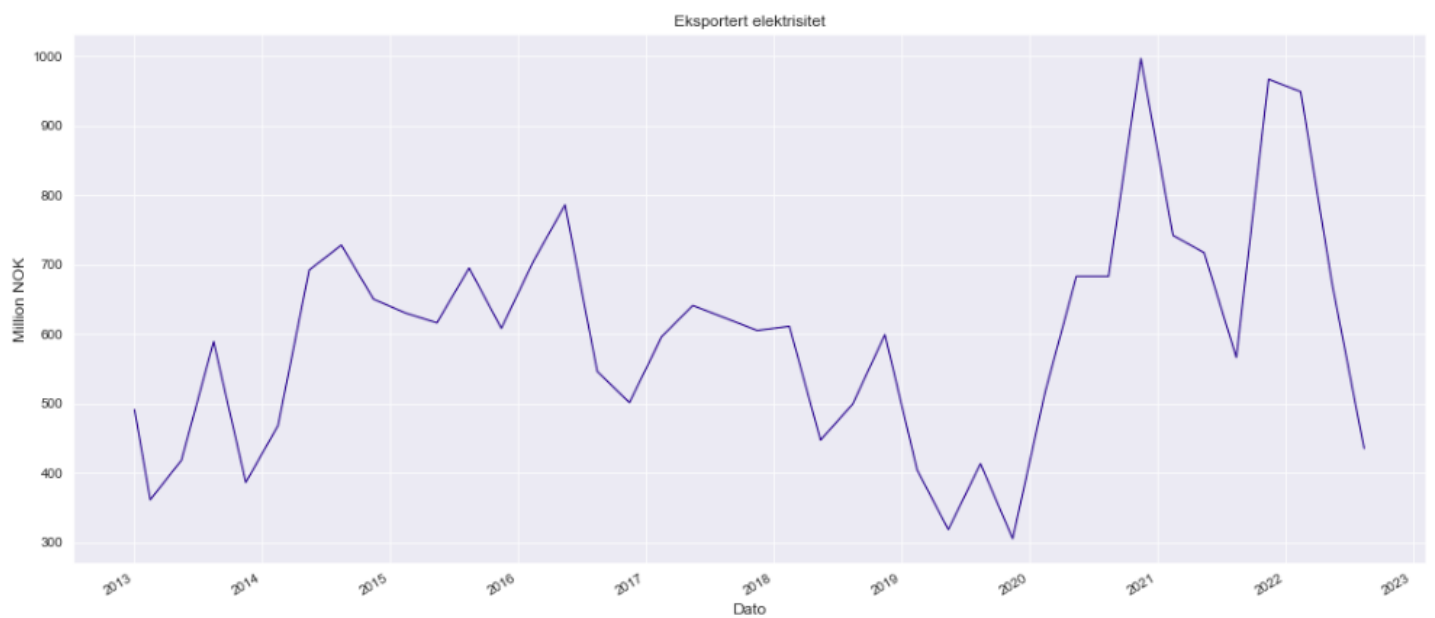
	Spotpris	Nedbør	Temperatur	Oljepris	Gasspris	USD/EUR	NOK/EUR	NOK/USD	CO2	Kullpris	Energiforbruk	Import	Eksport	Fyllingsgrad - NO1	Fyllingsgrad - Norge
Spotpris	1,0														
Nedbør	-0,1	1,0													
Temperatur	0,0	0,1	1,0												
Oljepris	0,4	-0,0	0,1	1,0											
Gasspris	0,9	-0,0	0,0	0,4	1,0										
USD/EUR	-0,3	0,0	-0,0	0,5	-0,3	1,0									
NOK/EUR	0,2	0,0	0,0	-0,5	0,2	-0,6	1,0								
NOK/USD	0,3	0,0	0,0	-0,5	0,3	-0,9	0,9	1,0							
Co2	0,7	-0,0	0,0	0,3	0,8	-0,4	0,6	0,6	1,0						
Kullpris	0,9	-0,0	0,0	0,5	0,9	-0,4	0,3	0,4	0,8	1,0					
Energiforbruk	-0,1	-0,1	-0,9	-0,1	-0,1	0,0	-0,1	-0,1	-0,1	-0,1	1,0				
Import	0,8	-0,0	-0,0	0,4	0,8	-0,4	0,2	0,4	0,7	0,9	-0,1	1,0			
Eksport	0,1	-0,0	-0,0	-0,1	0,1	-0,1	0,2	0,2	0,3	0,1	0,1	-0,1	1,0		
Fyllingsgrad - NO1	0,1	0,1	0,4	-0,1	-0,0	0,0	0,1	0,0	-0,0	-0,1	-0,5	-0,1	-0,0	1,0	
Fyllingsgrad - Norge	-0,1	0,1	-0,2	-0,1	-0,1	0,0	0,1	0,1	-0,1	-0,1	-0,3	-0,1	0,0	0,9	1,0

Figur 8.1: Korrelasjonsmatrise

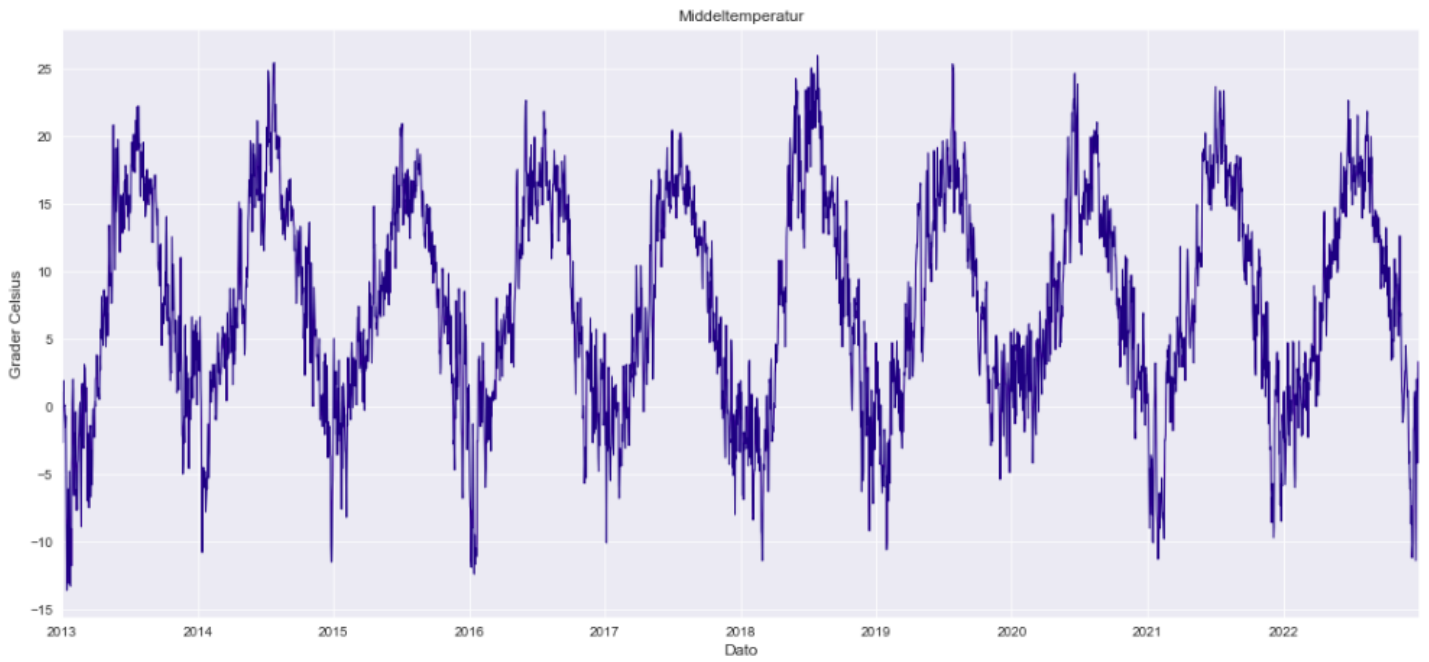
8.2 Resterende uavhengige variabler



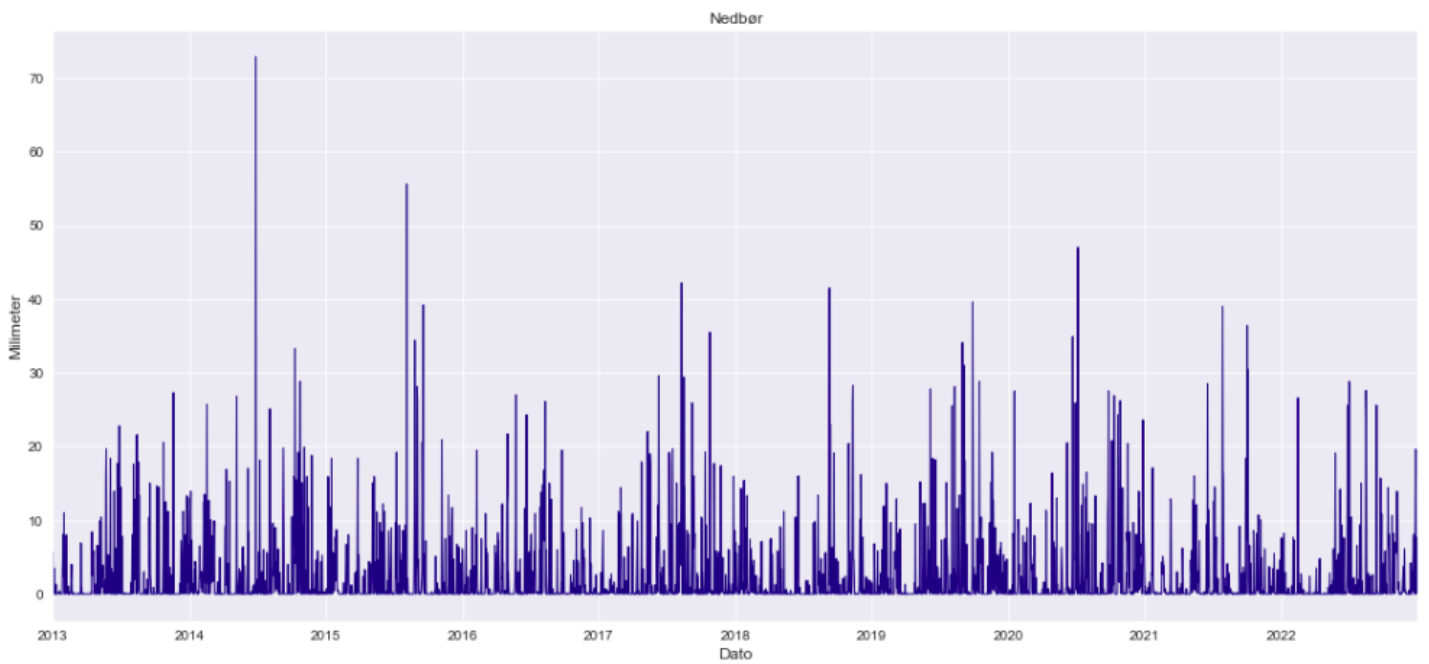
Figur 8.2: Oljepris 2013-2023



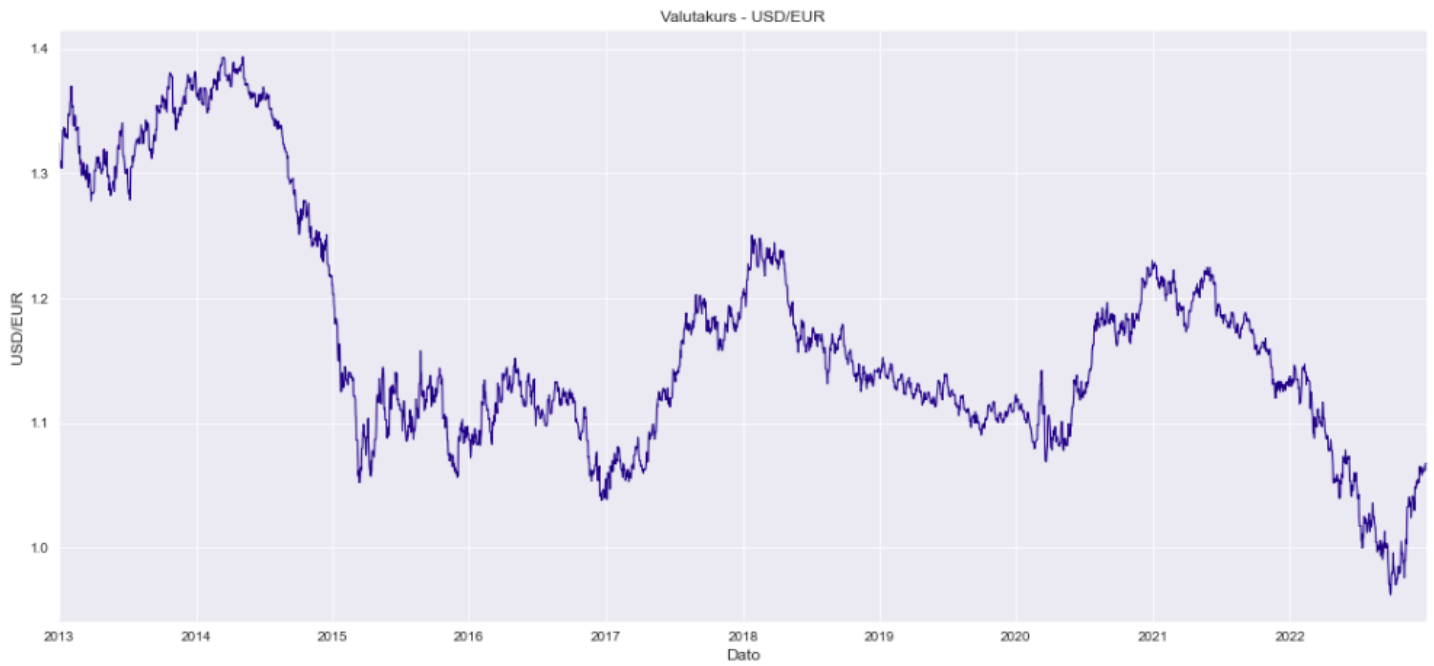
Figur 8.3: Eksport 2013-2023



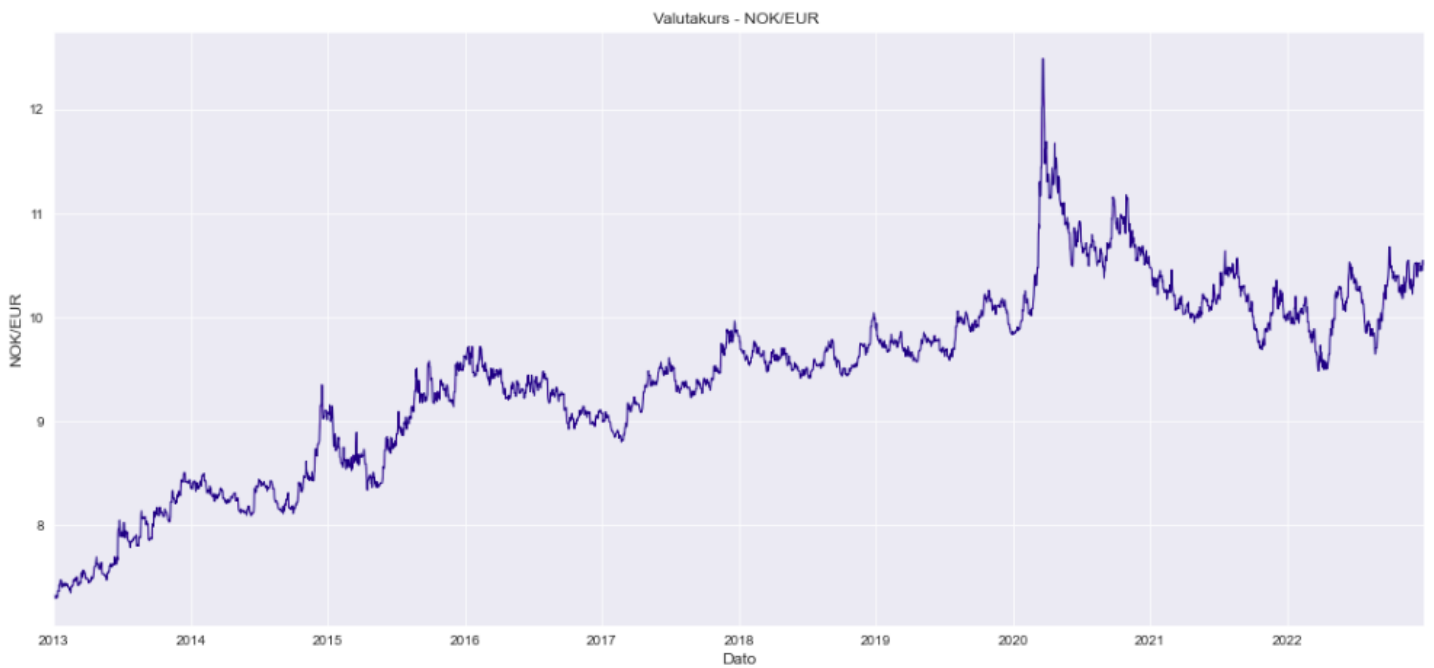
Figur 8.4: Middeltemperatur 2013-2023



Figur 8.5: Nedbør 2013-2023



Figur 8.6: USD til EUR 2013-2023



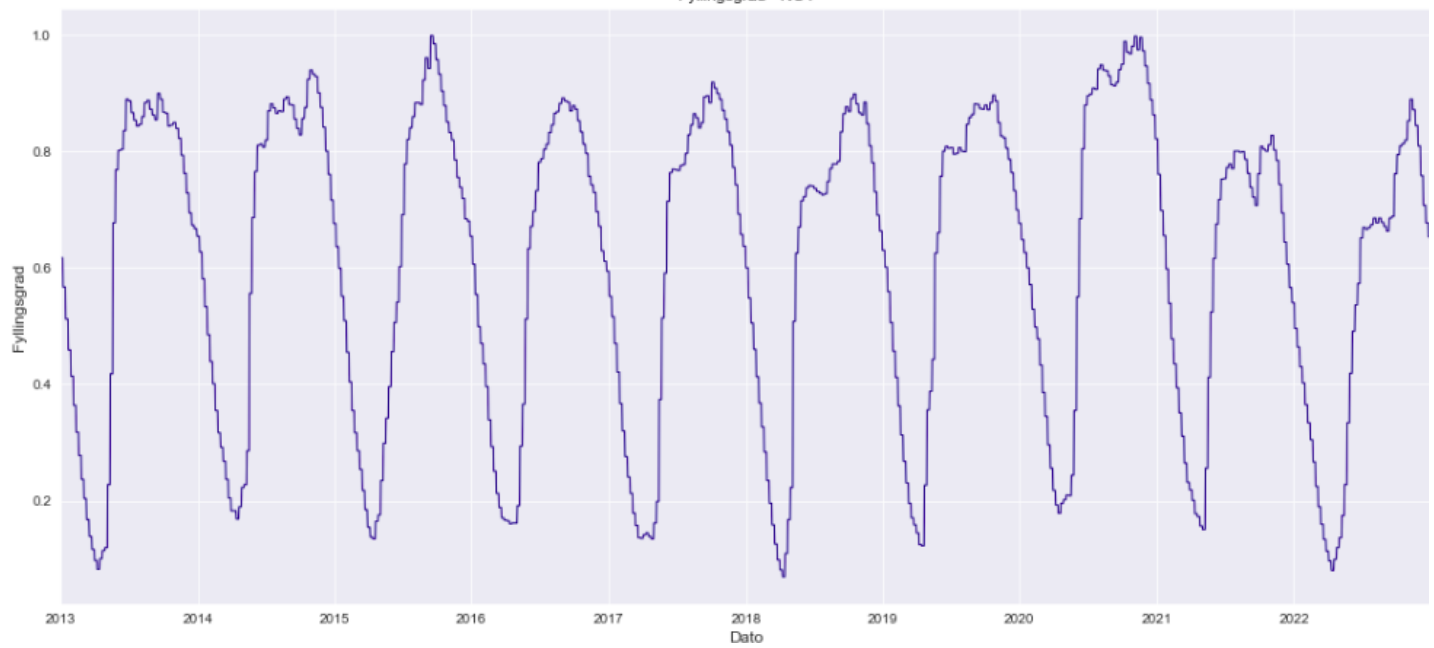
Figur 8.7: NOK til EUR 2013-2023

Valutakurs - NOK/USD



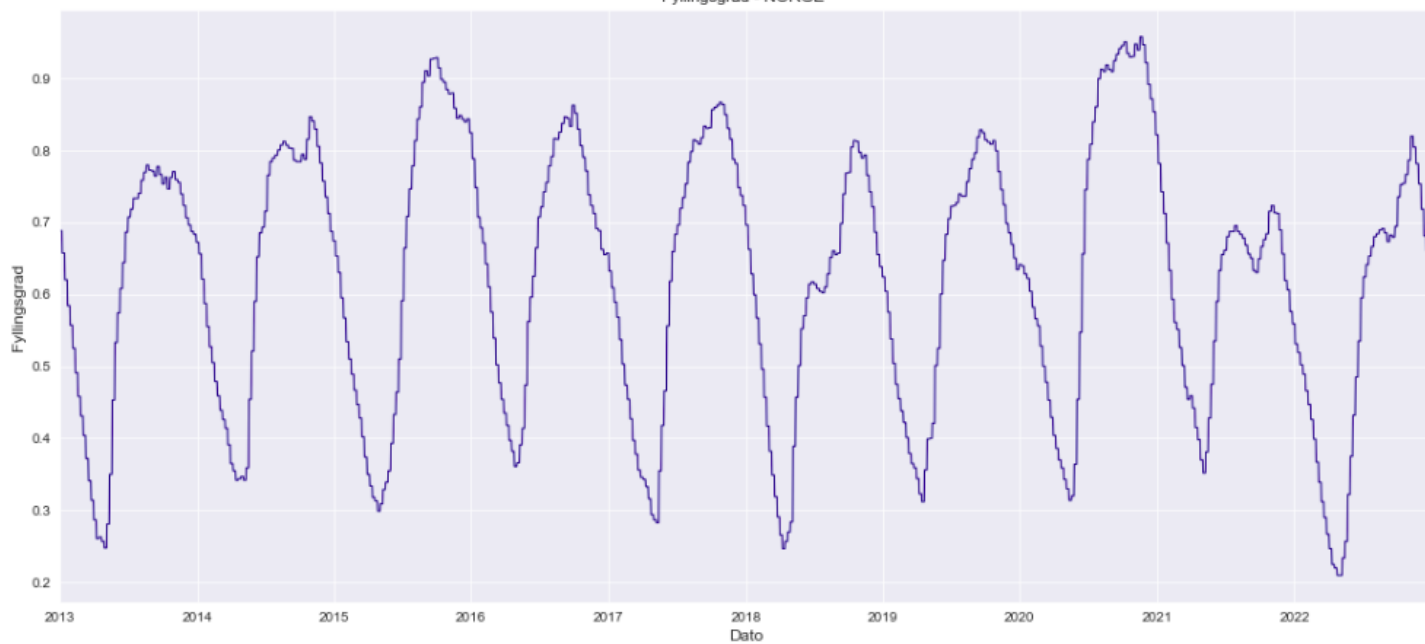
Figur 8.8: NOK til USD 2013-2023

Fyllingsgrad - NO1



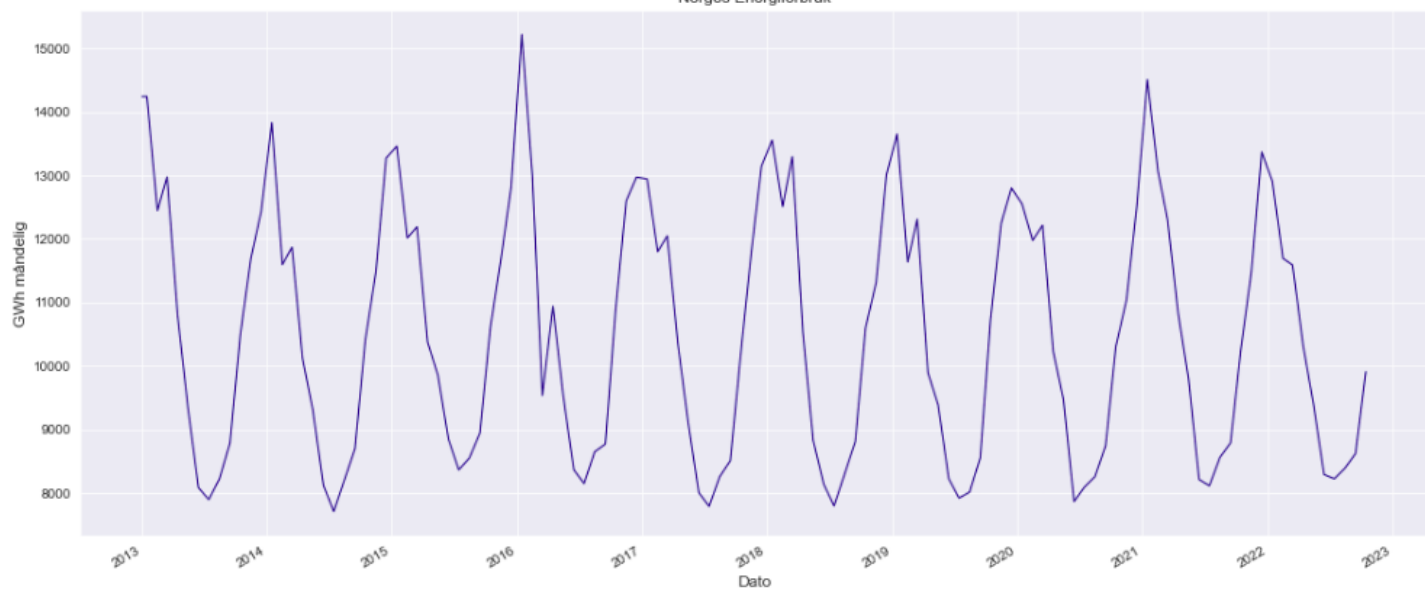
Figur 8.9: Fyllingsgrad NO1 2013-2023

Fyllingsgrad - NORGE



Figur 8.10: Fyllingsgrad - NORGE 2013-2023

Norges Energiforbruk



Figur 8.11: Norges energiforbruk 2013-2023

8.3 Zip-fil

Vedlagt ligger en zip-fil som inneholder tilhørende data og kode. Filen er kalt “Bachelor vedlegg”, og inneholder:

- “Bachelor med LR_NB_LG.ipynb”. Denne filen består av alle statistiske data når det gjelder datasettet, samt visualiseringer. Filen inneholder også maskinlæringsmodellene for Lineær regresjon, Naive Bayes Classification og Logistisk regresjon.
- “Bachelor Artificial neural network.ipynb”. Denne filen inneholder koden for det nevrale nettverket.
- “DailyEikon.csv”. Inneholder variabler fra Eikons database som oppdateres hver ukedag.
- “MonthlyEikon.csv”. Filen inneholder variabelen *Norges energiforbruk*
- “QuarterlyEikon.csv”. Fil som inneholder variablene *Import elektrisitet* og *eksport elektrisitet*.
- “DataSeKlima.csv”. Denne filen inneholder data hentet fra Norsk Klimaservicesenter og består av variablene *Middeltemperatur* og *Nedbør*
- “Fyllingsgrad.csv”. Fil bestående av variablene knyttet til fyllingsgrad hentet fra Noregs Vassdrags- og energidirektorat
- “Gjennomsnittlig spotpris og standardavvik.xlsx”. Excel-fil som inneholder figur 4.2

