



Predikerte boligpriser ved introduksjonen av Bybanen til Åsane

Hvordan vil Bybanen påvirke boligprisene langs traseen?

Kristoffer Døskeland og Lucas Francois

Veileder: Håkon Otneim

Masterutredning i økonomi og administrasjon

Hovedprofil: Finansiell Økonomi og Business Analytics

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

Forord

Denne masterutredningen er skrevet som avsluttende del av vårt masterstudium i økonomi og administrasjon ved Norges Handelshøyskole (NHH). Utredning utgjør 30 studiepoeng innenfor hovedprofilene Finansiell Økonomi og Business Analytics.

Formålet med utredningen er å bruke maskinlæring til å predikere hvordan utviklingen i boligprisene for perioden 2010-2019 mellom Byparken-Åsane ville vært dersom Bybanen til Åsane hadde vært ferdigstilt i 2010. Valg av tema er gjort med bakgrunn i interesse for boligpriser, i tillegg til et ønske om utvidet kunnskap tilknyttet maskinlæring og prediksjonsmodeller.

Vi vil benytte anledningen til å rette en stor takk til vår veileder Håkon Otneim for givende veiledning og konstruktive innspill gjennom hele skriveprosessen. Vi vil også takke Anders Lund fra Eiendomsverdi AS for utlevering av sentralt datagrunnlag.

Norges Handelshøyskole

Bergen, desember 2020

Sammendrag

Formålet med denne masterutredningen er å predikere hvordan boligprisutviklingen mellom Byparken og Åsane ville vært for perioden 2010-2019, i et tenkt scenario hvor Bybanen til Åsane hadde blitt ferdigstilt i 2010. Det finnes mange studier som redegjør for infrastrukturens påvirkning på boligprisnivå. Disse studiene har et økonometrisk perspektiv hvor formålet er å finne en presis fortolkning av hvordan avstand til offentlig transport påvirker boligprisene. Ulikt tidligere studier, søker vår utredning derimot å bygge en prediksjonsmodell som predikerer hvordan boligprisene vil utvikle ved introduksjonen av et offentlig transportmiddel som Bybanen.

I mangel på observasjoner mellom Byparken og Åsane som er påvirket av nærhet til Bybanen, overførte vi Bybanens effekt på boligprisene til Åsane-området ved å trene prediksjonsmodellene på et utvalg av boligtransaksjoner mellom Byparken og Flesland, hvor bybanens påvirkning er tilstede. Ettersom det ikke foreligger et testsett som er påvirket av Bybanen mellom Byparken og Åsane, kan vi ikke fastslå hvilken prediksjonsmodell som faktisk vil være den mest presise. Det foreligger imidlertid indikasjoner på at maskinlæringsmodellen Super Learner, uten å transformere noen variabler, vil predikere boligprisene i Åsane mer presist enn den lineære regresjonsmetoden Elastic Net. Videre finner vi at det vil være mest fornuftig å kun benytte observasjoner fra Byparken til Lagunen som treningssett for å best fange opp bybaneeffekten når vi predikerer i Åsane. Resultatene antyder også at prediksjonenes presisjon blir betydelig bedre når en variabel som hensyntar avstand til Bybanen inkluderes som en del av variabelgrunnet, alt annet like. Til slutt finner vi indikasjoner på at det er fornuftig å trene- og validere en prediksjonsmodell på observasjoner fra et område hvor Bybanen foreligger til å predikere boligpriser i et annet område hvor vi forestiller oss at Bybanen eksisterer

Når vi predikerer utviklingen i boligprisene mellom Byparken og Åsane for perioden 2010-2019, gitt Bybanens inntreden i 2010, finner vi at Bybanen predikeres å påvirke boligprisveksten i nærheten av Bybanen betydelig. Det predikeres videre å foreligge en økt vekst jo nærmere en bolig ligger et bybanestopp.

Nøkkelord – Maskinlæring, Prediksjon, Boligpriser

Innhold

| | | |
|----------|---|-----------|
| 1 | Innledning | 1 |
| 2 | Bakgrunn | 5 |
| 2.1 | Empiriske funn | 9 |
| 3 | Datasett | 11 |
| 3.1 | Den hedonistiske prisingsmodellen | 11 |
| 3.2 | Inkluderte variabler | 12 |
| 3.2.1 | Salgspris | 12 |
| 3.2.2 | Inflasjonsjustering og sesongjustering | 13 |
| 3.3 | Variabler fra den hedonistiske prisingsmodellen | 14 |
| 3.3.1 | Lokasjonsattributter | 14 |
| 3.3.1.1 | Innsamling av gåavstand fra hver bolig til nærmeste bybanestopp | 15 |
| 3.3.1.2 | Innsamling av kjøreavstand Bergen sentrum | 16 |
| 3.3.1.3 | Innsamling av gåavstand til nærmeste bykjerne | 16 |
| 3.3.2 | Miljømessige attributter | 17 |
| 3.3.3 | Strukturelle attributter | 18 |
| 3.3.4 | Makroøkonomiske variabler | 20 |
| 3.3.4.1 | Renter | 20 |
| 3.3.4.2 | BNP | 20 |
| 3.4 | Rensing og justering av data | 21 |
| 3.4.1 | Variabeltransformasjon | 21 |
| 3.4.2 | Kritikk mot manglende data | 22 |
| 3.5 | Overfitting og underfitting | 22 |
| 3.6 | Treningssett, valideringssett og testsett | 24 |
| 4 | Metode | 26 |
| 4.1 | Maskinlæring og prediksjon | 26 |
| 4.1.1 | Læringsprosessen - <i>supervised</i> og <i>unsupervised learning</i> | 26 |
| 4.2 | Kryssvalidering | 27 |
| 4.3 | Elastic Net | 28 |
| 4.3.1 | Ridge Regresjon | 29 |
| 4.3.2 | Lasso Regresjon | 30 |
| 4.3.3 | Elastic Net | 31 |
| 4.4 | Super Learner | 32 |
| 4.5 | RMSE og MAE - vurdering av presisjon | 33 |
| 5 | Analyse | 35 |
| 5.1 | Bybanens byggetrinn | 35 |
| 5.2 | Oppbygging av et helhetlig prediksjonsrammeverk | 36 |
| 5.2.1 | Variabeltransformasjon | 37 |
| 5.2.2 | Valg av trenings- og valideringssett | 40 |
| 5.2.3 | Bybaneeffektens overførbarhet | 48 |
| 5.2.4 | Betydningen av å inkludere variabelen <i>avstand til nærmeste bybanestopp</i> i våre prediksjoner | 50 |
| 5.3 | Prediksjoner av boligpriser tilknyttet Åsane-traseen | 52 |

| | | |
|----------|---|-----------|
| 5.4 | Prediksjoner av boligpriser tilknyttet Åsane-trasseen fra og med Bybanens ferdigstillelse og ni år frem i tid | 54 |
| 5.4.1 | Alle boliger innen 3000 meter fra nærmeste bybanestopp | 54 |
| 5.4.2 | Alle boliger innen 1000 meter fra nærmeste bybanestopp | 61 |
| 5.4.3 | Leiligheter og eneboliger innen 1000 meter fra nærmeste bybanestopp | 63 |
| 5.4.4 | Alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp | 67 |
| 5.4.5 | Kvantifisering av bybaneeffekten | 69 |
| 6 | Diskusjon | 72 |
| 6.1 | Prediksjonsrammeverkets relevans | 72 |
| 6.2 | Prediksjonenes informasjonsverdi | 74 |
| 6.3 | Politisk perspektiv og byutvikling rundt Åsane-traseen | 75 |
| 6.3.1 | Usikkerhet rundt Bybanen og boligutbygging i Åsane-området . . | 75 |
| 6.3.2 | Krav for utbygging rundt Bybanen | 76 |
| 6.4 | Prediksjonsmodellens svakheter | 77 |
| 6.5 | Forslag til videre forskning | 79 |
| 7 | Konklusjon | 81 |
| | Referanser | 84 |
| | Appendiks | 89 |
| A1 | Bakgrunnsinformasjon | 89 |
| A2 | Variabler | 91 |
| A3 | Prediksjoner av boligpriser rundt Åsane-trasseen | 96 |
| A3.1 | Årlige prediksjoner for alle boliger innen 3000 meter fra nærmeste bybanestopp | 96 |
| A3.2 | Årlige prediksjoner for alle boliger innen 1000 meter fra nærmeste bybanestopp | 98 |
| A3.3 | Årlige prediksjoner for alle leiligheter innen 1000 meter fra nærmeste bybanestopp | 99 |
| A3.4 | Årlige prediksjoner for alle eneboliger innen 1000 meter fra nærmeste bybanestopp | 101 |
| A3.5 | Årlige prediksjoner for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp | 102 |

Figurliste

| | | |
|------|--|----|
| 2.1 | Antall reisende med Bybanen fra første byggetrinn (Byparken-Nesttun) ble ferdigstilt til og med 2019. | 5 |
| 2.2 | Dagens Bybanetrasé med fremtidig utvidelse mot Fyllingsdalen og Vågsbotn, Åsane (Miljøloftet, 2020). | 6 |
| 3.1 | Korrelasjonsmatrise | 15 |
| 3.2 | Utvikling i kvadratmeterpris for inkluderte boligtyper mellom år 2000 og 2019 | 18 |
| 3.3 | Fordeling av de inkluderte boligtypene | 19 |
| 3.4 | Fordeling av boligtyper for ulike avstander fra nærmeste bybanestopp langs Byparken-Lagunen | 19 |
| 3.5 | Overfitting og underfitting (Fortmann-Roe, 2012) | 24 |
| 5.1 | Prediksjonsrammeverkets fire steg | 37 |
| 5.2 | Visualisering av årlig RMSE for de ulike modellene | 39 |
| 5.3 | Visualisering av RMSE for prediksjoner av Byparken-Åsane - Elastic Net | 43 |
| 5.4 | Visualisering av RMSE for prediksjoner av Byparken-Åsane - Super Learner | 44 |
| 5.5 | Gjennomsnittet av predikerte boligpriser mellom Byparken og Åsane i perioden 2001 til 2007 - Elastic Net. | 45 |
| 5.6 | Gjennomsnittet av predikerte boligpriser mellom Byparken og Åsane i perioden 2001 til 2007 - Super Learner. | 46 |
| 5.7 | Gjennomsnittet av predikerte boligpriser for alle boliger innen 3000 meter fra nærmeste Bybanestopp - Elastic Net | 54 |
| 5.8 | Gjennomsnittet av predikerte boligpriser for alle boliger innen 3000 meter fra nærmeste Bybanestopp - Super Learner | 59 |
| 5.9 | Gjennomsnittet av predikerte boligpriser for alle boliger innen 1000 meter fra nærmeste Bybanestopp - Elastic Net | 61 |
| 5.10 | Gjennomsnittet av predikerte boligpriser for alle boliger innen 1000 meter fra nærmeste Bybanestopp - Super Learner | 62 |
| 5.11 | Gjennomsnittet av predikerte boligpriser for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Elastic Net | 64 |
| 5.12 | Gjennomsnittet av predikerte boligpriser for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Super Learner | 64 |
| 5.13 | Gjennomsnittet av predikerte boligpriser for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net | 65 |
| 5.14 | Gjennomsnittet av predikerte boligpriser for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Super Learner | 66 |
| 5.15 | Gjennomsnittet av predikerte boligpriser for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net | 67 |
| 5.16 | Gjennomsnittet av predikerte boligpriser for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner | 68 |
| 5.17 | Utvikling av variable importance for alle dummy-avstander til Bybanen modellert med Elastic Net | 70 |
| A1.1 | Fremstilling av omtrentlig gåavstand fra Byparken. | 89 |
| A2.1 | Vekst av antall boliger per boligtype (2006-2019) | 92 |
| A2.2 | Fordeling - kvadratmeterpris med og uten transformasjoner | 92 |
| A2.3 | Fordeling - bygningsalder med og uten transformasjoner | 93 |
| A2.4 | Scatter plot - bygningsalder med og uten transformasjoner | 93 |

| | |
|---|----|
| A2.5 Fordeling - salgstid med og uten transformasjoner | 94 |
| A2.6 Scatter plot - salgstid med og uten transformasjoner | 94 |
| A2.7 Fordeling - tomteareal med og uten transformasjoner | 95 |
| A2.8 Scatter plot - tomteareal med og uten transformasjoner | 95 |

Tabelliste

| | | |
|-------|--|----|
| 4.1 | Metodene som velges manuelt til å inkluderes i Super Learner | 32 |
| 5.1 | Oversikt over Bybanens byggetrinn | 35 |
| 5.2 | Gjennomsnittlig RMSE og MAE for de ulike metodene | 38 |
| 5.3 | RMSE og MAE for predikerte boligpriser mellom Byparken og Åsane for perioden 2001-2007 | 42 |
| 5.4 | RMSE for predikerte boligpriser mellom Nesttun og Lagunen etter Bybanens ferdigstillelse. | 49 |
| 5.5 | RMSE for predikerte boligpriser mellom Nesttun og Lagunen før Bybanens ferdigstillelse. | 49 |
| 5.6 | RMSE og MAE for Elastic Net med og uten variabelen <i>avstand til nærmeste bybanestopp</i> | 51 |
| 5.7 | RMSE og MAE for Super Learner med og uten variabelen <i>avstand til nærmeste bybanestopp</i> | 51 |
| 5.8 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Elastic Net, historiske priser og treningssett | 57 |
| 5.9 | Koeffisienter for de inkluderte metodene i Super Learner | 60 |
| 5.10 | Oversikt over gjennomsnittlig boligprisvekst for vårt utvalg av prediksjoner - Elastic Net | 69 |
| 5.11 | Oversikt over gjennomsnittlig boligprisvekst for vårt utvalg av prediksjoner - Super Learner | 69 |
| A1.1 | Bybanenstraseenes egenskaper (Skyss, 2020a) | 90 |
| A2.1 | Variabelliste | 91 |
| A3.1 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett | 96 |
| A3.2 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett | 96 |
| A3.3 | Koeffisientene for prediksjonene av alle boliger innen 3000 meter fra nærmeste bybanestopp - Super Learner | 97 |
| A3.4 | Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 3000 meter nærmeste bybanestopp - Elastic Net | 97 |
| A3.5 | Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 3000 meter fra nærmeste bybanestopp - Super Learner | 97 |
| A3.6 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett | 98 |
| A3.7 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett | 98 |
| A3.8 | Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net | 99 |
| A3.9 | Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner | 99 |
| A3.10 | Gjennomsnittet av årlig prosentvis prisvekst for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett | 99 |

| | | |
|-------|---|-----|
| A3.11 | Gjennomsnittet av årlig prosentvis prisvekst for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett | 100 |
| A3.12 | Konfidensintervall for gjennomsnittet av predikerte leilighetpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net | 100 |
| A3.13 | Konfidensintervall for gjennomsnittet av predikerte leilighetpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner | 100 |
| A3.14 | Gjennomsnittet av årlig prosentvis prisvekst vekst for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett | 101 |
| A3.15 | Gjennomsnittet av årlig prosentvis prisvekst for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett | 101 |
| A3.16 | Konfidensintervall for gjennomsnittet av predikerte eneboligpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net | 102 |
| A3.17 | Konfidensintervall for gjennomsnittet av predikerte eneboligpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner | 102 |
| A3.18 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett | 102 |
| A3.19 | Gjennomsnittet av årlig prosentvis prisvekst for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett | 103 |
| A3.20 | Konfidensintervall for gjennomsnittet av predikerte boligpriser mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net | 103 |
| A3.21 | Konfidensintervall for gjennomsnittet av predikerte boligpriser mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner | 103 |

1 Innledning

Planleggingen av Bybanen i Bergen startet allerede på 1990-tallet, og har vist seg å bli en av de største infrastrukturelle nyvinningene på Vestlandet i nyere tid. I 2016 ble Bybanen fra Byparken til Vågsbotn i Åsane vedtatt i bystyret (Bergen Kommune, 2019a), og planleggingen av trasévalg og tilhørende byutvikling er nå et sentralt politisk tema. Fra et intervju med tidligere byråd for finans, innovasjon og eiendom i Bergen¹, Håkon Pettersen (H. Pettersen, 10. februar 2020), fremkommer det at politikerne i bystyret har et tydelig ønske om å legge til rette for utbygging av boliger nært Bybanen, samt begrense arealutnyttelsen i mer perifere områder. Den ønskede fortettingen rundt Bybanen er et resultat av at Bybanens hovedformål både er å fremme byutvikling og byens infrastruktur, samtidig som det miljømessige perspektivet settes høyt. Tilretteleggingen for eiendomsforvaltning rundt bybanetraseen er spesielt interessant for investorer som ser potensiale for gode investeringsmuligheter. I denne sammenheng er kommuneplanens arealplan (KPA) en sentral innretning å følge (H. Pettersen, 10. februar 2020). Bergen Kommune er ifølge Pettersen liberal når det gjelder frigjøring av areal til boligutbygging langs Bybanen.

Før Bybanens trasévalg bekreftes i Bergen kommune, legges forslaget frem for byrådet. I denne prosessen kan hvem som helst komme med innsigelser og bidra med aktiv lobbyvirksomhet for å fremme sitt synspunkt. Etttersom det er sterke interesser tilknyttet boligutbygging og boliginvestering langs Bybanen, vil tunge investorer ifølge Pettersen ha insentiv til å påvirke den endelige avgjørelsen av både trasévalg og arealplaner. På tross av at utbyggingen av Bybanen fra Byparken til Vågsbotn i Åsane, med estimert ferdigstillelse i 2031, er vedtatt, er det viktig å nevne at vedtaket ikke er endelig. Fra samtaler med Jarle Brattespe, medlem i utvalg for miljø og byutvikling i SV, fremgår det at prosjektets realisering hviler på en politisk overensstemmelse for både finansiering og trasévalg, spesielt rundt Bryggen (J. Brattespe, 15. august 2020). Dette medfører at ulike interessegrupper fremdeles har påvirkningskraft på utviklingen av både trasévalg og bystruktur mellom Byparken og Vågsbotn. På tross av usikkerheten tilknyttet Bybanens trasé til Åsane, er mange etablerte investorer allerede tungt investert i prosessen.

¹Håkon var i besittelse av stillingen fra november 2018 til oktober 2019

Ettersom introduksjonen av infrastrukturelle nyvinninger, som toglinjer og trikkelinjer, sannsynligvis vil påvirke boligprisene i område hvor det aktuelle godet introduseres, er det følgelig innsiktsfullt både for privatpersoner, investorer og politikere å besitte kvantitativ informasjon om hvordan boligprisene i et område vil utvikle seg ved utbyggingen av et offentlig gode som Bybanen. I denne utredningen ønsker vi derfor å øke informasjonsgrunnet bak beslutninger vedrørende byutvikling og investering tilknyttet den kommende Bybanen fra Byparken til Åsane. Dette vil vi gjøre ved å bygge prediksjonsmodeller for å svare på følgende problemstilling:

Hvordan ville utviklingen i boligprisene mellom Byparken og Åsane vært i perioden 2010-2019, i et tenkt scenario hvor Bybanen til Åsane hadde blitt ferdigstilt i 2010?

Vi ønsker å benytte maskinlæringsmetodene *Elastic Net* og *Super Learner* for å belyse denne problemstillingen.

Det er viktig å presisere at Bybanen til Åsane ikke eksisterer i dag. Gjennom utredningen ønsker vi kun å analysere hvordan boligprisene tilknyttet den planlagte Åsane-traseen ville utviklet seg *dersom* Bybanen til Åsane hadde vært et faktum fra 2010. På tross av at vi predikerer boligpriser for en foregående periode, er prediksjonene likevel relevante når vi ønsker å studere utvikling i boligpriser fra og med Bybanen til Åsane faktisk kommer og ni år frem i tid. Dette fordi den prosentvise utviklingen i boligprisene er uavhengig av det absolutte boligprisnivået på tvers av ulike perioder. Dette momentet vil utdypes ytterligere gjennom utredningen. For å hentyde til den predikerte boligprisutviklingen tilknyttet bybanetraseen mellom Byparken og Åsane for perioden 2010-2019 vil referanser som *“boligprisutviklingen fra og med Bybanens ferdigstillelse til Åsane og ni år frem i tid”* og *“med Bybanen”* benyttes i det følgende.

En sentral utfordring i utredningen er å bygge prediksjonsmodeller som fanger opp effekten Bybanen predikeres å ha på boligprisene rundt bybanetraseen til Åsane. Videre omtales denne effekten som *bybaneeffekten*. For at bybaneeffekten skal materialiseres i prediksjonene, må transaksjonsdata som benyttes for å trene og validere prediksjonsmodellene komme fra et område hvor Bybanen faktisk eksisterer. Ettersom det i dag kun eksisterer en bybane

mellom Byparken og Flesland, medfører dette at prediksjonsmodellene må bygges på observasjoner fra dette området. En utfordring med denne tilnærmingen er imidlertid at den planlagte bybanetrasseen mellom Byparken og Åsane ligger nord for Bergen sentrum, mens den eksisterende Bybanen mellom Byparken og Flesland ligger sør for sentrum. Det faktum at det alltid foreligger systematiske forskjeller mellom to ulike geografiske områder, er en utfordring i analysen. På grunn av områdespesifikk variasjon vil Bybanens påvirkning på prisen for et hus lokalisert i Fana² være ulik effekten Bybanen vil ha for prisutviklingen for et identisk hus i Åsane. Denne effekten er vanskelig å fange opp når modellen trenes på et annet geografisk område enn hva som predikeres for. Gjennom utredningen vil vi bygge et rammeverk som søker å løse problemet.

Selv om utredningens hovedformål er å øke informasjonsgrunnlaget bak beslutninger tilknyttet byutvikling og investering rundt Åsane-trasseen, kan også analysen benyttes som et rammeverk når formålet er å utarbeide prediksjoner hvor det kun foreligger tilgjengelig informasjon fra referanseprosjekter for andre geografiske områder. Dette vil eksempelvis være tilfelle hvor ny offentlig kommunikasjon planlegges.

Gjennom oppgaven har vi foretatt flere begrensninger som må utdypes. For det første er det utelukkende tatt høyde for å predikere priser på privatboliger omsatt gjennom det private eiendomsmarkedet. Hverken næringseiendommer eller eiendomstomter inkluderes i analysen. Dette innebærer at resultatene som fremgår av analysen ikke er direkte overførbare til andre boligtyper.

Videre tilhører prediksjonsmetodene som benyttes i analysen kun et utvalg av metoder innen to ulike maskinlæringsgrupper. Elastic Net er en lineær regresjonsmetode med mulighet for både *regularisering* og *variabelreduksjon*³, hvor Super Learner er en *samlemetode* som er satt sammen av en rekke ulike metoder, som følgelig spesifiseres i kapittel 4.4. Innenfor hver maskinlæringsgruppe finnes det et stort spekter av modelleringsmuligheter, i tillegg til at det foreligger en rekke maskinlæringsgrupper som ikke inkluderes i analysen. Metodevalget baseres på et ønske av å se om prediksjonenes presisjonen varierer når en kompleks metode som Super Learner benyttes sammenlignet med en relativt enkel og intuitiv lineær metode som Elastic Net. Det er også viktig å bemerke seg at utredningens formål er prediksjon og ikke *statistisk inferens*. Dette

²Bydel sør for Bergen som Bybanen fra Byparken til Flesland går gjennom.

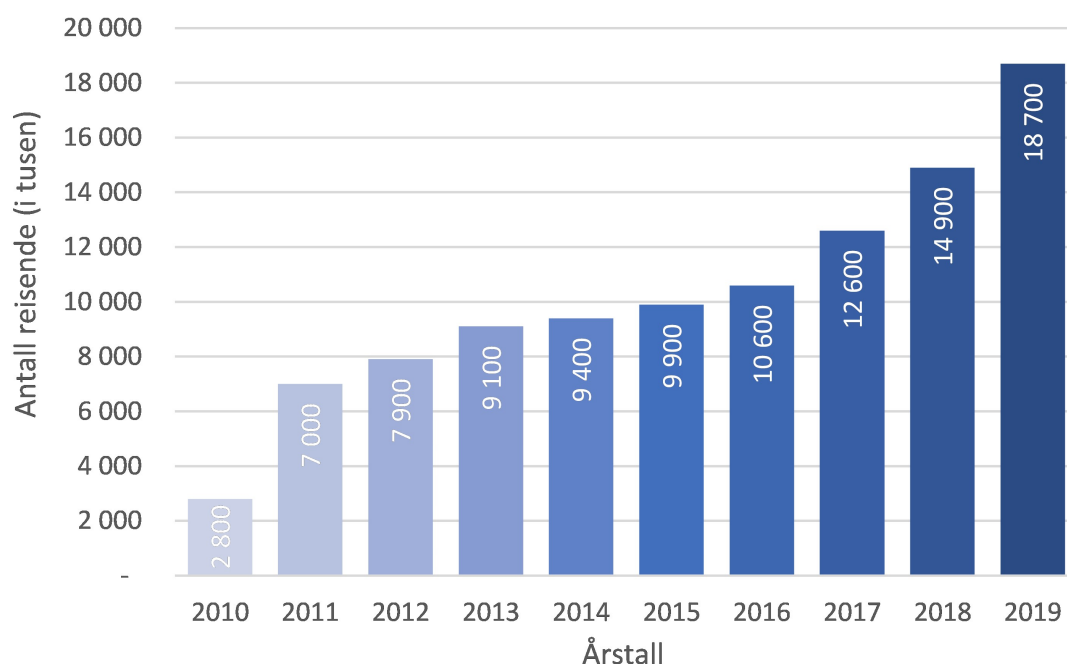
³Dette vil utdypes i kapittel 4.3

innebærer at koeffisientene som fremgår fra Elastic Net ikke kan tolkes *kausalt*. Utredningen vektlegger derfor andre statistiske momenter enn hva som vektlegges med et formål om kausale slutninger. Vi kan heller ikke utelukke at det finnes metoder som gir mer presise prediksjoner, eller høyere grad av intuisjon.

2 Bakgrunn

13. mars i år 2000 ble Bybanens første byggetrinn mellom Byparken og Nesttun vedtatt i byrådet, i januar 2008 ble traseen påbegynt og i juni 2010 åpnet linjen for offentlig drift. Fra ferdigstillelsen av første byggetrinn og frem til i dag har det vært en sterk vekst i antall reisende. Dette fremgår tydelig av Figur 2.1.

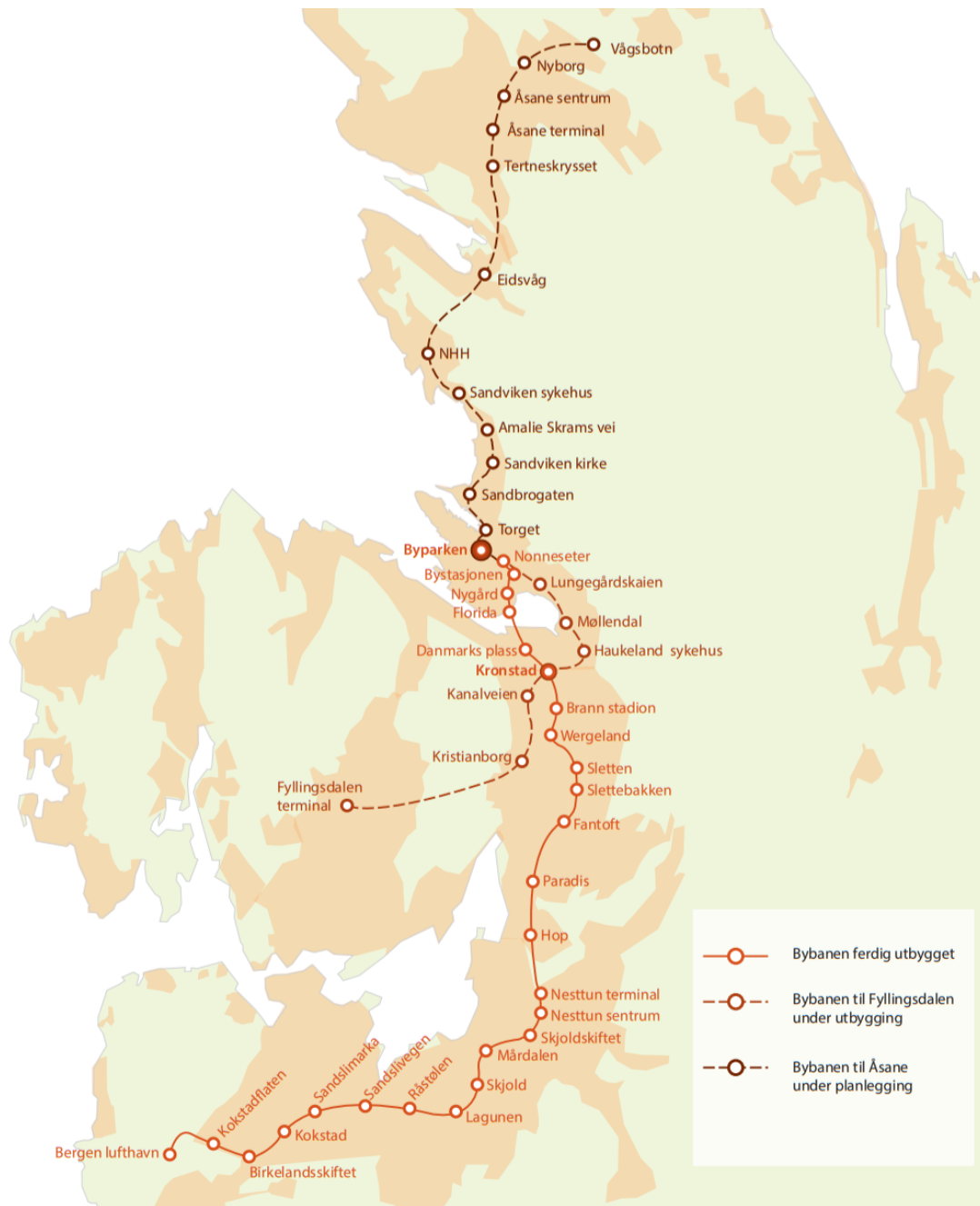
Figur 2.1: Antall reisende med Bybanen fra første byggetrinn (Byparken-Nesttun) ble ferdigstilt til og med 2019.



Fra figuren observerer vi at antall reisende nesten har syvdoblet seg fra 2010 til 2019. Selv om den sterke veksten delvis skyldes at Bybanen i dag dekker et større geografisk område enn i 2010, er det likevel liten tvil om at Bybanen har blitt en viktigere del av bybildet siden første byggetrinn ble ferdigstilt (H. Pettersen, 10. februar 2020). Tallene fra figuren er hentet fra Skyss sin årsrapport for 2019 (Skyss, 2020b)

I dag dekker Bybanen store deler av søndre Bergen. Utbyggingen og planleggingen for videre traseer, mot henholdsvis Fyllingsdalen og Åsane er viktige tema for videre byutvikling. Figure 2.2 viser oversikten både over ferdig utbygget og planlagt bybanetrasé for de aktuelle områdene.

Figur 2.2: Dagens Bybanetrasé med fremtidig utvidelse mot Fyllingsdalen og Vågsbotn, Åsane (Miljøløftet, 2020).



Eksisterende bybane starter ved Byparken og går til Bergen lufthavn (Flesland). Påbegynt trasé går fra Byparken til Fyllingsdalen terminal. Planlagt linje til Åsane går fra Byparken til Vågsbotn

Bybanetraseen fra Byparken til Åsane ble vedtatt 20. april 2016, men som nevnt i oppgavens innledning foreligger det politisk uenighet vedrørende Bybanens trasévalg rundt Bryggen⁴. Dette kan medføre at prosessen blir betraktelig forskjøvet og i verste fall skrinlagt (J. Brattespe, 15. august 2020). Alle holdeplasser er foreløpig bekreftet i

⁴Bryggen strekker seg fra Torget bybanestopp til Sandbrogaten bybanestopp.

byrådets vedteker, med unntak av Torgets, Sandbrogatens og Tertneskryssets eksakte plassering⁵. Estimert ferdigstilling er planlagt i 2031 (Matre, 2019). Fra tabell A1.1 i appendiks foreligger ytterligere informasjon om lengde, antall bybanestopp og reisetid for bybanetraseene som er ferdigstilt, samt planlagt trasé til Åsane.

Fra en artikkel publisert i nettavisen E24 (Lekve, 2013) uttalte eiendomsmegler og daglig leder i Krogsveen, Bergen, Fredrik Søreide, at en tre-roms leilighet i nærheten av Bybanen fort kunne være verdt 300 000 kroner mer enn en tilsvarende leilighet et stykke unna. Videre uttaler Asbjørn Myrstøl fra Garanti Eiendom at “Bergens nye kjæledegge, Bybanen, har gjort at prisene i søndre bydel har fått litt drahjelp. Bybanen gir to effekter, områdene blir kvitt store bilkøer og man kommer frem når man har planlagt” (Lekve, 2013). Disse artiklene, i tillegg til en stadig diskusjon rundt Bybanen og dets påvirkning på boligprisene, vekket vår interesse.

Vi fant ordlyden i uttalelsen fra Søreide noe uformell, og vi fikk derfor et inntrykk av at kommentaren var basert på generelle kunnskaper om boligmarkedet fremfor en presis kvantitativ analyse av den faktiske bybaneeffekt. Mistanken ble verifisert etter intervjuer med eiendomsmeglerne Fredrik Søreide fra Krogsveen (F. Søreide, 12. februar 2020) og Christina Erstad fra W Eiendomsmegling (C. Erstad, 15. februar 2020), hvor det fremkommer at estimatene av Bybanens innvirkning på boligprisene hovedsaklig er erfaringsbasert. Vi utforsket videre hvordan eiendomsmeglere verdivurderer boliger på en generell basis. Etter samtaler med eiendomsmeglerne fra Krogsveen og W Eiendomsmegling, supplert med informasjon fra DNB Eiendom sine nettsider (Iversen, 2019), fremgår det at boliger vurderes basert på prising av tilsvarende boliger gjennom Eiendomsverdi AS sin estimeringsmodell, i tillegg til meglers kunnskap om de spesifikke kvalitetene ved boligen. Fra samtaler med modell- og analysedirektør fra Eiendomsverdi AS, Anders Lund, kom det videre frem at deres estimeringsmodeller ikke inkorporer *bybanespesifikke variabler*⁶ (A. Lund, 25. Mars 2020). Eiendomsverdi AS sin estimeringsmodell vil diskuteres grundigere i kapittel 6.

Med bakgrunn i mangel på kvantitative analyser for Bybanens innvirkning på boligpriser, fant vi det som en interessant oppgave å tallfeste hvordan avstand til nærmeste bybanestopp

⁵Tertneskryssets plassering er ifølge Jarle Brattespe derimot ikke en politisk uoverensstemmelse som vil påvirke utbyggingen i stor grad.

⁶Vi definerer en bybanespesifikk variabel som en variabel som er direkte relatert til Bybanen.

har påvirket boligprisene tilknyttet allerede-eksisterende bybanetrasé mellom Byparken og Lagunen. Kvantifisering av bybaneeffekten var imidlertid allerede tallfestet og derfor mindre verdifull å analysere videre. Forskningen vi fant (Fredriksen, 2013) var skrevet i et økonometrisk perspektiv, hvor analysens formål var å reddykke Bybanens påvirkning på boligprisutviklingen i områdene rundt Bybanens første trasé mellom Byparken og Nesttun. Fredriksens analyse konkluderte med en positiv bybaneeffekt. Det vi derimot ikke kunne finne analyser av, hverken for Bergen eller for andre geografiske områder, var predikerte boligpriser ved en introduksjon av et offentlig gode som en toglinje eller en bybane.

I enhver verdivurdering hvor formålet er å prise et aktivum, som en bolig eller et verdipapir, står et godt informasjonsgrunnlag sentralt. De fleste finansielle prisingsformler avgjør dagens verdi basert på forventet fremtidig inntjening. En presis tallfesting av fremtidige verdier og inntjeningsmuligheter vil således være verdifullt. Eksempelvis vil nyheter om utbygging av nye tunneller trolig føre til økt boligprisvekst for nærliggende områder som et resultat av forventninger om positive eksternaliteter. Utfordringen med verdivurderinger av boliger, i tilfeller hvor det planlegges å bygge ut offentlig kommunikasjon, som Bybanen, er at det utarbeides få analytiske vurderinger av hvordan utbyggingen av et slikt offentlig gode vil påvirke boligprisene i tiden etter gode introduseres. I retrospekt kan vi gjennomføre økonometriske analyser for analysere hvordan prisene har utviklet seg, men er denne informasjonen direkte overførbar når en avgjørelse som påvirker fremtiden skal tas? Dersom eiendomsmeglere, politikere, investorer og privatpersoner med økt sikkerhet kan estimere boligprisenes prosentvise utvikling etter en strukturell endring som Bybanen, vil dette være informasjon som kan bidra til bedre verdivurderinger og investeringsstrategier.

Med bakgrunn i diskusjonen over er vår ambisjon å kunne gi eiendomsmeglere, politikere, investorer og privatpersoner kvantitativ informasjon om hvordan boligmarkedet rundt den planlagte Bybanetraseen mellom Byparken og Åsane vil utvikle seg etter Bybanens inntreden. Vi håper at dette kan tilføre innsikt til investorer og privatpersoner rundt deres investeringsstrategi, politikere i deres planleggingen av byutvikling og eiendomsmeglere i deres boligestimater. I tillegg håper vi at rammeverkets oppbygging kan være til hjelp i andre prediksjonssammenhenger hvor det også foreligger mangel på informasjon om observasjonene som skal predikeres.

2.1 Empiriske funn

Kvantitative analyser med fokus på infrastrukturens påvirkning på boligpriser er et sentralt forskningsemne. Blant annet finner Diaz og Mclean (1999) at en av de viktigste implikasjonene ved å introdusere nye trikkelinjer og toglinjer er dets påvirkning på eiendomsverdier for tilhørende områder. Cameron Murray (2016) finner videre at boliger innenfor en radius på 400 meter fra Gold Coast Light Rail (GCLR), Gold Coast, Australia, i snitt selges for 7,1% mer enn tilsvarende boliger lokalisert mer enn 400 meter unna. I likhet finner Jones (2015) i sitt metastudie at boliger som ligger innenfor en radius på 400 meter fra nærmeste lettbane⁷ vil oppleve en prisøkning på 9,5% sammenlignet med tilsvarende boliger som ligger lengre borte. Gallo (2018) påpeker at tilgjengeligheten til lettbaner og kvaliteten på linjene er avgjørende for boligprisene for det aktuelle område. I urbane områder hvor det er tilgang til transportsystemer av høy kvalitet vil boligprisene være høyere.

Gallo (2018) analyserer videre hvordan togsystemet i Napoli har påvirket boligprisene tilknyttet togstasjonene. Han finner at den prosentvise prisøkningen, som et resultat av nærhet til nærmeste togstasjon, hvor togene har avganger oftere enn seks ganger i timen, er stor for mange områder. I rapporten er det aktuelle område i Napoli delt inn i 60 soner, hvor det fremkommer at boligprisene er 20% høyere for to av sonene, 10% høyere for ni av sonene og 5% høyere for ni av sonene. For de resterende 37 sonene er effekten under 5%. For å konkretisere den totale effekten estimeres boligene i Napoli å øke med totalt 7,2 milliarder euro som en direkte effekt av toglinjene i byen, noe som i gjennomsnitt tilsvarer en økning på 23,034 euro per boenhet på 100 kvadratmeter. Denne verdistigningen tilfaller eiendomseierne, og innføringen av lokale skatter og avgifter fremlegges i analysen som en interessant tanke. Det Gallo ikke vektlegger i sin utredning er nærheten til togstasjonene og avstandens betydning for eiendomsprisene.

Dorantes et al. (2011) ser derimot på meter fra nærmeste togstasjon som en kontinuerlig variabel innenfor en radius på 2000 meter fra de respektive togstasjonene i utkanten av Madrid. Resultatene viser at hus som ligger 1000 meter fra Metrosurs toglinje i snitt selges for mellom 2.18% og 3.18% mindre enn et tilsvarende hus i umiddelbar nærhet til

⁷Lettbane er "light rail" på engelsk og er en form for en urban toglinje med lavere kapasitet enn en metro, men med høyere fart enn en tradisjonell t-bane (Boorse, 2001). Bybanen kategoriseres som en lettbane.

stasjonene. Umiddelbar nærhet er i denne sammenheng definert som avstand under 100 meter. Tilsvarende gjelder for hus som ligger nærme Cercanias toglinje i Madrid, bare at her er påvirkningen fra 3.38% til 5.17%, avhengig av det spesifikke område. Golub et al. (2012) finner at avstanden til lettbanestasjoner generelt har en positiv effekt på boligpriser, og at effekten avtar med økt avstand fra stasjonen. Prisen for eneboliger som ligger helt nært til baneavgangene blir derimot lavere. Aliyu et al. (2018) finner gjennom et metastudie at nærhet til offentlig skinnetransport som oftest har en positiv effekt, men henviser til tilfeller hvor effekten har vært negativ eller neglisjerbar. Det konkluderes med at dette ofte er et resultat av negative eksternaliteter som dårlig luft, lyd og visuell forurensning.

Det foreligger således mye empiri som viser til at nærhet til toglinjer har en positiv effekt på boligprisene. Noe som underbygger at dette også er tilfellet i Bergen, er Fredriksens masterutredning om Bybanens trasé mellom Byparken og Nesttun sin påvirkning på boligprisene for område tilknyttet den aktuelle trasseen (Fredriksen, 2013). Fredriksen (2013) finner at Bybanen har ført til en økning i boligprisene på 5.41% for boliger som ligger innenfor en radius på 100 meter fra Bybanens holdeplasser, mens boliger som ligger innenfor 1000 meters radius fra holdeplassene har 3.98% høyere pris. Når det gjelder inkluderte variabler, er Fredriksens analyse noe mangelfull. Vi ønsker derfor i denne analysen å se på et større spekter av sentrale variabler for å øke prediksjonenes presisjon⁸.

Det er viktig å presisere at all empiri nevnt over søker å forklare den *kausale* effekten av nærhet til togtransport på boligprisene. Vi ønsker derimot i denne utredningen å *predikere* boligpriser mellom Byparken og Vågsbotn for perioden 2010-2019, gitt Bybanens inntreden til Åsane i 2010. Som nevnt innledningsvis er det en sentral utfordring at prediksjonene baseres på boligtransaksjoner mellom Byparken og Flesland ettersom boligene i dette området er de eneste observasjonene som påvirkes av Bybanen. Dette tilfører forskningsfeltet rundt boligpriser og offentlig transport ny innsikt om hvordan kunnskap kan overføres fra et geografisk område til et annet når et offentlig gode som Bybanen introduseres. Det er viktig å presisere at denne utredningen ikke tar høyde for *statistisk inferens*, noe som impliserer at en kvantifisering av Bybanens påvirkning på boligprisene ikke kan tolkes som et direkte resultat. Det vektlegges i midlertid å skape presise prediksjoner for boligprisene tilknyttet den kommende Åsane-traseen.

⁸Dette vil gjennomgås i kapittel 4

3 Datasett

Fra Fredriksens utredning (2013) om Bybanens påvirkning på boligprisene i området mellom Byparken og Nesttun, ble den *hedonistiske prisingsmodellen*, introdusert av Sherwin Rosen (Rosen, 1974), benyttet som rammeverk for variabelutvelgelse. På tross av at Fredriksens formål er statistisk inferens, kan det samme rammeverket for variabelutvelgelse også gi oss en pekepinn på hvilke variabler som er sentrale selv om formålet er prediksjon.

I dette kapitlet presenteres først den hedonistiske prisingsmodellen, før de ulike datakildene og de inkluderte variablene beskrives. Hvordan data er renset og justert, i tillegg til potensielle sammenhenger mellom forklaringsvariablene og avhengig variabel, blir videre presentert og analysert. Deretter vil utfordringer tilknyttet overfitting diskuteres, før vi avslutningsvis diskuterer datasettets inndeling i treningssett, valideringssett og testsett.

3.1 Den hedonistiske prisingsmodellen

Mariano Gallo (2018) påpeker at nesten all forskning som søker å forklare eiendomsverdi og dets korrelasjon med transportsystemers tilgjengelighet benytter en hedonistisk prisingsmodell som et rammeverk for variabelutvelgelse. Dette underbygges av Osland (2001) som også påpeker at modellen er mye anvendt i studier av prisdannelse i boligmarkedet. En hedonistisk prisingsmodell søker å forklare prisen på et sammensatt gode, i vårt tilfellet boliger, ved hjelp av sammensetningen av egenskapene, omtalt som attributtene, som utgjør godet.

Attributtene i den hedonistiske prisingsmodellen deles ofte inn i tre samlevARIABLER: *Lokasjonsattributter*, *miljømessige attributter* og *strukturelle attributter*. Strukturelle attributter innebærer alle variabler som utgjør selve boligen. Informasjon om boareal, tomtestørrelse og antall rom står her sentralt. Lokasjonsattributter inkluderer variabler som beskriver nærområdet, hvor nabolagets levekår er viktig. Miljømessige karakteristikker omhandler blant annet variabler som luftkvalitet, nærhet til sentrum og infrastruktur (Osland, 2001). Alle inkluderte variabler vil bli presentert i detalj fra kapittel 3.2.

3.2 Inkluderte variabler

For å oppnå presise prediksjoner er det sentralt at variablene som inkluderes i datagrunnlaget vårt dekker flest mulig aspekter som er styrende for hvordan en bolig verdivurderes i markedet. I tillegg er det viktig at observasjonene som inkluderes i opptreningen og valideringen av modellene inneholder tilstrekkelig variasjon. Tilstrekkelig variasjon er avgjørende for at modellene skal fange opp reelle sammenhenger for hvordan boligpriser oppstår.

For oversikten sin del er inkluderte variabler delt inn i underkategorier i henhold til den hedonistiske prisingsmodellen. Vi har også inkludert to makroøkonomiske variabler: Vekst i BNP og renter.

3.2.1 Salgspris

Vi presiserte i kapittel 1 at fokuset i denne utredningen er å predikere salgspris på boliger omsatt gjennom det private eiendomsmarkedet. Transaksjonsdata, som utgjør priser for solgte boliger, strekker seg fra 1. januar 2000 til 31. desember 2019 og er hentet fra Eiendomsverdi AS, et privateid uavhengig selskap som til enhver tid har et sanntidsbilde av boligprisene for alle landets boligmarkeder (Eiendomsverdi, 2020a). For å sammenligne boligpriser uavhengig av boligenes størrelse har vi dividert total salgspris på antall kvadratmeter for å oppnå kvadratmeterpris.

Boligtransaksjonene mellom år 2000 og 2019 er videre delt inn i tre ulike datasett, basert på geografisk plassering. Det første datasettet inneholder transaksjonspriser for solgte boliger lokalisert innen 3000 meter fra nærmeste bybanestopp tilknyttet bybanetraseen mellom Byparken og Nesttun. Det andre datasettet inneholder boligtransaksjoner for solgte boliger som ligger nærmere enn 3000 meter fra nærmeste bybanestopp tilknyttet andre bybanetrasé (Nesttun-Lagunen), mens det siste omfatter transaksjonsdata for boliger som ligger innen 3000 meter fra tredje bybanetrasé (Lagunen-Flesland).

Grunnen til at boliger som ligger mer enn 3000 meter unna nærmeste bybanestopp ekskluderes fra analysens datagrunnlag, er at Eiendomsverdi konkluderer med at toglinjen mellom Viken fylke⁹ og Oslo sin påvirkning på boligpriser er neglisjerbar for boliger

⁹Viken er tidligere Akershus fylke

lokalisert mer enn 2000 meter fra nærmeste togstasjon (Eiendomsverdi, 2020b). Vi kan derfor anta at vi ikke ekskluderer sentrale observasjoner ettersom vi ønsker predikere boligpriser hvor bybaneeffekten potensielt er økonomisk signifikant.

3.2.2 Inflasjonsjustering og sesongjustering

Ettersom vi observerer salgspris over tid, må vi vurdere om boligprisene skal inflasjonsjusteres eller ikke. Litman Donovan et al. (2008) fremlegger viktigheten av å se på reell prisstigning, hvilket krever inflasjonsjusterte priser. På den andre siden legger Diamond et al. (1997) vekt på at folk er mest opptatt av nominelle priser og således at en inflasjonsjustering ikke er hensiktsmessige i alle tilfeller. I vår sammenheng kan det argumenteres for at en inflasjonsjustering er nødvendig. Dette begrunnes med at data strekker seg over en lang periode i tillegg til at boligtransaksjonenes størrelse utgjør en vesentlig investering for folk flest. Det vil ikke være rimelig å sammenligne nominelle priser for to boliger som er solgt med 10 års mellomrom når realverdien av disse boligene trolig avviker betydelig.

Spørsmålet blir videre hvilket inflasjonsmål det skal justeres for. Kihwan Seo (2016) anbefaler å benytte en boligprisindeks som justeringsmål for å fange opp den generelle boligprisveksten. SSB har en egen boligprisindeks for Bergen som strekker seg fra 2003 til 2019 (SSB, 2020b). Ettersom transaksjonsdata starter allerede i år 2000, har vi lagt til grunn den landsdekkende boligprisindeksen for de tre første årene. Dette anses som en god tilnærming ettersom boligprisindeksene for Bergen og for Norge har vært tilnærmet identiske fra 2003 til 2019. Ettersom det foreligger sesongvariasjoner i salgspris har vi også valgt å sesongjustere observasjonene etter SSB sine justeringsmål. Denne justeringen er imidlertid lagt til som et valg når boligprisindeksen til SSB benyttes. I tillegg til å justere boligprisene i henhold til veksten i boligmarkedet, er det også nødvendig å justere for den generelle kjøpekraften. Dette er gjort ved å justere boligtransaksjonene med SSB sin konsumprisindeks (SSB, 2020a).

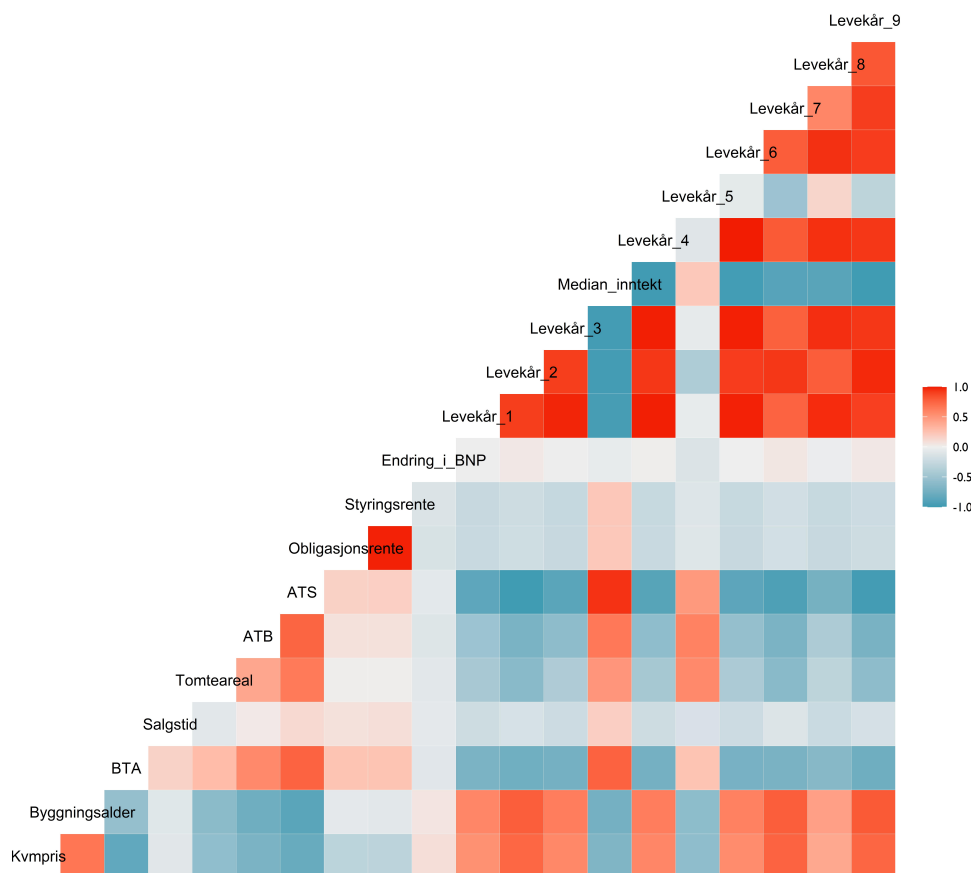
Oppsummert er boligprisene i denne utredningen både sesongjustert, i tillegg til å være inflasjonsjustert etter både konsumprisindeksen og boligprisindeksen til 2019-priser. Dette gjør det mulig å sammenligne transaksjoner på tvers av ulike perioder.

3.3 Variabler fra den hedonistiske prisingsmodellen

3.3.1 Lokasjonsattributter

Tilgjengelighet er en nøkkelfaktor når lokasjonsattributter skal analyseres (Osland, 2001). Med tilgjengelighet menes hvor lett lokale fasiliteter kan nås fra en bolig. Her står avstand til både nærmeste forretningsområde og offentlig transport sentralt (Srouf et al., 2002). Etersom Bergen er preget av store topografiske ujevnheter vil luftavstand mellom bosted og interessepunkt være lite representativt for tiden det tar å komme seg fra en bolig til en bestemt lokasjonen. Vi har derfor utarbeidet kategoriske variabler som representerer gåavstand fra hver bolig til nærmeste bybanestopp og kjøreavstand til sentrum. I tillegg har vi opprettet dummyvariabelen *nærhet til nærmeste bykjerne*, hvor en bolig anses som nær dersom den er lokalisert innen 1000 meter fra en bykjerne. I denne utredningen er følgende definert som bykjerne: Byparken Bybanestopp, Åsane senter, Lagunen senter, Nesttun senter og Sletten senter.

En korrelasjonsmatrise er et nyttig verktøy for å studere de inkluderte variablenes samvariasjon, og er presentert i Figur 3.1. I figuren indikerer blå farge negativ korrelasjon, eller samvariasjon, hvor rød farge impliserer positiv korrelasjon. Fra matrisen fremgår det eksempelvis at avstand til sentrum (ATS) og avstand til nærmeste bybanestopp (ATB) er negativt korrelert med kvadratmeterpris (Kvmpris).

Figur 3.1: Korrelasjonsmatrise

Tabell A2.1 i appendiks gir en helhetlig presentasjon over de inkluderte variablene, hvor forkortelser som benyttes i korrelasjonsmatrisen forklares. Her fremlegges variablenes opprinnelse, måleenhet og kilde. Levekårsvariablene refererer til de miljømessige attributtene presentert i kapittel 3.3.2.

3.3.1.1 Innsamling av gåavstand fra hver bolig til nærmeste bybanestopp

Ettersom det ikke foreligger noen form for oversikt over gåavstand fra hver bolig i Bergen til nærmeste bybanestopp, måtte vi utarbeide dette på egenhånd. Gåavstand mellom et start- og endpunkt kan manuelt beregnes gjennom Google Maps. Ettersom vi hadde behov for å gjennomføre 41 077 beregninger mellom Byparken og Flesland, i tillegg til 17 517 beregninger mellom Byparken og Vågsbotn, var det nødvendig å automatisere innsamlingen. Ved å opprette en funksjon i Google Maps Platform som henter ut informasjon direkte fra Google sine servere, fikk vi muligheten til å gjennomføre mangfoldige beregninger

gjennom Google Sheets. Koden som ble brukt til å hente ut og beregne gåavstandene ble utarbeidet i JavaScript.

Det er flere forhold som gjorde datainnsamlingen til en langtekkelig prosess. For det første legger Google sin API sterke begrensninger i antall daglige beregninger som kan gjennomføres. For det andre finner ikke Google Maps automatisk nærmeste Bybanestopp fra den aktuelle boligen. Google Maps beregner kun gåavstand mellom to geografiske punkter, og det er tilnærmet umulig å vite hvilket bybanestopp som ligger nærmest hver bolig. For å finne nærmeste bybanestopp for hver bolig, sorterte vi først inn boligene etter postnummer. Deretter fant vi hvilke tre bybanestopp som lå nærmest området tilknyttet hvert postnummer. Gåavstanden fra hver bolig tilhørende det aktuelle postnummeret ble følgelig beregnet mot de tre nærmeste bybanestoppene. Det bybanestoppet med kortest gåavstand fra den aktuelle boligen ble videre inkludert i datasettet. På denne måten komprimerte vi antall kalkulasjoner, samtidig som vi sikret oss korrekte avstandsberegninger. Likevel medførte prosessen et stort mangfoldige estimeringer, hvor hver beregning krevde betydelig datakapasitet.

For å lokalisere de fremtidige bybanestoppene mellom Byparken og Vågsbotn, benyttet vi kart med oversikten over bybanestoppene utarbeidet av Norconsult og Viak (2020) som utgangspunkt.

3.3.1.2 Innsamling av kjøreavstand Bergen sentrum

Ettersom avstand til sentrum har en vesentlig innvirkning på boligpriser (Srouf et al., 2002) valgte vi også å utarbeide en egen avstandsvariabel for kjøreavstand fra hver bolig til sentrum. Byparken bybanestopp er i denne utredningen definert som Bergen sentrum. For å innhente nødvendig informasjon benyttet vi samme fremgangsmåte som beskrevet i kapittel 3.3.1.1. Vi endret imidlertid transportmiddel fra “gange” til “bil” ettersom Google Maps, ved bruk av funksjonen “gange”, ofte beregner ruter som ikke vil bli brukt dersom avstanden til sentrum blir for stor.

3.3.1.3 Innsamling av gåavstand til nærmeste bykjerne

Gåavstand til nærmeste bykjerne ble også beregnet gjennom Google Maps Platform, hvor hver bolig ble delt inn etter postnummer for å finne nærmeste bykjerne.

3.3.2 Miljømessige attributter

Den hedonistiske prisingsmodellen vektlegger videre viktigheten av å analysere hvordan miljømessige attributter påvirker boligprisene. Miljømessige variabler beskriver karakteristikken av området hvor boligen ligger. Det finnes imidlertid ingen felles konsensus om hvilke variabler som bør inkluderes for å fremme effekten av nærmiljø på boligpriser (Osland, 2001). I denne analysen har vi inkludert en rekke levekårsvariabler for de ulike bydelene i Bergen. Variablene er hentet fra en helse- og levekårsrapport utarbeidet av Bergen Kommune (2016). Variablene er som følger: Andel med dårlig norsk språk, barn med enslige foreldre, antall barn i lavinntektshusholdninger, median inntekt, unge uførepensjonister, barn med barnevernstiltak, kommunale utleieboliger per 1000, anmeldelser per 1000 og barneutflyttinger 0-5 år. Variablene er bydel-spesifikke, noe som gjør det mulig å skille mellom levekår i ulike bydeler.

Rapporten fra Bergen kommune påpeker at de ovenfornevnte variablene kan være korrelert med levekår og økonomisk stabilitet. Ifølge rapporten kan en økning i både innvandring og andelen innbyggere med dårlig norsk språk påvirke økonomi og helse negativt. I tillegg vil oppvekst med én forelder stå i sammenheng med dårligere prestasjoner på skolen, frafall og lavere kjøpekraft. Forskning gjennomført av SSB viser til at 22% av totalt forbruk brukes på bolig (Vegard, 2018) og median inntekt er derfor en sentral variabel å inkludere. Det samme gjelder andel barn i lavinntektshusholdninger.

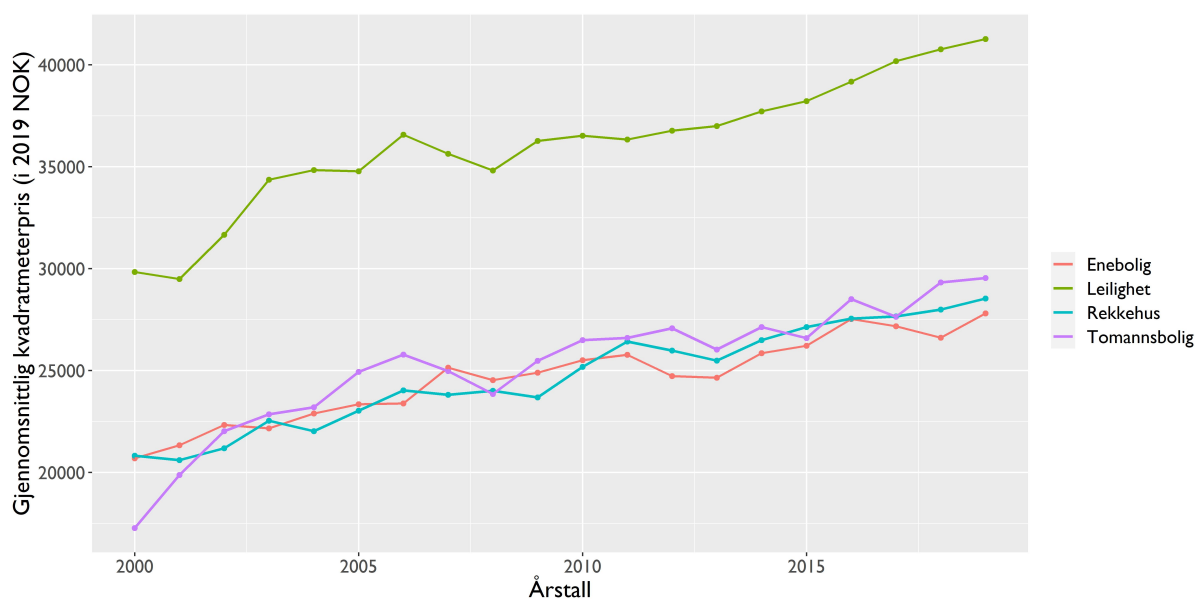
Videre sier andel unge på uføretrygd, definert som andel personer mellom 18 og 44 år som mottar uføretrygd, andel barn med barnevernstiltak, definert som andelen barn mellom 0 og 17 år med barnevernstiltak i løpet av 2014, anmeldelser per 1000 innbygger og barneutflytting også noe om levekår. Hypotesen er at disse variablene er negativt korrelert med boligprisene ettersom husholdninger som befinner seg i en eller flere av disse kategoriene har lavere levevilkår noe som kan tyde på lavere kjøpekraft. Sammenhengene nevnt over viser seg å stemme godt overens med hva vi observerer i korrelasjonsmatrisen (Figur 3.1).

3.3.3 Strukturelle attributter

Strukturelle attributter er egenskaper som utgjør den direkte verdien av en bolig. De mest fremtredende strukturelle attributtene er bruksareal, tomteareal og byggeår, og er alle hentet fra Eiendomsverdi sine databaser. Når disse variablene analyseres gjennom korrelasjonsmatrisen er det sentralt å bemerke seg at avhengig variabel er kvadratmeterpris, og ikke boligens totale salgspris. Dette innebærer eksempelvis at tomteareal er negativt korrelert med kvadratmeterpris, noe som er naturlig ettersom den marginale betalingsviljen for en ekstra kvadratmeter boareal er synkende. Boligene er videre delt inn etter følgende boligtyper: Enebolig, tomannsbolig, rekkehus og leilighet.

Figur 3.2 viser historiske priser for de respektive boligtypene fra vårt datasett, hvor leiligheter fremstår som klart dyrest. Alle boligpriser i utredningen er sesongjustert, i tillegg til å være inflasjonsjustert etter både konsumprisindeksen og boligprisindeksen til 2019-priser.

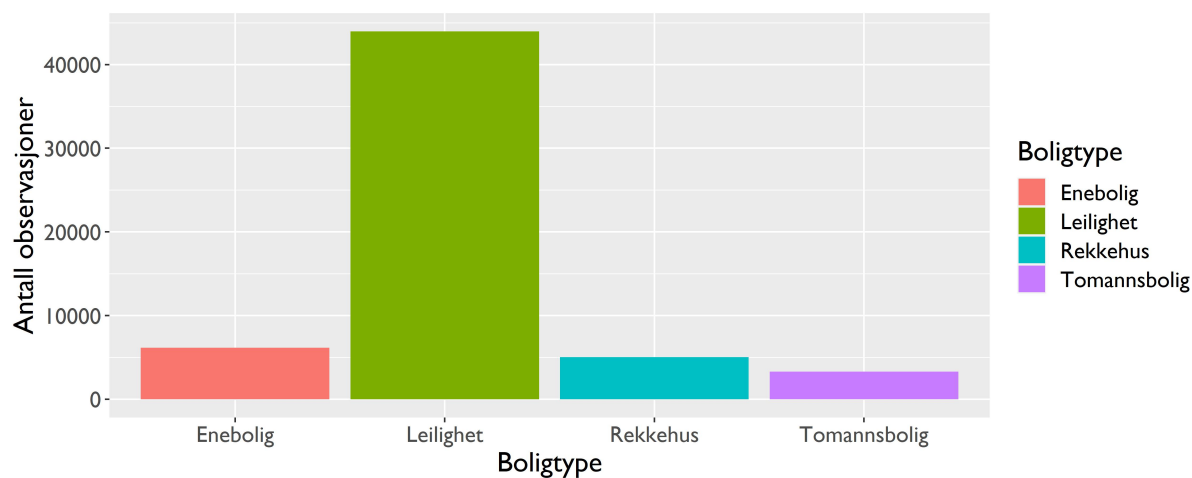
Figur 3.2: Utvikling i kvadratmeterpris for inkluderte boligtyper mellom år 2000 og 2019



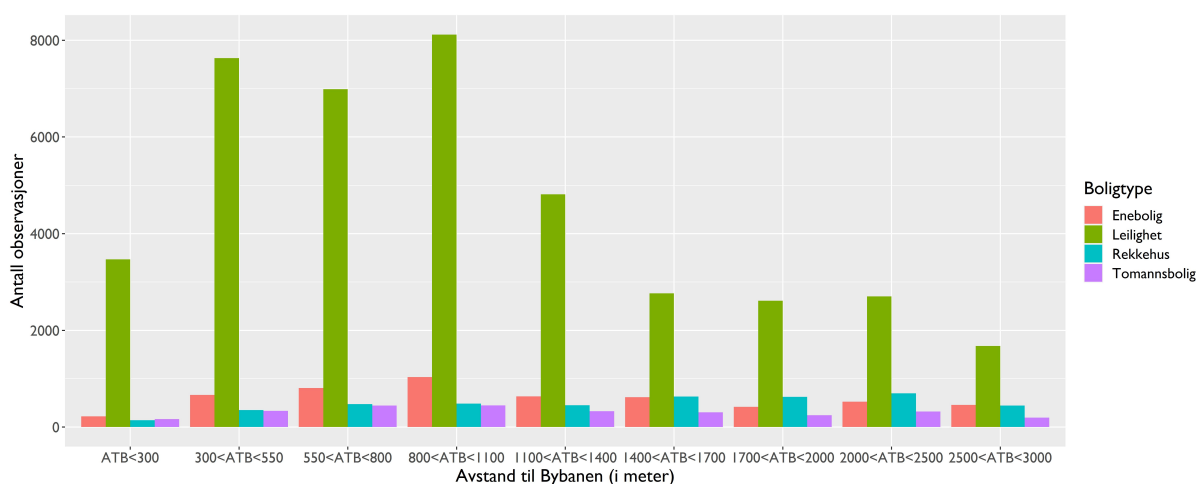
Prisnivået representerer gjennomsnittlig kvadratmeterpris for alle boliger innen 3000 meter henholdsvis fra den eksisterende bybanetraseen mellom Byparken og Flesland, og den kommende traseen mellom Byparken og Åsane

Figur 3.3 illustrerer videre en viktig svakhet med datasettet vårt. Fra figuren fremgår det at størsteparten av det totale datagrunnlaget består av leiligheter.¹⁰ Figur 3.4 viser videre fordelingen for de ulike boligtypene kategorisert etter avstand til nærmeste bybanestopp. For alle boligtyper, bortsett fra leilighet, foreligger det relativt få observasjoner innen mange av de inkluderte avstandsdummyene. Dette kan i sin tur påvirke prediksjonenes presisjon ettersom modellene ikke fanger opp relevant variasjon i datagrunnlaget. Denne svakheten vil diskuteres videre i kapittel 6.4.

Figur 3.3: Fordeling av de inkluderte boligtypene



Figur 3.4: Fordeling av boligtyper for ulike avstander fra nærmeste bybanestopp langs Byparken-Lagunen



¹⁰Overvekten av leiligheter er trolig et resultat av høyere omløpshastighet i tillegg til at det finnes en relativt stor andel leiligheter innen 3000 meter fra Bybanen.

3.3.4 Makroøkonomiske variabler

Ved å supplere den hedonistiske prisingsmodellen med makroøkonomiske variabler, kan vi potensielt oppnå bedre prediksjoner. Derfor er renter og bruttonasjonalprodukt (BNP) også inkludert som en del av datagrunnlaget.

3.3.4.1 Renter

Ifølge Sutton et al. (2017) har kortsiktig utlånsrente en sterk og vedvarende effekt på boligpriser. Kortsiktige renter er i tråd med Sutton et al. (2017) basert på en tremåneders statskasseveksler. Etttersom boligkjøpere ofte binder rentene for en lengre periode, inkluderes ofte også langsiktig rente med utgangspunkt i 10-årige statsobligasjoner i datagrunnlaget hvor effekter på boligpriser analyseres (Sutton et al., 2017).

Utfordringen med kortsiktige renter er det ikke finnes datamateriale for tremåneders statskassevekslere før 2003. Som en god tilnærming benytter vi i stedet styringsrenten ettersom den både er styrende for tremåneders statskassevekslere og tilgjengelig for hver periode fra år 2000 til 2019. Fra korrelasjonsmatrisen antydes derimot en svak korrelasjon mellom både styringsrenten og den langsiktige renten mot kvadratmeterpris. Alle data for renter er hentet fra Norges Bank (2020).

3.3.4.2 BNP

Videre er bruttonasjonalprodukt (BNP) et sentralt mål på økonomisk aktivitet og benyttes ofte som et kvantitativt mål på utviklingen i et lands økonomi (SSB, 2017). BNP betegnes som summen av de varer og tjenester som produseres for inneværende år, minus kostnader tilknyttet tilhørende produksjon. BNP er omtalt som det viktigste målet på verdiskaping i et land, og endring i BNP benyttes ofte som et mål på nedgang og oppgang i et lands økonomiske tilstand (SSB, 2017). Et lands økonomiske tilstand er videre knyttet til husholdningenes privatøkonomiske situasjon, som i sin tur henger sammen med boligpriser (Vegard, 2018). BNP er hentet fra SSB sin database og inflasjonsjustert til reelle tall ved hjelp av konsumprisindeksen.

3.4 Rensing og justering av data

For å bruke datasettet til å bygge prediksjonsmodeller, ble alle observasjoner med ufullstendig adresse eller bruttoareal lik null fjernet. Tomtearealet for alle tomannsboliger ble videre delt i to ettersom Eiendomsverdi i sine databaser kun presenterte sammenslått areal. Mange av tomtene eies trolig i fellesskap mellom to sammenhengende tomannsboliger. Det kan tenkes at en tomt som eies i fellesskap har lavere verdi per kvadratmeter tomteareal sammenlignet med en tomt som besittes av én eier. Ettersom vi ikke har informasjon om hvilke tomter som eies i fellesskap og hvilke som ikke gjør det for de ulike tomannsboligene, valgte vi å dele tomteareal i to for alle tomannsboliger. Dersom det foreligger en signifikant prisforskjell mellom tomannsboliger med egen tomt og tomannsboliger med felles tomteareal, og tomteareal oppfattes av prediksjonsmodellene som sentrale, vil presisjonen til prediksjonene potensielt bli dårligere enn hvis vi hadde klart å tildele alle tomannsboligene korrekt tomteareal. Dette er en mulig svakhet med analysen, som trolig vil ha liten betydning for presisjon. Totalt antall observasjoner innen 3000 meter fra nærmeste bybanestopp, for både eksisterende bybane og kommende bybane er 58594.

3.4.1 Variabeltransformasjon

Elastic Net, en av de to prediksjonsmetodene vi benytter, antar lineære sammenhenger mellom avhengig og uavhengig variabel. Dersom sammenhengen i virkeligheten er ikke-lineær, vil en lineær modell ha problemer med å fange opp den sanne variasjonen i datasettet. Dette kan løses ved å transformere variablene slik at de fremstår på en lineær form. Figur A2.2 til A2.8 i appendiks viser fordelingen til de variablene hvor variabeltransformasjon førte til økt linearitet. Det fremgår tydelig at en logaritmisk transformasjon av *kvadratmeterpris*, *salgstid* og *tomteareal* fører til mer normalfordelte observasjoner hvor en lineær linje representerer observasjonene på en bedre måte. For *bygningssalder* ser det imidlertid ut til at en kvadrert transformasjon gir best lineær tilpasning. Rent intuitivt gir det også mening at disse variablene burde transformeres. Eksempelvis er det rimelig å anta at kvadratmeterpris vil falle med økt boligstørrelse, noe som fører til at en logaritmisk transformasjon er hensiktsmessig. I resultatdelen vil vi se om prediksjonsmodellene oppnår høyere presisjon når variablene bygningssalder, salgstid

og tomteareal transformeres, sammenlignet med presisjonene vi oppnår når vi predikerer uten transformerte variabler.

3.4.2 Kritikk mot manglende data

Datasettet vårt dekker ifølge den hedonistiske prisingsmodellen mange sentrale aspekter ved prisdannelsen av en bolig. Likevel foreligger det flere variabler som kunne vært inkludert for å oppnå høyere presisjon. Eksempelvis ville flere avstandsparametre, boligattributter og miljømessige attributter potensielt ført til bedre prediksjoner. I oppbyggingen av en prediksjonsmodell er det alltid en avveining mellom et økt antall variabler og modellkompleksitet, noe som vil diskuteres i neste delkapittel. Likevel innehar både Elastic Net og Super Learner karakteristikker som gjør det mulig å overgå problemet tilknyttet overfitting når flere variabler inkluderes. Dette fører til at en inkludering av flere variabler er mulig uten at det går på bekostning av modellenes presisjon. En annen utfordring er at det foreligger få boligtransaksjoner for enkelte boligtyper. Dette kan skape ujevne resultater, noe som vil diskuteres i kapittel 6.

3.5 Overfitting og underfitting

I en tid med god tilgang på data, i tillegg til en stadig utfoldelse av nye og mer avanserte maskinlæringsmetoder, er mulighetene for å oppdage komplekse datastrukturer bedre enn noen gang. På tross av at maskinlæringsmetoder uten forhåndsprogrammerte innstillinger åpner opp for å oppdage komplekse strukturer i data uten omfattende programmeringskunnskaper, kreves det likevel at sluttbrukeren er observant på sentrale statistiske sammenhenger. I de tilfellene hvor sluttbruker er opptatt av å produsere presise prediksjoner er avveiningen mellom økt kompleksitet og høyere test-residualer ¹¹ et sentralt hensyn å ta.

Modellens kompleksitet fremgår ulikt for ulike modeller innen maskinlæring. Ofte vil modellens kompleksitet være et resultat av antall funksjoner eller vilkår som inkluderes, om modellen er lineær eller ikke-lineær og hvor kompleks den underliggende algoritmen er (Tayo, 2019). Komplekse modeller kjennetegnes ved at de er mindre tolkbare, ofte krever økt beregningskraft i tillegg til at de står i fare for å overtilpasse prediksjonene

¹¹Test-residualer er forskjellen mellom den predikerte verdien, \hat{y} og den sanne verdien, y fra testsettet.

"out-of-sample"¹² til treningsdata. Dersom modellens kompleksitet øker, eksempelvis ved å legge til et parameter i en lineær regresjon, vil modellens tilpasning til treningsdata alltid bli bedre, uavhengig av hvor betydningsfull den inkluderte variabelen er for å forklare den sanne sammenhengen i data. Dette er en grunnleggende egenskap ved statistiske modeller (Fortmann-Roe, 2012). På tross av at en modell har god tilpasning for treningsdata, impliserer ikke dette nødvendigvis en god tilpasning for data out-of-sample. Forholdet mellom prediksjonsavvik for treningsdata og testdata kan presenteres på følgende måte¹³:

$$\text{Det sanne prediksjonsavviket} = \text{Residualene til treningsdata} + f(\text{modellens kompleksitet}) \quad (3.1)$$

Ligningen over viser at prediksjonsavviket er et resultat av en avveining mellom residualer for treningsdata og modellens kompleksitet, hvor økt kompleksitet fører til lavere residualer for treningsdata, men høyere prediksjonsavvik. Dette medfører at kompleksitet kun er ønskelig frem til et visst nivå. Dersom modellens kompleksitet blir for høy, vil den slite med å skille støy¹⁴ fra underliggende signal¹⁵ (Fortmann-Roe, 2012). Denne utfordringen omtales som *overfitting*. Underfitting oppstår når det motsatte er tilfellet - at modellen er for enkel og ikke fanger opp underliggende signal fra treningssettet.

Diskusjonen rundt overfitting og underfitting er elementær innen maskinlæring ettersom komplekse modeller står i fare for å overtilpasse sine parametre og dermed oppnå lav presisjon out-of-sample. Samtidig er det ikke ønskelig at modellen blir så simpel at underliggende trend ikke fanges opp (Fortmann-Roe, 2012). Fra figur 3.5 illustreres avveiningen mellom modellkompleksitet og overfitting. Eksempelet tar for seg tre forskjellige modelleringer hvor lykke estimeres med velstand som uavhengig variabel. Modellens økte kompleksitet er et resultat av at flere høyere-ordens variabler for velstand inkluderes. De tre scatterplottene øverst illustrerer denne sammenhengen hvor modellens kompleksitet øker fra venstre mot høyre. Med økt kompleksitet fremkommer det tydelig at treffsikkerhet in-sample øker og at modellen i teorien kan oppnå neglisjerbare residualer "in-sample". Det store spredningsplottet viser at det er hensiktsmessig å øke kompleksiteten

¹²Out-of-sample referer til data modellen ikke ser, som et testsett. In-sample er derimot data som benyttes for å tilpasse modellens spesifikasjoner slik at den stemmer godt overens med data out-of-sample.

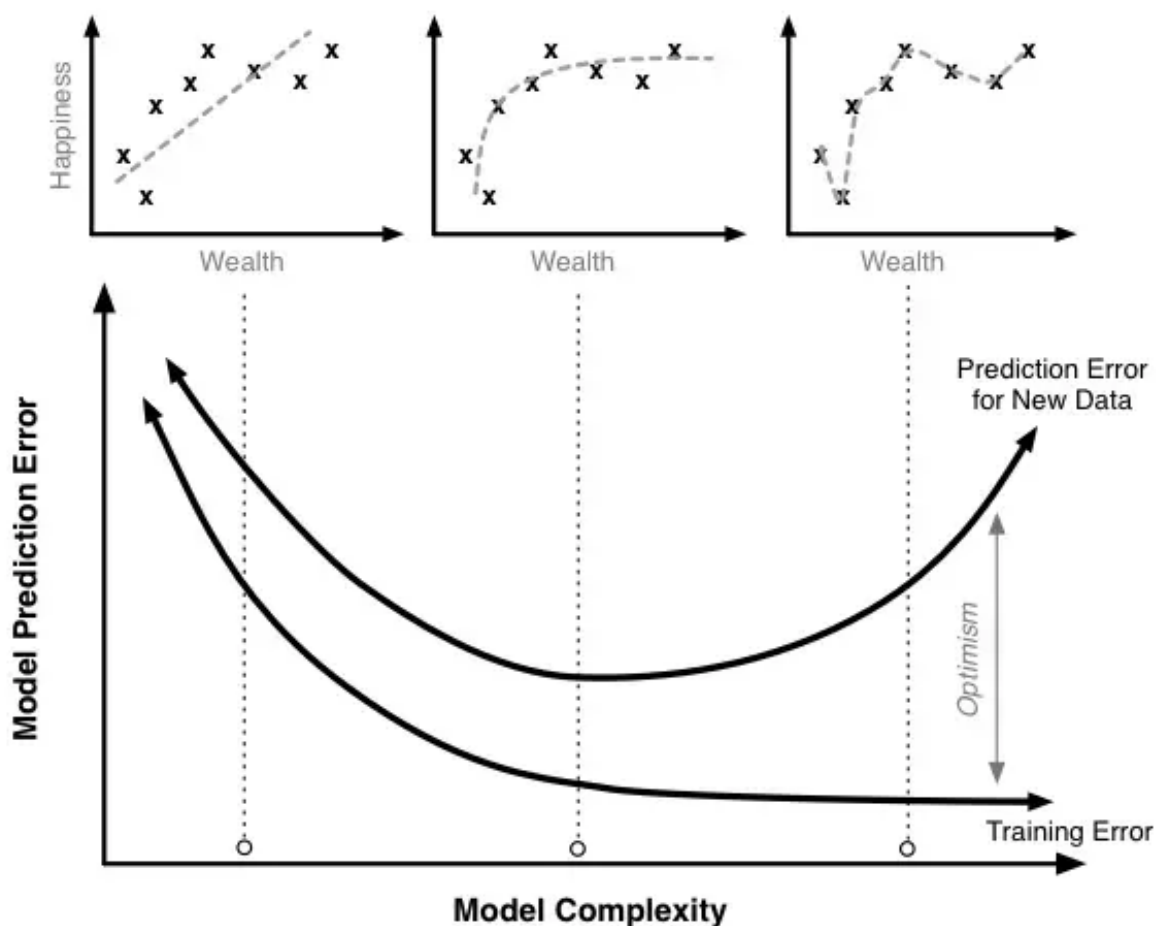
¹³Trening -og testdata vil defineres under neste delkapittel.

¹⁴Støy er irrelevant informasjon eller tilfeldig data som ikke bidrar til å forklare sanne sammenhenger i datasettet. Støy i modellen fører til at prediksjonsmodellen vil treffe dårlig på ny data.

¹⁵Signal omfatter det sanne underliggende mønsteret en ønsker å trene modellen på.

inntil modellkompleksiteten ikke kan økes ytterligere uten å gå på bekostning av dårligere prediksjoner i testsettet. Alle punkter hvor prediksjonsavviket øker med høyere kompleksitet, er hvor modellen overfitter. Alle punkter før dette, er derimot underfitting.

Figur 3.5: Overfitting og underfitting (Fortmann-Roe, 2012)



For å bygge robuste og presise prediksjonsmodeller, står det sentralt å forebygge overfitting. I kapittel 4 vil vi gå nærmere inn på hvordan både Elastic Net og Super Learner er bygget for å redusere problemet.

3.6 Treningssett, valideringssett og testsett

For å evaluere en modells evne til å produsere presise prediksjoner deles det helhetlige datasettet inn i et treningssett og et testsett. Treningssettet deles deretter inn i to ytterligere datasett: Et til trening og et til validering. Treningssettet utgjør den delen av det helhetlige datasettet som brukes for å konfigurere og tune modellens parametre, hvor valideringssettet

benyttes for å evaluere modellens tilpasning. Testsettet består derimot av resterende datagrunnlag, og benyttes for å validere at modellens parametre er korrekt spesifisert out-of-sample. Når modellen konfigureres holdes testsettet skjult, og fungerer derfor som en troverdig indikator på modellens prediksjonsevne. Dersom prediksjonsmodellen produserer prediksjoner som stemmer godt overens med observasjonene fra testsettet, vil modellen oppnå lave prediksjonsavvik.

Den eksakte fordelingen mellom trenings- og testsett vil variere for ulike formål. Det anbefales ofte å ha 20% i testdata og 80% i treningsdata (Hyndman og Athanasopoulos, 2018). Det finnes mange forskjellige metoder for inndeling i treningssett, valideringssett og testsett. I vårt tilfelle er denne inndeling en kompleks oppgave. Dette kommer av at Bybanen fra Byparken til Flesland ble planlagt, bygget og ferdigstilt i ulike perioder. I tillegg ønsker vi å predikere boligpriser for et område hvor Bybanen ikke foreligger, trent og validert på observasjoner hvor Bybanen faktisk eksisterer. Dette medfører at vi ikke har tilgjengelig et testsett fra Åsane hvor Bybaneeffekten innkalkuleres. Inndelingen av datasettet vil drøftes videre i kapittel 4 og 5.

4 Metode

4.1 Maskinlæring og prediksjon

Prediktive analyser omfatter en rekke statistiske metoder som benyttes for å predikere fremtidige utfall, hvor maskinlæring inngår som en klasse av slike metoder. De fleste prediksjonsmodeller tilpasser data etter forhåndsbestemte sammenhenger mellom outputvariabel og inkluderte input-variabler. Eksempelvis vil vanlig lineær regresjon anta lineære sammenhenger mellom de uavhengige variablene og avhengig variabel. Dersom sammenhengene i data ikke er lineære, bør variablene transformeres for å bedre tilpasses den lineære spesifikasjonen. I slike tilfeller er modellen begrenset til å kun bli like god som dets forhåndspesifiserte innstillinger, uavhengig av hvilke sammenhenger som faktisk finnes i data (Jung et al., 2018).

Maskinlæringsmetoder kan derimot oppdage komplekse strukturer i et datagrunnlag uten forhåndprogrammerte spesifikasjoner (Samuel, 1959). Dette gir datamaskinen mulighet til å utarbeide analytiske modeller basert på avanserte algoritmer uten at sluttbruker spesifiserer potensielle sammenhenger. Ved å benytte en slik fremgangsmåte, også omtalt som algoritmisk modellering, kan sluttbruker finne mønstre i data som passer bedre med virkeligheten enn mønstrene som finnes ved å benytte metoder som krever forhåndsbestemte spesifikasjoner. Når formålet er å oppnå prediksjoner med høy presisjon er vi opptatt av å finne en funksjon som beskriver den avhengige variabelen på en presis måte. Forholdet mellom avhengig variabel og uavhengige variabler er derfor mindre relevant, og en praktisk fortolkning av prediksjonsmodellen står således ikke like sentralt for sluttbruker.

Maskinlæring deles inn i en læringsprosess og en algoritme. Læringsprosessens formål er å finne den beste tilpasningen for de uavhengige variablene, hvor algoritmen modellerer sammenhengen mellom de uavhengige variablene og avhengig variabel basert på læringsprosessen (Jung et al., 2018).

4.1.1 Læringsprosessen - *supervised* og *unsupervised learning*

Læringsprosessen deles videre inn i *supervised* og *unsupervised learning*. Supervised learning refererer til problemer hvor modellens output er klart identifisert. Formålet er

dermed å finne den riktige kvantifiseringen av avhengig variabel med bakgrunn i modellens input. Unsupervised learning har derimot ingen definert output. Her er algoritmens mål å finne mønstre basert på en rekke observasjoner og ut fra mønstrene kategorisere de uavhengige variablene (James et al., 2013). Metodene benyttet i denne utredningen baseres på supervised learning hvor vi har en klart definert output - kvadratmeterpris.

4.2 Kryssvalidering

For å optimalisere prediksjonsmodellens presisjon out-of-sample, benytter alle maskinlæringsmodeller ulike metoder for å tune modellens parametre. Formålet med tuningen er å unngå problemer tilknyttet overfitting. Kryssvalidering er en fremgangsmåte som sørger for optimal parameterspesifikasjon for den aktuelle modellen. Det finnes ulike former for kryssvalidering, hvor vi i denne utredningen har valgt å benytte *k-fold kryssvalidering*. Ved *k-fold kryssvalidering* deles treningssettet inn i k like store, ikke-overlappende grupper. Alle gruppene (k), med unntak av én, benyttes først som treningssett. Gruppen som ikke inngår i treningssettet legges bort og betegnes som valideringssett. For treningssettet benyttes ulike kombinasjoner av relevante parametre for å tune og validere modellen. Denne prosessen estimeres k ganger, hvor gruppen som utgjør valideringssettet endres for hver modellering. Valget av parameterverdier baseres videre på hvilke verdier som gir lavest prediksjonsavvik mot valideringssettet gjennom *k-fold kryssvalidering*. Modellen som oppnår best tilpasning in-sample benyttes videre til å predikere observasjoner out-of-sample. Observasjonene out-of-sample evalueres deretter mot testsettet for å vurderer modellens presisjon.

Det finnes mange alternative metoder for kryssvalidering. *Leave-one-out-kryssvalidering* er et eksempel på en av mange metoder som kunne vært benyttet i denne utredningen. I denne metoden tilsvarer antall grupper, k , antallet observasjoner i datasettet, i . Etersom metoden krever stor prosesseringskraft for omfattende datasett og komplekse prediksjonsmodeller, anser vi *leave-one-out-kryssvalidering* som mindre hensiktsmessig i denne utredningen. Dette på tross av metodens potensial til å gi noe økt presisjon.

4.3 Elastic Net

Elastic Net (Zou og Hastie, 2005) er en maskinlæringsalgoritme som kombinerer Ridge og Lasso regresjon, som henholdsvis beskrives i neste delkapittel, med formål å oppnå bedre prediksjoner enn hva som er mulig ved å benytte metodene hver for seg. Både Lasso og Ridge regresjon er lineære metoder, som i likhet med ordinær lineær regresjon (OLS) bygger på å estimere koeffisientene ($\hat{\beta}_p$) som minimerer den kvadrerte avstanden mellom prediksjonene (\hat{y}_n) og de observerte verdiene (y_n). Avstanden mellom \hat{y}_n og y_n omtales som residualer.

Sammenhengen mellom avhengig variabel og de uavhengige variablene kan presenteres gjennom en ligning for multipl regressjon:

$$y_n = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \mathcal{E}_n \quad (4.1)$$

Her representerer β_p den sanne, men uobserverbare påvirkningen variabelen X_p har på avhengig variabel. Parameteren α representerer verdien av avhengig variabel når alle uavhengige variabler er lik 0. \mathcal{E}_n er definert som feilleddet for hver prediksjon n . \hat{y}_n er videre den predikerte verdien for de sanne observasjonene, y_n , og $\hat{\beta}_p$ den beste estimatoren til den sanne parameteren β_p .

En utfordring med OLS i et prediksjonsperspektiv er å velge en modell gjør en god avveining mellom modellkompleksitet og økt varians, som diskutert i kapittel 3.5. Ved å inkludere for mange, eller korrelerte forklaringsvariabler¹⁶ i en ordinær lineær regresjon, vil modellkompleksiteten øke og overfitting kan bli et problem. Elastic Net kombinerer dimensjonsreduksjon fra Ridge og variabelseleksjon fra Lasso for å utarbeide en prediksjonsmodell som tar høyde for problematikken tilknyttet korrelerte variabler, overfitting og underfitting. Disse karakteristikene utdypes under.

For å få en økt forståelse for hvordan Elastic Net er sammensatt, er det hensiktsmessig å først presentere Lasso og Ridge regresjon hver for seg.

¹⁶Situasjonen som oppstår når en eller flere forklaringsvariabler er lineært korrelert omtales som multikorrelasjon.

4.3.1 Ridge Regresjon

Ridge regresjon, introdusert av Hoerl og Kennard (1970), har i likhet med OLS som formål å minimere de kvadrerte residualene. Ulikt ordinær lineær regresjon åpner Ridge for å regularisere koeffisientestimatene ved å introdusere en straffeterm i minimeringsproblemet. Regularisering innebærer å krympe koeffisientestimatene tilhørende korrelerte variabler. Dette medfører at påvirkningen korrelerte variabler har på avhengig variabel minker ved at variabler som forklarer den samme variansen tildeles lavere koeffisientestimer enn de ville hatt ved OLS. Ridge regresjon vil ikke krympe koeffisientestimatene til å bli eksakt lik null. Dette medfører at ingen variabler vil bli helt eliminert fra modellen ved bruk av denne metoden.

Optimeringsproblemet er som følger:

$$\hat{\beta} = \arg \min_{\hat{\beta}_j} \left[\underbrace{\sum_{i=1}^n (Y - X\hat{\beta})^2}_{\text{RSS}} + \lambda \underbrace{\sum_{j=1}^p (\hat{\beta}_j)^2}_{\text{Ridge penalty}} \right] \quad (4.2)$$

Y representerer verdien av avhengig variabel og X verdien av de uavhengige variablene. Videre er n antall observasjoner og p antall forklaringsvariabler i modellen. Akkurat som minste OLS søker Ridge regresjon å minimere RSS¹⁷, definert i det første leddet av ligningen. Størrelsen på det andre leddet, straffetermen, også omtalt som “shrinkage penalty”, styres av hvor stor vekten kompleksitetsparameteren i Ridge, λ , skal være. Dette avgjøres gjennom kryssvalidering. Eksempelvis vil $\lambda=0$ medføre at Ridge regresjon sammenfaller med ordinær lineær regresjon ettersom straffeparameteren forsvinner. Dersom $\lambda \rightarrow \infty$ vil derimot de estimerte forklaringsvariablene gå mot null. Ulikt OLS, som kun generer ett sett av koeffisientestimer, vil Ridge regresjon generere et sett av koeffisientestimer for hver verdi av λ . Derfor er valget av en passende λ viktig. I praksis fører straffeparameteren til at korrelerte variabler får redusert sine estimerte koeffisienter i den endelige modellen, noe som fører til lavere varians.

¹⁷RSS er en forkortelse for residual sum of squares, oversatt til summen av de kvadrerte residualene på norsk.

4.3.2 Lasso Regresjon

Lasso regresjon ble for første gang introdusert av Tibshirani (1996) som en utvidelse av OLS og et alternativ til Ridge regresjon. Metoden er nesten identisk med Ridge med unntak av en viktig forskjell: Hvor Ridge kun har muligheten til å *krympe* koeffisientestimatene for korrelerte variabler, kan Lasso eliminere koeffisientene helt.

Optimeringsproblemet er som følger:

$$\hat{\beta} = \arg \min_{\beta_j} \left[\underbrace{\sum_{i=1}^n (Y - X\hat{\beta})^2}_{\text{RSS}} + \lambda \underbrace{\sum_{j=1}^p |\hat{\beta}_j|}_{\text{Lasso penalty}} \right] \quad (4.3)$$

Den første delen av ligningen representerer minimeringsproblemet for OLS, hvor den andre delen er Lasso sin straffeterm. I motsetning til Ridge regresjon kvadreres ikke $\hat{\beta}_j$ i Lasso regresjon. Denne forskjellen medfører at Lasso regresjon åpner for at en tilstrekkelig stor λ vil føre til at koeffisientestimatene blir lik null og dermed forsvinne helt fra den endelige modellen. Dette omtales som variabelseleksjon. Variabelseleksjon har sine fordeler og ulemper. Fordelen er at modellen blir enklere å tolke, samt at variansen vil reduseres. Ulempen er at modellen ikke hensyntar hvilken eller hvilke av de korrelerte variablene som blir ekskludert fra modellen. Dette kan medføre at viktige forklaringsvariabler blir utelatt. For eksempel kan avstand til sentrum ekskluderes fra modellen på grunn av sterk negativ korrelasjon med bygningsalder. Dette kan påvirke modellens prediksjoner ettersom avstand til sentrum trolig har en egen effekt på boligpris som ikke kan modelleres ved hjelp av bygningsalder alene.

Den ene metoden vil ikke dominere den andre uavhengig av prediksjonsformålet. På et overordnet nivå vil Lasso ofte predikere bedre dersom mange av prediksjonsvariablene har små koeffisientestimater, hvor resterende koeffisienter er av betydelig størrelse. Ridge vil på den andre siden kunne predikere bedre dersom responsvariabelen er en funksjon av mange uavhengige variabler, hvor tilhørende koeffisienter er av relativt lik størrelse. (Oleszak, 2019)

4.3.3 Elastic Net

Elastic Net ble fremlagt av Zou og Hastie (2005) som en sammenslåing av Ridge og Lasso regresjon for å muliggjøre både dimensjonsreduksjon og variabelseleksjon. Elastic Net utfører simultant automatisk krymping og variabelutvelgelse, og muliggjør derfor å beholde grupper av korrelerte variabler. Simuleringsstudier viser at Elastic Net ofte oppnår høyere presisjon enn Lasso regresjon (Zou og Hastie, 2005).

Optimeringsproblemet er som følger:

$$\hat{\beta} = \arg \min_{\beta_j} \left[\underbrace{\sum_{i=1}^n (Y - X\hat{\beta})^2}_{\text{RSS}} + \lambda \sum_{j=1}^p \left[\underbrace{(1 - \alpha) (\hat{\beta}_j)^2}_{\text{Ridge}} + \underbrace{\alpha |\hat{\beta}_j|}_{\text{Lasso}} \right] \right] \quad (4.4)$$

Første ledd av ligningen over beskriver minimeringsproblemet for OLS. Det andre leddet er straffetermen for Ridge og Lasso regresjon vektet med α . α er mellom 0 og 1, hvor α lik 0 tilsvarer Ridge regresjon og α lik 1 tilsvarer Lasso regresjon. λ avgjør størrelsen på det totale straffeledet, hvor $\lambda = 0$ impliserer ordinær lineær regresjon. På denne måten styrer α den relative vektning av Ridge og Lasso, som avgjør hvor sentralt dimensjonsreduksjon og variabelseleksjon skal være for å redusere variansen i den valgte modellen.

Elastic Net er spesielt godt egnet til å håndtere situasjoner hvor det er parvis korrelasjon mellom parametre. Som nevnt, vil vi ved å kun benytte Lasso regresjon stå i fare for å velge bort sentrale forklaringsvariabler i de tilfeller hvor det foreligger korrelasjon mellom uavhengig variabel. Ved å inkludere Ridge regresjon, vil Elastic Net kunne beholde viktige variabler selv om det foreligger korrelasjon mellom dem. At Elastic Net evner å beholde korrelerte variabler kan være en viktig egenskap i vårt tilfelle. Dette fordi variabler som er sterkt korrelert enkeltvis kan være viktige for å predikere boligpriser.

Vi har benyttet R-pakken¹⁸ *glmnet* for å bygge prediksjonsmodeller med Elastic Net (Friedman et al., 2009). Pakken har en egen funksjon for å gjennomføre k -fold kryssvalidering kalt *cv.glmnet*. I vår analyse settes $k=10$ både som et resultat av at vi ikke ønsker for høy kostnad ved hver beregning, i tillegg til at $k=10$ gir en god avveining mellom økt presisjon og økt varians (James et al., 2013). *Cv.glmnet* finner videre optimal

¹⁸R er et programmeringsspråk som i hovedsak benyttes til statistisk analyse. Pakkene som inngår i R består av samlinger av funksjoner som enkelt kan lastes ned (R core team, 2020).

verdi for λ ved å minimere residualenes størrelse for prediksjonene og valideringssettet. α ble også valgt på bakgrunn av å minimere RMSE in-sample.

4.4 Super Learner

Super Learner, introdusert av Van der Laan et al. (2007), er en samlingsmetode som søker å oppnå presise prediksjoner gjennom en optimal vektning av ulike prediksjonsmetoder. Kryssvalidering benyttes for å avgjøre vektningen av de inkluderte metodene, hvor den optimale modellkombinasjonen velges på bakgrunn av den metodesammensetningen som minimerer avvikene for prediksjonene in-sample (Van der Laan et al., 2007). Fordelen med å benytte Super Learner i analysen, er at metoden, ved å inkludere et bredt modellspekter med ulike karakteristikk, muliggjør en god tilpasning til det aktuelle prediksjonsproblemet.

Metodene som potensielt bør være med i prediksjonsmodellen velges manuelt av sluttbruker. Etter at metodene er valgt ut, vil utallige kombinasjoner av de inkluderte metodene med tilhørende tuning-parametre vurderes basert på oppnådd prediksjonspresisjon in-sample. Dette betyr at dersom vi eksempelvis inkluderer ordinær lineær regresjon, Nevrale Nettverk og Elastic Net, vil Super Learner i prinsippet velge den vektningen av metoder som gir de mest presise prediksjonene av boligprisene i valideringssettet. Ulempen med Super Learner er imidlertid at metoden krever stor datakapasitet, spesielt når komplekse ikke-lineære metoder inkluderes.

R-pakken *SuperLearner* (Polley og van der Laan, 2017) benyttes i denne sammenheng for å utarbeide prediksjonene. De inkluderte metodene er presentert i følgende tabell:

Tabell 4.1: Metodene som velges manuelt til å inkluderes i Super Learner

| | |
|--------------|--|
| SL.Mean | Sample Mean |
| SL.lm | Ordinary Least Square |
| SL.glmnet | Elastic Net |
| SL.nnet | Neural Network |
| SL.kernelKnn | Kernel k-Nearest Neighbour |
| SL.polymars | Multivariate Adaptive Polynomial Spline Regression |

De inkluderte prediksjonsmetodene inneholder et bredt spekter av ulike karakteristikk, hvor noen er enkle og lineære, mens andre er mer komplekse og håndterer ikke-lineære

sammenhenger på en bedre måte. Ved å inkludere et bredt spekter av metoder, håper vi å oppnå økt presisjon.

4.5 RMSE og MAE - vurdering av presisjon

For å kunne vurdere modellenes treffsikkerhet, er det viktig at prediksjonenes presisjon evalueres ved bruk av pålitelige beregningsmetoder. I denne sammenheng ønsker vi å fokusere på to kvantitative måltall for å analysere prestasjonen til de ulike modellene.

Root Mean Squared Error (RMSE) er et standard mål som benyttes for å måle residualenes størrelse når prediksjonene out-of-sample sammenlignes med de observerte verdiene fra test-settet (Moody, 2019). Fra formelen under fremkommer det at RMSE uttrykkes som gjennomsnittlig kvadrert verdi av residualene for hele test-perioden.

$$RMSE = \sqrt{\left(\sum_{i=1}^n (\hat{y}_n - y_n)^2\right) / n} \quad (4.5)$$

Ettersom RMSE kvadrerer residualene før gjennomsnittet beregnes, vil metoden straffe store avvik relativt hardt sammenlignet med små avvik. Videre er det verdt å bemerke at residualene er absolutte. Dette resulterer i at residualer av positiv og negativ størrelse tildeles lik vektning. Prediksjonsmodellen med lavest RMSE betegnes som den mest presise. Ettersom mange eiendomsinvestorer baserer sine investeringsstrategier på en høy belåningsgrad (Steinert og Crowe, 2001) er det sentralt at modellen straffer store avvik. Før eiendomsinvestorer får tilbakebetalt sin egenkapital, har långivere rett på sine innskutte midler. Dersom et prosjekt ender med tap, skal alltid långivere få tilbakebetalt sin andel før investors tap dekkes. Dette betyr at investors kapitalkrav er residualt. Dersom investor har en høy belåningsgrad vil risikoen tilknyttet investeringens egenkapital øke ettersom prosjektets tap ikke trenger å være stort før det går på bekostning av investors egenkapital. Dette medfører at investeringer basert på prediksjoner med stor feilmargin kan føre til at investor står i fare for å tape store andeler av sin egenkapital og i verste fall ende som insolvent.

Alternativt kan Mean Absolute Error (MAE) brukes for å måle prediksjonsevne. I denne utredningen brukes målet som et sekundært evalueringskriterium. Fra formelen under

fremkommer det at MAE tallfester gjennomsnittet av absoluttverdiene til residualene.

$$MAE = \left(\sum_{i=1}^n |\hat{y}_n - y_n| \right) / n \quad (4.6)$$

Metoden måler feilprediksjoner i absolutte termer, akkurat som RMSE, men straffer derimot ikke større avvik hardere enn små avvik. MAE er lettere å tolke og anslår i denne analysen absolutt gjennomsnittlig avvik mellom faktisk kvadratmeterpris og predikert kvadratmeterpris.

Oppsummert er MAE lettere å tolke, mens RMSE vil være fordelaktig dersom det er sentralt å straffe større feilprediksjoner. Dersom det forekommer tilfeller hvor to modeller har relativt lik MAE, men ulik RMSE er dette et tegn på at den ene modellen har færre, men større avvik sammenlignet med modellen med lavere RMSE. Det kan således være interessant å inkludere begge måltallene for å få et bedre sammenligningsgrunnlag.

5 Analyse

5.1 Bybanens byggetrinn

Før vi går i dybden bak prediksjonsrammeverkets oppbygging, ønsker vi først å presentere Bybanens foreløpige utbyggingstrinn.

Tabell 5.1: Oversikt over Bybanens byggetrinn

| Byggetrinn nummer | Planlegging | Utbygging | Etter ferdigstillelse |
|----------------------|---------------------|---------------------|-----------------------|
| 1 (Byparken-Nesttun) | Okt. 2000-Jan 2008 | Jan. 2008-Juni 2010 | Juni 2010-Des. 2019 |
| 2 (Nesttun-Lagunen) | Mars 2008-Jan 2011 | Jan. 2011-Juni 2013 | Jan. 2013-Des. 2019 |
| 3 (Lagunen-Flesland) | Okt. 2011-Juli 2013 | Juli 2013-Aug. 2017 | Aug. 2016-Des. 2019 |

Første kolonne representerer Bybanens første, andre og tredje byggetrinn. Disse strekker seg fra henholdsvis Byparken til Nesttun, Nesttun til Lagunen og Lagunen til Flesland. Planleggingsfasen utgjør perioden fra hvert respektive byggetrinn vedtas i bystyret til utbyggingen starter. Utbyggingsfasen er definert som perioden fra utbyggingen starter til bybanetraseens ferdigstillelse. “Etter ferdigstillelse” er perioden fra ferdigstillelse til 31. desember 2019.

Fra Tabell 5.1 fremkommer det at Bybanens foreløpige byggetrinn har blitt vedtatt og planlagt, bygget og ferdigstilt i ulike perioder¹⁹. For eksempel ble Bybanen fra Byparken til Nesttun planlagt fra 2000 til 2008, bygget ut mellom 2008 og 2010 og ferdigstilt i 2010.

For tydelighetens skyld ønsker vi å presisere noen formuleringer som benyttes i kommende analyse. Vi benytter eksempelvis formuleringer som: “Byparken-Nesttun etter Bybanens inntreden og seks år frem i tid benyttes som treningssett, og Nesttun-Lagunen fra og med Bybanens ferdigstillelse og seks år frem i tid som testsett”. Dette antyder at vi bruker verdiene på boligtransaksjonene som har forekommet innen 3000 meter fra nærmeste bybanestopp mellom Byparken og Nesttun for perioden 2010-2016 som treningssett, og verdiene på boligtransaksjonene som har forekommet innen 3000 meter fra nærmeste bybanestopp mellom Nesttun og Lagunen for perioden 2013-2019 som testsett.

¹⁹Opprinnelig ble Bybanens trasé mellom Lagunen og Flesland bygget i to omganger, hvor det første byggetrinn gikk fra Lagunen til Birkelandsskiftet. Dette byggetrinn ble ferdig i underkant av et år før traséen ble videreført til Flesland. Det er likevel vanlig å anse dette som et sammenhengende byggetrinn (Bergen Kommune, 2019a), noe vi også gjør i denne utredningen.

5.2 Oppbygging av et helhetlig prediksjonsrammeverk

Det er mange momenter som gjør det utfordrende å predikere boligprisene mellom Byparken og Åsane for perioden 2010-2019, i et tenkt scenario hvor Bybanen til Åsane hadde blitt ferdigstilt i 2010²⁰. Som nevnt innledningsvis er et av hovedproblemene at eksisterende bybane ikke går gjennom det geografisk område vi ønsker å predikere for. På grunn av mangel på observasjoner mellom Byparken og Åsane som er påvirket av nærhet til Bybanen, ønsker vi å utarbeide et prediksjonsrammeverk, som gjennom fire steg, innhenter nok informasjon til å ta kvalifiserte avgjørelser om hvilken modell som trolig vil predikere boligprisene i Åsane med Bybanen mest presist. Innsikten fra prediksjonsrammeverket vil deretter benyttes for å predikere boligprisene mellom Byparken og Åsane for perioden 2010-2019, hvor vi forutsetter at Bybanen ble ferdigstilt i 2010. Før vi forklarer rammeverket i detalj, ønsker vi kort å beskrive hovedlinjene i de fire stegene.

Første steg er å teste om variablene *bygningsalder*, *salgstid* og *tomteareal* bør transformeres i henhold til beskrivelsen fra kapittel 3.4.1 når vi predikerer boligprisene tilknyttet Åsane-traseen. I mangel på observasjoner fra Åsane-området som er påvirket av bybaneeffekten, er neste steg å avgjøre hva som vil være det beste trenings- og valideringssettet for å trene og tune modellens parametre til å predikere i Åsane for perioden etter Bybanen kommer der. Her er målet å finne det mest presise datagrunnlaget til å trene modellene som skal predikere boligprisene i Åsane med Bybanen. Med et presist datagrunnlag, menes data som i størst mulig grad gir samsvar mellom treningsdata og de reelle forholdene i Åsane. Deretter vil vi analysere om vi oppnår høyere presisjon når vi predikerer boligprisene i et område hvor Bybanen foreligger, ved å trene og validere modellene på data fra et annet geografisk område som også er påvirket av nærhet til Bybanen, sammenlignet med presisjonen vi oppnår om treningsdata er upåvirket av Bybanen. Dette er et viktig moment ettersom vi må vite om vi kan trene en modell basert på observasjoner fra et område hvor Bybanen foreligger (Byparken-Lagunen), til å predikere boligpriser i et annet område, hvor Bybanen kommer på et senere tidspunkt (Byparken-Åsane). Gjennom det siste steget i prediksjonsrammeverket vil vi analysere hvilken innvirkning variabelen *avstand til nærmeste bybanestopp* har på presisjonen vi oppnår når vi predikerer boligpriser innen

²⁰Som nevnt innledningsvis vil vi benytte referanser som “fra og med Bybanens ferdigstillelse til Åsane” eller “med Bybanen” for å hentyde til disse prediksjonene

3000 meter fra Bybanen. Prediksjonsrammeverket oppsummeres i Figur 5.1.

Figur 5.1: Prediksjonsrammeverkets fire steg



5.2.1 Variabeltransformasjon

Ettersom vi ikke vet hvordan boligprisene vil utvikle seg rundt Åsane-traseen i tiden etter Bybanens inntreden, kan vi heller ikke fastslå med sikkerhet hvilken prediksjonsmodell som vil oppnå høyest presisjon. Det første vi ønsker å vurdere er derfor om det foreligger indikasjoner på at variablene *bygningssalder*, *salgstid* og *tomteareal* bør transformeres i henhold til beskrivelsen fra kapittel 3.4.1 for å oppnå økt presisjon når vi predikerer boligprisene rundt Åsane-området.

Vi tester dette ved å predikere boligprisene mellom Nesttun og Lagunen fra Bybanens ferdigstillelse og seks år frem i tid ved å benytte fire ulike modeller. Elastic Net, henholdsvis med og uten transformerte variabler utgjør to av modellene. Super Learner²¹, med og uten transformerte variabler utgjør de to resterende modellene. Alle modellene benytter observasjoner mellom Byparken og Nesttun som trenings- og valideringssett for tiden etter Bybanens ferdigstillelse for det respektive område og seks år frem i tid. Dersom Elastic Net med transformerte variabler predikerer mer presist enn Elastic Net uten transformerte variabler, vil modellen uten variabeltransformasjon utelukkes fra videre analyse, og motsatt. Det samme gjelder for Super Learner. Fremgangsmåten bygger på at den modellen, for henholdsvis Elastic Net og Super Learner, som oppnår høyest presisjon når vi tester på et område hvor Bybanen foreligger i dag, indikerer høyest presisjon når vi skal predikere boligpriser rundt Åsane-traseen etter Bybanens inntreden der.

²¹Som tidligere nevnt er Elastic Net en lineær metode, og ved å benytte transformerte variabler kan vi potensielt oppnå bedre tilpasning på grunn av ikke-lineære sammenhenger i datagrunnlaget. Ettersom vi har inkludert lineære metoder i Super Learner, vil vi også teste variabeltransformasjon for denne metoden.

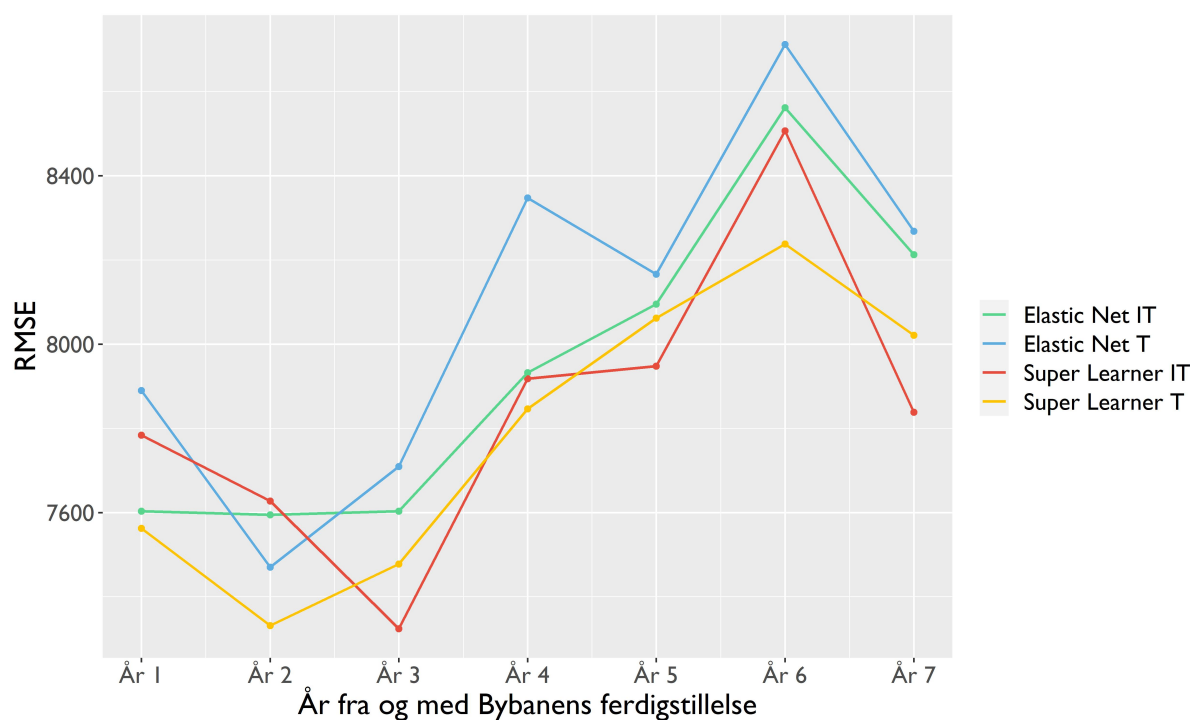
Det er ikke tilfeldig at Byparken-Nesttun benyttes som treningssett og Nesttun-Lagunen som testsett. For det første er boligprisene både i trenings- og testsettet påvirket av Bybanen. For det andre er det sentralt å teste modellenes treffsikkerhet når testsett består av observasjoner tilhørende et geografisk område som er ulikt området som inkluderes i treningssettet. Disse momentene er viktige ettersom utredningens formål er å predikere boligpriser for et området hvor Bybanen kommer (Byparken-Åsane), basert på et treningssett som består av observasjoner fra et annet området hvor Bybanen foreligger (Byparken-Flesland). Tabell 5.2 viser gjennomsnittlig RMSE og MAE for de fire ulike modellene.

Tabell 5.2: Gjennomsnittlig RMSE og MAE for de ulike metodene

| | Elastic Net IT | Elastic Net T | Super Learner IT | Super Learner T |
|------|-----------------------|----------------------|-------------------------|------------------------|
| RMSE | 8121.4 | 8464.4 | 8021.9 | 8032.9 |
| MAE | 6110.1 | 6321.0 | 5959.6 | 5919.4 |

Elastic Net IT indikerer Elastic Net uten transformerte variabler (ikke-transformert) og Elastic Net T Elastic Net med transformerte variabler (transformert). Det samme gjelder for Super Learner IT og Super Learner T.

Basert på begge de inkluderte evalueringskriteriene, hvor målet er lavest mulig verdi, oppnår Elastic Net uten transformerte variabler en RMSE og en MAE på henholdsvis 4.0% og 3.3% under Elastic Net med transformerte variabler. Dette på tross av at den grafiske fremstillingen av de transformerte variablene (Figur A2.2-A2.8) antydte at variabeltransformasjon ville føre til mer lineære sammenhenger i data. Videre oppnår Super Learner IT en RMSE som er 0.1% lavere enn Super Learner T, men derimot en MAE som er 0.7% høyere. Forskjellen i presisjonen som oppnås med Super Learner IT og Super Learner T er lite signifikant. I denne sammenheng virker det derfor som at variabeltransformasjon spiller en marginal rolle når Super Learner benyttes.

Figur 5.2: Visualisering av årlig RMSE for de ulike modellene

Ettersom Bybanen mellom Nesttun og Lagunen, som i denne sammenheng benyttes som testsett, ble ferdigstilt i 2013, tilsvarer År 1 her 2013, År 2 2014 og så videre. Observasjonene som utgjør treningssettet (Byparken-Nesttun) strekker seg derimot fra 2010-2016. År 1 vil således tilsvare 2010 og År 2 2011.

Ved å illustrere årlig RMSE grafisk får vi en oversikt over modellenes presisjon for hvert år etter Bybanens ferdigstillelse. Grafen bekrefter at både Super Learner IT og Super Learner T gjennomgående oppnår høyere presisjon enn Elastic Net IT og Elastic Net T. Det eneste unntaket er Super Learner T sin presisjon i År 1 og År 2 fra og med Bybanens inntreden. Det faktum at Super Learner ikke bare presterer best i gjennomsnitt, men relativt godt for hver inkluderte periode er fordelaktig. Dersom metoden eksempelvis hadde predikert særdeles bra for tre av periodene, men dårlig for resterende perioder, ville dette medført høyere usikkerhet tilknyttet boliginvesteringer hvor investor ikke har et ønske om å besitte eierskap til investeringen gjennom hele perioden.

Med grunnlag i diskusjonen over, ønsker vi å benytte både Super Learner og Elastic Net uten transformerte variabler i videre modelloppbygging. På tross av at Elastic Net IT presterer dårligere enn Super Learner IT, beholdes begge modellene i videre analyse. Dette fordi vi ønsker å sammenligne modellpresisjon også i andre sammenhenger for å

få et bedre grunnlag for å avgjøre hvilken metode som trolig vil prestere best når vi predikerer boligpriser i Åsane etter Bybanens ferdigstillelse. Videre i analysen benytter vi betegnelsene “Elastic Net” og “Super Learner” for å beskrive henholdsvis Elastic Net og Super Learner uten transformerte variabler.

En svakhet med fremgangsmåten vi bruker for å avgjøre om variablene bør transformeres eller ikke, er at treningssettet som benyttes ligger på samme side av Bergen sentrum som testsettet. I tillegg ligger mange av observasjonene i treningssettet i samme bydel som testsettet, nemlig Fana. Strekningen mellom Byparken og Åsane ligger derimot på den andre siden av Bergen sentrum, og er ikke en del av Fana bydel. Dette medfører at modellene som benyttes for å avgjøre om variablene burde transformeres eller ikke, trenes og testes på observasjoner som trolig bærer preg av å være relativt like. Dette er sannsynligvis ikke tilfellet for treningsdata (Byparken-Lagunen)²² og de faktiske observasjonene som predikeres mellom Byparken og Åsane i den endelige modellen. Vi vil således ikke vite med sikkerhet om modeller som bygger på ikke-transformerte variabler vil føre til høyest presisjon når vi predikerer boligprisene i Åsane med Bybanen. Det foreligger likevel indikasjoner på at variablene ikke bør transformeres når vi predikerer boligpriser i Åsane med Bybanen.

5.2.2 Valg av trenings- og valideringssett

Som kjent, er en utfordring i analysen at det ikke foreligger et datasett fra Åsane-område som er påvirket av Bybaneeffekten²³. For å predikere boligprisene i Åsane fra og med Bybanens ferdigstillelse må vi derfor benytte et utvalg fra boligtransaksjonene mellom Byparken og Flesland for å trene og validere modellene som benyttes. I denne sammenheng står det sentralt å finne et trenings- og valideringssett som gir grunnlag for høy presisjon når vi predikerer boligprisene i Åsane med Bybanen.

Det treningssettet fra det området som fører til at modellene oppnår høyest presisjon når vi predikerer boligprisene i Åsane før Bybanen kommer, indikerer at et treningsgrunnlag fra det samme området også vil føre til de mest presise prediksjonene i tiden etter Bybanens ferdigstillelse. Dersom vi eksempelvis oppnår høyere presisjon ved å benytte et treningssett mellom Byparken og Nesttun, enn ved å benytte et treningssett mellom

²²Bakgrunnen for at Byparken-Lagunen benyttes som trenings- og testsettet utdypes i neste delkapittel

²³Bybaneeffekten er definert i kapittel 1

Byparken-Lagunen når boligprisene mellom Byparken og Åsane i perioden før Bybanen predikeres, indikerer dette at observasjoner mellom Byparken og Nesttun også vil føre til mest presise prediksjoner i tiden etter Bybanen kommer. I så tilfelle ville vi brukt Byparken-Nesttun som treningsgrunnlag for å predikere Bybanens påvirkning på boligprisene i Åsane mellom for perioden 2010-2019.

Ettersom boligprisene i Åsane-området er upåvirket av bybaneeffekten, er det fornuftig å teste denne sammenhengen med et trenings- og valideringssett som også er upåvirket av bybaneeffekten. I første omgang vil vi derfor identifisere et utvalg av observasjonene fra Byparken til Flesland som kan antas å være tilnærmet upåvirket av bybaneeffekten. Vi ser fra Tabell 5.1 at første byggetrinn ble vedtatt i år 2000. Utbyggingen av tilhørende trasé ble videre påbegynt i 2008. Ifølge teorien om effisiente markeder skal prisen av et aktivum, som en aksje eller en bolig, reflektere all relevant informasjon som påvirker prisingen av aktivumet (Malkiel og Fama, 1970). Dette medfører at boligprisene vil stige allerede idet øyeblikket nyheten om Bybanen blir publisert. Dette momentet vil diskuteres ytterligere i kapittel 6.4, men vi antar inntil videre at Bybanens påvirkning på boligprisene mellom Byparken og Nesttun er marginal i perioden 2001-2007²⁴. Vi velger derfor å bruke treningsdata fra perioden 2001-2007.

Spørsmålet blir videre om treningsdata skal bestå av alle observasjonene fra Byparken til Flesland, eller kun av et utvalg av dem. Vi deler derfor datasettet fra Byparken til Flesland inn i tre grupper, alle for perioden 2001-2007. De tre gruppene består henholdsvis av observasjoner fra områdene mellom Byparken-Nesttun, Byparken-Lagunen og Byparken-Flesland. Alle de tre gruppene blir brukt som trenings- og valideringssett for å predikere boligprisene mellom Byparken og Åsane for perioden før Bybanens første byggetrinn ble påbegynt (2001-2007). Som diskutert over, bygger intuisjonen vår på at det treningsgrunnlaget som fører til høyest presisjon i tiden før Bybanens påvirkning på boligprisene var signifikant, indikerer at et treningssett fra det samme område følgelig vil medføre høyest presisjon også i tiden etter at Bybanen til Åsane kommer.

I vår endelige prediksjonsmodell ønsker vi å predikere priser på boliger som strekker seg over to bydeler - Bergenhus og Åsane - med tilhørende varierende avstand fra sentrum. Vi

²⁴2001-2007 utgjør perioden fra og med ett år etter at første byggetrinn ble vedtatt, til og med året før utbyggingen startet.

benytter derfor et randomisert testsett²⁵ mellom Byparken og Åsane når vi identifiserer hvilket datasett som skal utgjøre det beste treningssettet. Tabell 5.3 tallfester presisjonen til Elastic Net og Super Learner for prediksjonene beskrevet over.

Tabell 5.3: RMSE og MAE for predikerte boligpriser mellom Byparken og Åsane for perioden 2001-2007

| Elastic Net | | | |
|----------------------|------------------|------------------|-------------------|
| | Byparken-Nesttun | Byparken-Lagunen | Byparken-Flesland |
| RMSE | 9881.7 | 9772.8 | 9807.0 |
| MAE | 7230.6 | 6977.7 | 7010.9 |
| Super Learner | | | |
| | Byparken-Nesttun | Byparken-Lagunen | Byparken-Flesland |
| RMSE | 10122.2 | 9400.1 | 11052.3 |
| MAE | 7324.3 | 6818.0 | 7479.5 |

Treningssettet som benyttes er henholdsvis Byparken-Nesttun, Byparken-Lagunen og Byparken-Flesland for perioden mellom 2001 og 2007.

RMSE lik 9772.8 representerer Elastic Net sin presisjon når observasjonene mellom Byparken og Lagunen for perioden 2001-2007 benyttes som trenings- og valideringssett for å predikere boligprisene mellom Byparken og Åsane for samme periode. RMSE lik 9807.0 angir derimot Elastic Net sin presisjon når observasjonene mellom Byparken og Flesland brukes for å trene og validere modellene som predikerer boligprisene rundt Åsane-området. Tilsvarende gjelder for Super Learner.

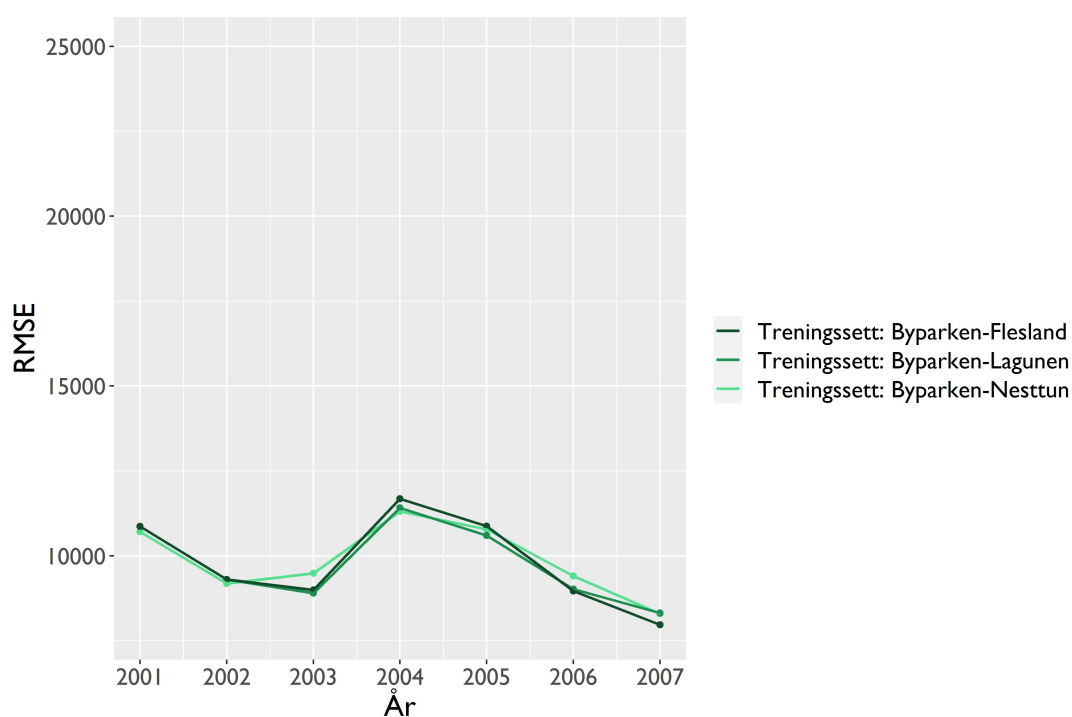
Resultatene som fremgår fra tabellen indikerer at presisjonen gjennomgående blir høyest når Byparken-Lagunen benyttes som treningssett for å predikerer boligprisene i Åsane. I relative termer oppnår Elastic Net marginale forbedringer i presisjon når Byparken-Lagunen benyttes som treningsgrunnlag i stedet for Byparken-Flesland, hvor RMSE og MAE er henholdsvis 0.3% og 0.5% lavere. Presisjonsforbedringen for Elastic Net er derimot større når vi sammenligner presisjonen som oppnås ved å bruke Byparken-Lagunen som treningssett, sammenlignet med Byparken-Nesttun. Her er RMSE og MAE henholdsvis 1.1% og 3.6% lavere. Når Super Learner trenes og valideres på observasjonene rundt Byparken-Lagunen er RMSE 17.6% lavere og MAE 9.7% lavere enn presisjonen som oppnås når Byparken-Flesland utgjør treningsgrunnlaget. Sammenligner vi presisjonen Super Learner oppnår når Byparken-Lagunen benyttes som trenings- og valideringssett i

²⁵Antall observasjoner i testsettet utgjør her 30% av observasjonene som benyttes for å trene og validere modellene

stedet for Byparken-Nesttun, er RMSE og MAE henholdsvis 7.7% og 7.4% lavere. Det foreligger således antydninger til at presisjonen gjennomgående blir høyest når Byparken-Lagunen benyttes som trenings- og valideringssett for begge metodene.

Sammenligner vi videre presisjonsforskjellen mellom Elastic Net og Super Learner når Byparken-Lagunen benyttes som treningssett for begge metodene, ser vi at Super Learner oppnår en RMSE og en MAE som henholdsvis er 4.0% og 2.3% lavere enn Elastic Net. Dette indikerer at Super Learner utgjør den mest presise modellen når boligprisene i Åsane skal predikeres i tiden før Bybanen. For å få en oversikt over årlig presisjon, fremstilles utviklingen i RMSE grafisk.

Figur 5.3: Visualisering av RMSE for prediksjoner av Byparken-Åsane - Elastic Net



Figur 5.4: Visualisering av RMSE for prediksjoner av Byparken-Åsane - Super Learner

Den grafiske fremstillingen av RMSE viser at prediksjonene for 2001, og til dels 2004 og 2005, er ustabile. Videre er det interessant å se at Super Learner predikerer langt bedre i år 2001 når Byparken-Lagunen benyttes som treningssett. Når vi studerer predikerte boligpriser i år 2001 ser vi at prediksjonene utført med Super Learner oppnår en RMSE på omtrent 22500 når Byparken-Flesland benyttes som trenings- og valideringssett. Gjennom en dypere analyse av prediksjonene, observerer vi at 129 av 873 prediksjoner har en RMSE på over 40000 når Byparken-Flesland benyttes som treningssett. Det er usikkert hva som er kilden til de store feilprediksjonene. Det kan imidlertid tenkes å være et resultat av et utilstrekkelig antall observasjoner mellom Nesttun og Flesland dette året. Når maskinlæring benyttes for å predikere kan ustabile prediksjoner oppstå på grunn av mangel på datapunkter i treningsdata for visse typer observasjoner (Hastie et al., 2009). Etersom vårt fokus er å predikere boligpriser rundt Åsane-området med Bybanen, og å diskutere implikasjonene dette medfører, må vi avgrense hva vi går i dybden på. I denne sammenheng er vi derfor ikke så interessert i å analysere hvorfor visse treningssett gir mer ustabile prediksjoner, men derimot hvilket treningssett som gir grunnlag for mest presise prediksjoner. Når Byparken-Lagunen utgjør trenings- og valideringssettet ser vi videre at Super Learner presterer gjennomgående jevnt sammenlignet med Elastic Net,

som i relativ forstand viser til mer volatile resultater.

Det er viderer viktig å påpeke at begrenset tilgang til boligtransaksjoner og et begrenset detaljnivå i disse transaksjonene er en utfordring når man trener maskinslæringsmodeller (Hastie et al., 2009). På tross av at vi har inkludert mange sentrale variabler fra den hedonistiske prisingsmodellen, i tillegg til sentrale makroøkonomiske variabler, finnes det flere variabler som med fordel kunne vært inkludert for å oppnå høyere presisjon. For eksempel vil variabler som tallfester boligens avstand til nærmeste skole, boligens utsikt, samt boligens standard potensielt medført økt presisjon. Slik data er imidlertid ikke lett tilgjengelig, og fører til en større usikkerhet i prediksjonene. Denne svakheten er drøftet i kapittel 6.4. Utviklingen av predikerte gjennomsnittlige boligpriser for alle modellene illustreres grafisk i Figur 5.5 og 5.6.

Figur 5.5: Gjennomsnittet av predikerte boligpriser mellom Byparken og Åsane i perioden 2001 til 2007 - Elastic Net.



Figur 5.6: Gjennomsnittet av predikerte boligpriser mellom Byparken og Åsane i perioden 2001 til 2007 - Super Learner.



Det er viktig å bemerke at hvert årlige prediksjonspunkt i Figur 5.5 og 5.6 representerer et gjennomsnitt av mangfoldige enkeltprediksjoner. Prediksjonsavviket er definert som den absolutte forskjellen mellom hver predikerte boligpris og faktisk boligpris på individnivå, og ikke som differansen mellom gjennomsnittlig prediksjon og faktisk gjennomsnitt. Derfor kan man ikke se hvilken modell som oppnår høyst presisjon med bakgrunn i den grafiske fremstillingen over. Likevel er det interessant å se at Elastic Net ser ut til å undervurdere gjennomsnittlig prisnivå i større grad enn Super Learner. Begge modellene følger imidlertid endringen i prisnivå rimelig godt.

Vi har således sett at måltallene for presisjon er entydig i favør av at observasjonene mellom Byparken og Lagunen skal utgjøre trenings- og valideringssettet både når Elastic Net og Super Learner benyttes til å predikere boligprisene i Åsane med Bybanen. Videre indikerer foregående resultater at Super Learner predikerer boligprisene mellom Byparken og Åsane mest presist for perioden 2001-2007 når Byparken-Lagunen benyttes som trenings- og valideringssett.

I tillegg til at Byparken-Lagunen viser seg å gi mest presise prediksjonene uten Bybanen, er det flere aspekter som tilsier at Bybanen vil ha en lignende påvirkning på boligprisene langs traseen til Åsane som den har hatt på boligprisene langs strekket mellom Byparken og Lagunen. Et viktig momentet i denne sammenheng er at den planlagte traseen mellom Byparken og Åsane har flere likhetstrekk med strekket fra Byparken til Lagunen enn hva som er tilfelle for andre inndelinger av treningssettet. Dersom vi trener en modell på et område hvor bybaneeffekten kan antas å være relativt lik det vi antar den også vil være for observasjonene vi predikerer, vil dette følgelig føre til mer nøyaktige prediksjoner.

Det finnes mange argumenter for å hevde at bybanetraseen mellom Byparken og Lagunen er relativt lik den planlagte traseen til Åsane. For det første ser vi fra Tabell A1.1 i appendiks at lengden på traseene mellom Byparken-Lagunen og Byparken-Åsane henholdsvis er 13.4 kilometer og 12.8 kilometer. Reisetiden fra start- til endestopp er også relativt lik. Det foreligger derimot fem flere bybanestopp for strekket Byparken-Lagunen enn hva som planlegges for traseen mellom Byparken og Åsane. Videre fremkommer det fra samtaler med medlem i utvalg for miljø- og utvikling i SV, Jarle Brattespe, og medlem i utvalg for miljø- og byutvikling i Høyre, Charlotte Spurkeland, at formålet med Bybanen til både Lagunen og til Åsane er å knytte områdene rundt som ligger i utkanten av sentrum mer mot sentrum. Det samme gjelder imidlertid også for Byparken-Nesttun, men utfordringen med å kun benytte Byparken-Nesttun som treningssett er at lengden på traseen er vesentlig kortere. Formålet med å forlenge strekket fra Lagunen til Flesland var derimot å knytte flyplassen og arbeidsplassene på Kokstad sammen med sentrum og boligområdene rundt Lagunen og Nesttun. Det faktum at både selve bybanetraseen og traseens formål sammenfaller godt, er momenter som indikerer at boligprisene vil oppleve en likere utvikling enn dersom traseene og formålet med traseene hadde vært mer ulike.

Med bakgrunn i diskusjonen over har vi belegg for å tro at Byparken-Lagunen vil utgjøre det beste treningssettet av følgende grunner. For det første er presisjonen høyest når dette treningssettet benyttes. For det andre foreligger det indikasjoner som tyder på at Bybanens påvirkning på boligprisene mellom Byparken og Vågsbotn vil ligne påvirkningen Bybanen har hatt på boligprisene mellom Byparken og Lagunen. Selv om vi finner Super Learner mest presis også i dette delkapittelet, ønsker vi av samme grunn som nevnt i kapittel 5.2.1 likevel å beholde Elastic Net i videre analyse.

5.2.3 Bybaneeffektens overførbarhet

Så langt i analysen har vi funnet indikasjoner på at både Elastic Net og Super Learner vil predikere boligprisene mellom Byparken og Åsane mest presist uten transformerte variabler. Vi har også identifisert hvilket treningssett som vil danne det videre grunnlaget for prediksjonene av boligprisene mellom Byparken og Åsane.

Det er videre viktig å vite hvorvidt bybaneeffekten kan antas å være overførbar mellom to ulike geografiske områder. Med *bybaneeffektens overførbarhet* mener vi om boligprisene i et område hvor Bybanen foreligger predikeres mer presist når vi bruker et treningssett fra et annet geografisk område som også er påvirket av nærhet til Bybanen, sammenlignet med presisjonen vi oppnår om vi benytter treningsdata som er upåvirket av Bybanen. Dette er viktig informasjon ettersom vi må vite om vi kan trene en modell basert på observasjoner fra et område hvor Bybanen foreligger (Byparken-Lagunen), til å predikere boligpriser i et annet område hvor Bybanen kommer på et senere tidspunkt (Byparken-Åsane).

Vi refererer videre til transaksjoner som registreres før starten på en ny bybaneutbygging som *før* Bybanen. Transaksjoner som derimot forekommer etter Bybanens ferdigstillelse, refereres til som *etter* Bybanen. For å teste bybaneeffektens overførbarhet vil vi predikere boligprisene tilknyttet området som strekker seg fra Nesttun til Lagunen etter Bybanens inntreden og seks år frem i tid. Vi vil gjennomføre denne prediksjonen basert på to ulike treningssett. Det ene vil bestå av boligtransaksjoner mellom Byparken og Nesttun etter Bybanens ferdigstillelse og seks år frem i tid. Det andre vil derimot bestå av boligtransaksjoner mellom Byparken og Nesttun før Bybanen. For å underbygge at resultatene ikke skyldes tilfeldig støy i datagrunnlaget, ønsker vi også å predikere boligprisene mellom Nesttun og Lagunen før Bybanen med bruk av tilsvarende treningssett.

Hypotesen vår er at prediksjonsmodeller hvor treningsdata består av transaksjoner etter Bybanen vil resultere i høyere presisjon enn treningsdata fra tiden før Bybanen, når vi predikerer observasjoner for tiden etter Bybanen. Dette valideres ved at også det motsatte er tilfellet: At treningsdata før Bybanen vil predikere observasjoner som også er fra tiden før Bybanen mer nøyaktig enn en modell basert på et treningssett som er fra tiden etter Bybanen. Dette på tross av at trenings- og testsettet er hentet fra to ulike områder.

Tabell 5.4: RMSE for predikerte boligpriser mellom Nesttun og Lagunen etter Bybanens ferdigstillelse.

| | Elastic Net | Super Learner |
|---------------|--------------------|----------------------|
| Før Bybanen | 10760.4 | 8384.8 |
| Etter Bybanen | 10355.0 | 8352.6 |
| Differanse | 405.4 | 32.2 |

Treningssettene består av observasjoner henholdsvis fra perioden før og etter Bybanens ferdigstillelse mellom Byparken-Nesttun

RMSE lik 10760.4 representerer Elastic Net sin presisjon når observasjoner mellom Byparken og Nesttun før Bybanens ferdigstillelse benyttes til å predikere boligprisene mellom Nesttun og Lagunen for perioden etter ferdigstillelse. RMSE lik 10355.0 beskriver derimot Elastic Net sin presisjon når observasjoner mellom Byparken og Nesttun etter Bybanens ferdigstillelse benyttes til å predikere boligprisene mellom Nesttun og Lagunen for perioden etter ferdigstillelse. Tilsvarende gjelder også for Super Learner.

Presisjonen er således bedre for både Elastic Net og Super Learner når vi predikerer boligpriser for perioden etter Bybanens ferdigstillelse basert på et treningssett som også er fra perioden etter Bybanens inntreden. Sammenligningsgrunnlaget er presisjonen vi oppnår når boligprisene etter ferdigstillelse predikeres basert på treningsdata fra perioden før ferdigstillelse. Når Elastic Net benyttes er differansen i RMSE 405.4, noe som tilsvarer en forbedring på 3.9%. For Super Learner er differansen kun 0.4%. En slik marginal forbedring tyder enten på tilfeldigheter, eller en veldig liten, men reell effekt.

Tabell 5.5: RMSE for predikerte boligpriser mellom Nesttun og Lagunen før Bybanens ferdigstillelse.

| | Elastic Net | Super Learner |
|---------------|--------------------|----------------------|
| Før Bybanen | 9178.0 | 6412.2 |
| Etter Bybanen | 10112.3 | 8245.4 |
| Differanse | 934.3 | 1833.2 |

Treningssettene består av observasjoner henholdsvis fra perioden før og etter Bybanens ferdigstillelse mellom Byparken-Nesttun.

Dersom vi predikerer boligprisene før Bybanens inntreden med et trenings- og valideringssett fra tiden før Bybanen ble ferdigstilt, oppnår vi mer entydige resultater. Tabell 5.5 viser nettopp dette: RMSE er lavere når vi predikerer Nesttun-Lagunen før

Bybanen når vi trener og validerer på observasjoner fra Byparken-Nesttun før Bybanen, sammenlignet med hva vi oppnår når boligprisene før Bybanen predikeres ved hjelp av et datagrunnlag fra perioden etter Bybanen. Begge modellene peker i samme retning. Differansen i RMSE er henholdsvis 943.3 for Elastic Net og 1833.2 for Super Learner, noe tilsvarer en forbedring på henholdsvis 10.2% og 28.6%.

Måltallene for presisjon viser at presisjonen øker når vi predikerer observasjoner som er påvirket av bybaneeffekten mellom Nesttun og Lagunen ved å benytte et treningssett som også er påvirket av Bybanen mellom Byparken og Nesttun, sammenlignet med hvilken presisjon vi oppnår for tilsvarende prediksjoner hvor treningsdata ikke er påvirket av den samme effekten. Det finnes således indikasjoner på at bybaneeffekten er overførbar mellom to geografiske områder.

Merk at det imidlertid foreligger flere usikkerhetsmomenter bak disse indikasjonene. For det første er det vanskelig å kontrollere for andre forhold som ikke er tilknyttet Bybanen, og som potensielt medfører høyere presisjon når observasjoner predikeres basert på et treningssett fra samme periode. I så tilfelle vil det være vanskelig å si om det er Bybanen som fører til bedre prediksjoner eller om det er andre forhold ved data som medfører økt presisjon. For det andre er det viktig å påpeke at området mellom Byparken og Nesttun trolig har flere likhetstrekk med området tilknyttet Nesttun-Lagunen, enn hva som gjelder for området mellom Byparken og Lagunen mot Byparken-Åsane. Dette skyldes at bybanetraseen mellom Byparken-Nesttun og Nesttun-Lagunen begge ligger på samme side av Bergen sentrum.

5.2.4 Betydningen av å inkludere variabelen *avstand til nærmeste bybanestopp* i våre prediksjoner

Videre ønsker vi å analysere viktigheten av å inkludere en *bybanespesifikk* variabel når boligpriser i nærheten av Bybanen predikeres. I denne utredningen er variabelen *avstand til nærmeste bybanestopp* (ATB) benyttet som en bybanespesifikk variabel.

For å analysere effekten ATB har på prediksjonenes presisjon, har vi utarbeidet to ulike prediksjonsmodeller. Begge modellene predikerer boligprisene mellom Nesttun og Lagunen etter Bybanens inntreden. Treningssettet for begge modellene består av observasjoner mellom Byparken og Nesttun for perioden etter Bybanens ferdigstillelse for dette området.

Det eneste som skiller de to modellene er at variabelen ATB kun er inkludert i én av dem. Dersom presisjonen øker når vi predikerer boligprisene mellom Nesttun og Lagunen, hvor ATB er en del av variabelgrunnlaget, er dette en indikasjon på at det er viktig å inkludere en bybanespesifikk variabel når vi predikerer boligprisene i Åsane.

Tabell 5.6: RMSE og MAE for Elastic Net med og uten variabelen *avstand til nærmeste bybanestopp*

| | Uten ATB | Med ATB |
|------|----------|---------|
| RMSE | 8806.0 | 8348.9 |
| MAE | 6622.2 | 6314.4 |

Tabell 5.7: RMSE og MAE for Super Learner med og uten variabelen *avstand til nærmeste bybanestopp*

| | Uten ATB | Med ATB |
|------|----------|---------|
| RMSE | 9693.5 | 8939.8 |
| MAE | 7500.5 | 6762.5 |

Fra tabellene over fremkommer det at Elastic Net oppnår en RMSE og MAE som henholdsvis er 5.2% og 4.9% lavere når den bybanespesifikke variabelen inkluderes i datagrunnlaget. For Super Learner er RMSE og MAE henholdsvis 7.8% og 9.8% lavere når ATB inkluderes. Dersom vi hadde inkludert variabler i treningssettet som hadde fanget opp den samme variasjonen som den bybanespesifikke variabelen gjør, ville vi stått i fare for at Elastic Net hadde ekskludert en eller flere av de korrelerte variablene for å unngå overfitting. I dette tilfellet ble den bybanespesifikke variabelen beholdt, i tillegg til at presisjonen økte. Dette tyder på at avstand til Bybanen står sentralt i vår analyse. Vi vil i kapittel 5.4.5 tallfeste hvor viktig nærhet til Bybanen estimeres å være når vi predikerer boligprisene i Åsane.

5.3 Prediksjoner av boligpriser tilknyttet Åsane-traseen

Før vi presenterer predikerte boligpriser tilknyttet bybanetraseen mot Åsane, vil vi kort repetere og diskutere analysens foreløpige funn. For det første har vi besluttet at variablene som inkluderes i Elastic Net og Super Learner ikke skal transformeres når boligprisene tilknyttet Åsane-traseen predikeres. For det andre har vi funnet ut at boligtransaksjonene mellom Byparken og Lagunen vil utgjøre prediksjonsmodellens trenings- og valideringssett. Videre finnes det antydninger til at bybaneeffekten er overførbar mellom to ulike geografiske områder. I tillegg foreligger det indikasjoner på at det kan være sentralt å inkludere en bybanespesifikk variabel, som ATB, for å oppnå høyere presisjon når formålet er å predikere boligpriser i områder hvor et offentlig gode som Bybanen skal introduseres.

Vi ønsker å inkludere disse funnene i en prediksjonsmodell som skal predikere utviklingen i boligprisene rundt bybanetraseen mellom Byparken og Vågsbotn fra og med 2010 til og med 2019, forutsatt at Bybanen til Åsane hadde vært ferdigstilt i 2010. Dette betyr at prediksjonsmodellene som fremgår fra og med kapittel 5.4 benytter ikke-transformerte variabler, trenes og valideres på observasjoner mellom Byparken og Lagunen og inkluderer variabelen ATB.

En viktig observasjon fra foregående analyse er at Super Learner har predikert mer presist enn Elastic Net gjennom hvert steg i prediksjonsrammeverkets oppbygging. Spesielt interessant er det at metoden viser til høyest presisjon i to spesifikke tilfeller. For det første ser vi fra kapittel 5.2.1 og 5.2.3 at Super Learner oppnår høyest presisjon når vi trener og validerer prediksjonsmodellene på et område hvor Bybanen eksisterer til å predikere boligprisene i et annet område hvor Bybanen også foreligger. Dette impliserer at metoden er mer presis enn Elastic Net når observasjoner som er påvirket av Bybanen benyttes for å predikere andre observasjoner som også er påvirket av Bybanen. For det andre oppnår Super Learner høyest presisjonen også når vi predikerer boligprisene rundt Åsane-traseen før Bybanens inntreden (2001-2007), trent og validert på observasjoner fra Byparken-Lagunen for perioden før Bybanen (2001-2007). Disse momentene indikerer at Super Learner er en mer presis metode enn Elastic Net når vi bruker et trenings- og valideringssett fra Byparken-Lagunen til å predikere boligprisene i Åsane med Bybanen. Selv om Super

Learner trolig utgjør den mest presise modellen for prediksjoner tilknyttet Byparken-Åsane med Bybanen, ønsker vi likevel å inkludere Elastic Net i videre prediksjoner. Dette gjøres for å sammenligne hvordan boligprisutviklingen predikeres når to forholdsmessig forskjellige metoder benyttes.

Som nevnt i utredningens innledning, ønsker vi å predikere boligpriser for en foregående periode. Dette kommer av at det er vanskelig å si noe om fremtidig boligprisutvikling for vårt datasett. I tillegg vet vi ikke når Bybanen til Åsane faktisk blir ferdigstilt. Prediksjoner av historiske priser er likevel relevante når vi ønsker å studere utvikling i boligpriser fra og med Bybanen til Åsane faktisk kommer og ni år frem i tid. Dette fordi boligprisenes utvikling er uavhengig av det absolutte prisnivået på tvers av perioder. I fortolkningene av prediksjonene som fremgår videre er vi derfor hovedsaklig opptatt av *boligprisenes nivåforskjeller og utvikling* i reell prisvekst.

Prediksjonene tilknyttet Åsane-traseen vil sammenlignes med følgende referanser:

1. Historiske boligpriser fra og med 2010 til og med 2019 mellom Byparken og Åsane. Dette er priser som er upåvirket av Bybanen. Vi vil ofte benytte referansen “boligpriser uten Bybanen” når vi diskuterer historiske priser. I figurene som fremgår fra og med kapittel 5.4 vil datasettet kalles “Historiske kvm.priser (B-Å)”, hvor B-Å indikerer Byparken-Åsane.
2. Prisutviklingen for området mellom Byparken og Lagunen fra og med 2010 til og med 2019. Dette representerer treningsdata. Dette datasettet vil omtales som “Treningsdata (B-L)”, hvor B-L representerer boligtransaksjoner mellom Byparken og Lagunen.

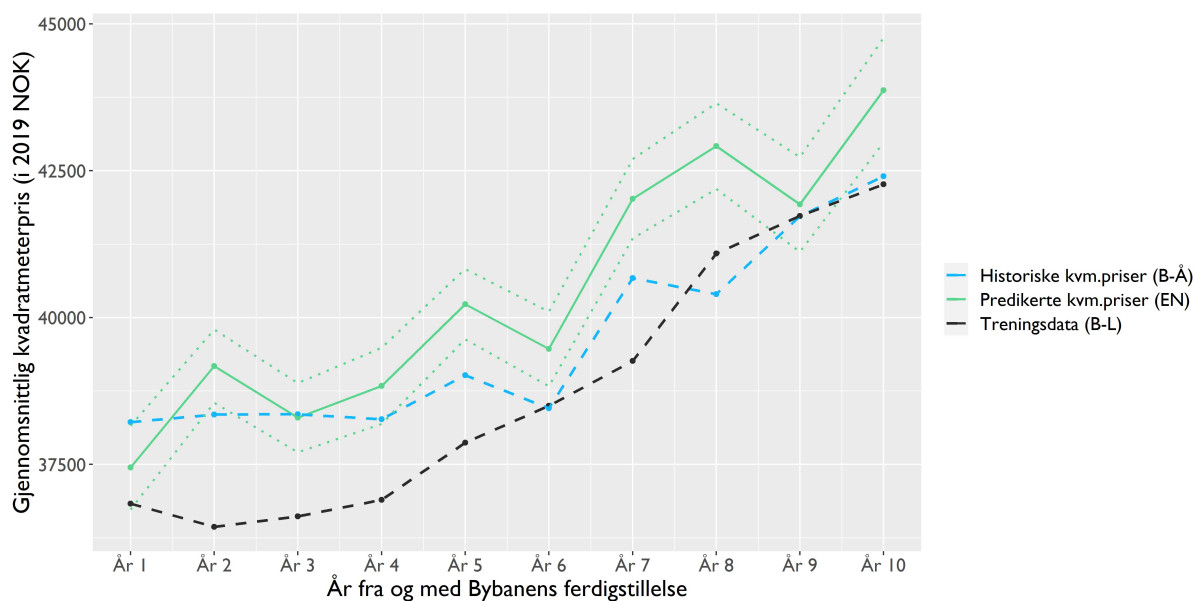
Grunnen til at vi ønsker å vurdere prediksjonene tilknyttet Åsane-traseen mot historiske boligpriser, er for å muliggjøre en sammenligning av boligprisene med og uten Bybanen. Dette kan videre brukes i en kvantifisering av verdien Bybanen predikeres å tilføre boligprisutviklingen i Åsane fra den kommer og ni år frem i tid. Treningsdata er inkludert for å illustrere utviklingen mellom Byparken og Lagunen. I følgende analyse vil Elastic Net og Super Learner i enkelte tilfeller forkortes med henholdsvis *EN* og *SL*.

5.4 Prediksjoner av boligpriser tilknyttet Åsane-trasseen fra og med Bybanens ferdigstillelse og ni år frem i tid

I første omgang vil vi predikere boligpriser for alle boliger²⁶ som er lokalisert innen 3000 meter fra nærmeste bybanestopp tilknyttet den fremtidige bybanetrasseen til Åsane. Deretter predikerer vi boligpriser for boliger henholdsvis innen 1000 meter, og mellom 1000 og 3000 meter fra nærmeste bybanestopp. Dette prediksjonsutvalget er gjort for å analysere om det predikeres å være en betydelig forskjell i prisutvikling for boliger som ligger nært Bybanen kontra boliger lengre borte. For å videre undersøke om prisveksten predikeres å være ulik for eneboliger og leiligheter, vil vi også utarbeide prediksjoner for disse boligtypene hver for seg. Avslutningsvis vil vi illustrere hvordan den kategoriske variabelen *gåavstand til nærmeste bybanestopp* påvirker predikert kvadratmeterpris over tid.

5.4.1 Alle boliger innen 3000 meter fra nærmeste bybanestopp

Figur 5.7: Gjennomsnittet av predikerte boligpriser for alle boliger innen 3000 meter fra nærmeste Bybanestopp - Elastic Net



²⁶Alle boligtyper inkluderer alle eneboliger, tomannsboliger, rekkehus og leiligheter i datasettet

Før vi studerer prediksjonene som fremgår av Figur 5.7 i detalj, ønsker vi å utdype noen momenter som er sentrale for en helhetlig forståelse av gjenværende analyse.

I figuren presenteres tre ulike grafer. Grønn linje, *Predikerte kvm.priser (EN)*, representerer gjennomsnittet av predikerte boligpriser tilknyttet Åsane-trasseen fra og med ferdigstillelsen av Bybanen til Åsane og ni år frem i tid, gitt at Bybanen hadde eksistert i det aktuelle området fra 2010 til 2019.

Videre representerer den blå linjen, *Historiske kvm.priser (B-Å)*, historiske gjennomsnittlige kvadratmeterpriser for det utvalget av boliger i Åsane som vi ønsker å predikere med bybaneeffekten. Historiske priser representerer således boligprisene rundt Åsane-trasseen før Bybanen var et faktum, og som nevnt, vil vi referere til de historiske prisene som *boligpriser uten Bybanen*. Ved å sammenligne grønn linje, som representerer predikerte kvadratmeterpriser tilhørende boliger i Åsane-området med Bybanen, mot historiske boligpriser, derav boligpriser uten Bybanen, får vi tallfestet hvor viktig Bybanen predikeres å være for utviklingen i boligpriser rundt Åsane når den kommer. Differansen i vekst mellom blå og grønn linje er således effekten Bybanen predikeres å ha på boligprisene langs Åsane-trasseen fra og med Bybanens ferdigstillelse og ni år frem i tid.

Sort linje, *Treningsdata (B-L)*, representerer som nevnt utviklingen for gjennomsnittlig kvadratmeterpris for treningsdata. Linjen illustrerer følgelig utviklingen i kvadratmeterpris for boliger tilknyttet bybanetrasseen mellom Byparken og Lagunen fra og med Bybanens ferdigstillelse til Åsane og ni år frem i tid.

Verdiene langs den horisontale akse "År fra og med Bybanens ferdigstillelse" har ulik betydning for grafene som representeres i Figur 5.7. For blå linje er År 1 definert som 2010 og År 10 som 2019. For grønn linje representerer År 1 og År 10 henholdsvis det året vi forestiller oss at Bybanen til Åsane ble ferdigstilt og det siste året i prediksjonsperioden. Dette tilsvarer 2010 og 2019.

Som presentert i Tabell 5.1 ble traseene mellom Byparken og Nesttun, og mellom Nesttun og Lagunen ferdigstilt i henholdsvis 2010 og 2013. Dette medfører at År 1 for den sorte linjen, "Treningsdata (B-L)", representerer året da Bybanen ble ferdigstilt henholdsvis mellom Byparken og Nesttun, og mellom Nesttun og Lagunen. Ettersom de respektive bybanetraseene ble ferdigstilt på ulike tidspunkter, tilsvarer År 1 2010 for observasjoner

mellom Byparken og Nesttun, men derimot 2013 for observasjoner mellom Nesttun og Lagunen. På grunn av at Bybanen mellom Nesttun og Lagunen ikke har stått ferdig i mer enn sju år (2013-2019), består treningsdata fra År 8 til og med År 10 etter Bybanens ferdigstillelse kun av observasjoner tilknyttet trasseen mellom Byparken og Nesttun. Dette kan påvirke prediksjonene fra År 8 og utover ettersom vi i kapittel 5.2.2 konkluderte med at Byparken-Lagunen utgjør et bedre trenings- og valideringsgrunnlag enn Byparken-Nesttun.

For en ytterligere vurdering av prediksjonene, er det også utarbeidet prediksjonsintervaller for gjennomsnittet av de predikerte kvadratmeterprisene. Et 95% prediksjonsintervall indikerer at prediksjonene i 95 % av tilfellene vil ligge innenfor det oppgitte intervallet (Hyndman, 2013). Et smalt konfidensintervall indikerer derfor at prediksjonene er mer konsistente. Fra figuren over fremgår det at konfidensintervallenes spredning er relativt lik gjennom hele perioden. Konfidensintervallene tallfestes i Tabell A3.4.

Et sentralt aspekt ved prediksjonene er at vi ikke har inkludert observasjoner fra Åsane i opptreningen eller valideringen av modellene. Det vil derfor foreligge Åsane-spesifikke effekter modellene ikke fanger opp. Vi har søkt å løse dette ved å velge ut et datasett som representerer observasjonene rundt Åsane-området, og kommet frem til at boligtransaksjonene mellom Byparken og Lagunen utgjør en god tilnærming. Utvelgelsen av lignende observasjoner vil derimot aldri eksakt replikere observasjonene som foreligger rundt Åsane. Dette medfører at resultatene må tolkes varsomt og at de trolig er mer upresise enn hva som ville vært tilfelle dersom vi hadde inkludert observasjoner rundt Åsane i opptreningen og valideringen av modellene. I mangel på testsett vet vi heller ikke hvor presise prediksjonene faktisk er. Denne svakheten vil drøftes videre i kapittel 6.4.

Når vi analyserer prosentvis utvikling for predikerte kvadratmeterpriser trenger vi derimot ikke å hensynta absolutte prisforskjeller mellom Åsane-området og treningssettet. Selv om modellen ikke nødvendigvis treffer så godt på prisnivå, kan den likevel gi oss verdifull informasjon i form av prosentvis endring. Den gjennomsnittlige prosentvise årlige prisendringen som fremgår av prediksjonene i Figur 5.7 illustreres i tabellen under.

Tabell 5.8: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Elastic Net, historiske priser og treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 4.61 % | 0.34 % | -1.07 % |
| År 2-3 | -2.24 % | 0.00 % | 0.49 % |
| År 3-4 | 1.41 % | -0.21 % | 0.77 % |
| År 4-5 | 3.58 % | 1.95 % | 2.64 % |
| År 5-6 | -1.89 % | -1.44 % | 1.65 % |
| År 6-7 | 6.48 % | 5.76 % | 1.99 % |
| År 7-8 | 2.14 % | -0.67 % | 4.66 % |
| År 8-9 | -2.32 % | 3.28 % | 1.55 % |
| År 9-10 | 4.63 % | 1.63 % | 1.29 % |
| Snitt | 1.82 % | 1.18 % | 1.55 % |

Kolonnen “Predikert vekst - EN (Med Bybanen)” representerer årlig vekst i predikert kvadratmeterpriser mellom 2010 og 2019. Dette tilsvarer årlig prosentvis endring for grønn linje i Figur 5.7. “Historiske priser (Uten Bybanen)” illustrerer derimot prosentvis utvikling i faktiske inflasjonsjusterte boligpriser mellom Byparken og Åsane fra og med 2010 til og med 2019 (blå linje). Eksempelvis predikeres den prosentvise prisutviklingen det første året etter Bybanens inntreden til Åsane å være 4.61%. Historiske boligpriser mellom Byparken og Åsane opplevde derimot en vekst på 0.34% fra 2010 til 2011, som representerer År 1 til År 2 i Figur 5.7. Nederste rad i tabellen representerer gjennomsnittlig prosentvis prisvekst for alle årene fra og med Bybanens ferdigstillelse.

Gjennomsnittlig årlig boligprisvekst dersom Bybanen til Åsane hadde vært et faktum fra 2010, predikeres å være 1.82% for alle boliger lokalisert innen 3000 meter fra nærmeste bybanestopp. Gjennomsnittlig historisk årlig vekst, derav priser uten Bybanen, har derimot vært 1.18% mellom 2010 og 2019. Ved å studere differansen mellom predikerte kvadratmeterpriser og historiske priser, finner vi hvor stor påvirkning Bybanen predikeres å ha på boliger tilknyttet den nye bybanetrasseen til Åsane. Differansen mellom gjennomsnittlig årlig vekst med Bybanen (1.82%) og uten Bybanen (1.18%) er 0.64 prosentpoeng. Dette impliserer at den årlig boligprisveksten i gjennomsnitt predikeres å være 54.2% høyere når Bybanen kommer, sammenlignet med hva den historiske prisveksten har vært gjennom perioden 2010-2019.

Vi ønsker å legge frem et hypotetisk eksempel for å tallfeste meravkastningen som i gjennomsnitt predikeres å foreligge når Bybanen kommer til Åsane. I eksempelet kjøper en investor en leilighet for 2 millioner kroner idet Bybanen til Åsane ferdigstilles. Leiligheten selges ni år senere og investoren får utbetalt 2.35 millioner kroner²⁷. Dette impliserer en avkastning på 350 000 kroner, eller 17.5%. Dersom Bybanen ikke hadde vært et faktum ville den samme boligen oppnådd en pris på 2.22 millioner kroner²⁸. Dette tilsvarer en avkastning på 220 000 kroner eller 10.0%. En avkastning på henholdsvis 17.5% og 10.0% over en periode på 10 år virker med første øyekast lavt. Dersom vi imidlertid tar høyde for to sentrale forhold bak tallene, fremkommer det at avkastningen ikke er så lav. For det første er avkastningen inflasjonsjustert både etter konsumprisindeksen og boligprisindeksen for Bergensområdet. De 350 000 kronene representerer derfor avkastning utover den gjennomsnittlige boligprisveksten i Bergen, i 2019-kroner. Dersom Bybanen kommer til Åsane, vil investor, ved å kjøpe en bolig innen 3000 meter fra et bybanestopp, i gjennomsnitt oppnå en avkastning som er 350 000 kroner høyere enn hva investoren ville oppnådd dersom han eller hun hadde kjøpt en bolig til 2 millioner kroner som hadde opplevd en gjennomsnittlig prisvekst i Bergensområdet. For det andre er det vanlig praksis at store deler av et boligkjøp er lånefinansiert. Fra 2011 og frem til i dag har maksimal belåningsgrad²⁹ for nedbetalingslån vært 85% (Finanstilsynet, 2020). Gjennomsnittlig belåningsgrad har derimot vært omtrent 65% de siste årene (Finanstilsynet, 2020). Dersom investoren i eksempelet har en belåningsgrad på 85% vil innskutt egenkapital kun utgjøre 300 000 kroner av boligens totale verdi på 2 millioner kroner. Dette impliserer at investorer som investerer i bolig i nærheten av Bybanen i gjennomsnitt vil oppleve at deres innskutte egenkapital vil øke med en realverdi på 350 000 kroner, eller 116.67% mer enn en gjennomsnittlig bolig ellers i Bergen. Ser vi derimot på vekstdifferansen med og uten Bybanen for den samme boligen, vil økningen i reell egenkapital være 130 000 kroner høyere for boligen etter at Bybanen kommer. Eksempelet vårt illustrerer at selv relativt små forskjeller i prosentvis vekst kan medføre store forskjeller i avkastning når vi hensyntar både inflasjonsjustering og belåningsgrad.

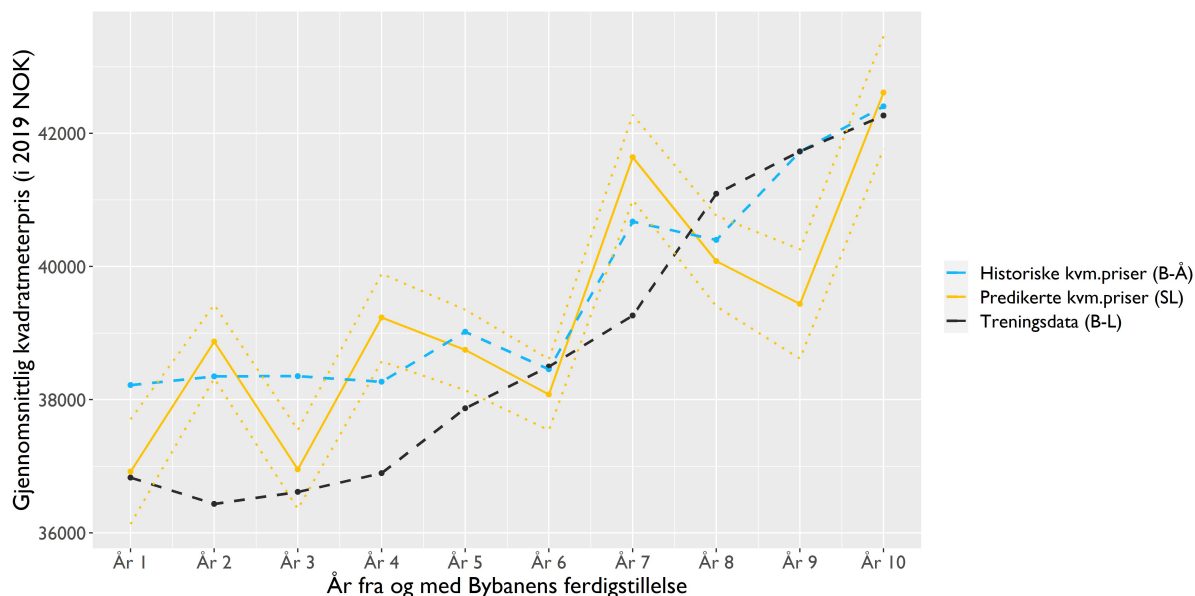
²⁷ $2\,000\,000 \times 1.0182^9$

²⁸ $2\,000\,000 \times 1.0118^9$

²⁹Belåningsgrad representerer lånets totale størrelse sammenlignet med boligens verdi

Figur 5.8 illustrerer tilsvarende prediksjoner som Figur 5.7, bortsett fra at Super Learner benyttes for å modellere.

Figur 5.8: Gjennomsnittet av predikerte boligpriser for alle boliger innen 3000 meter fra nærmeste Bybanestopp - Super Learner



Fra den grafiske fremstillingen over fremkommer det at gjennomsnittet av predikerte boligpriser utført med Super Learner er mer volatile enn hva som er tilfellet når Elastic Net benyttes for å predikere. På tross av at vi her observerer høy volatilitet, kom vi frem til at Super Learner trolig vil levere mer presise prediksjoner enn Elastic Net når vi predikerer boligpriser i Åsane for tiden etter Bybanens ferdigstillelse. Ettersom vi ikke har et tilgjengelig testsett fra Åsane, har vi ikke grunnlag for å fastslå hvilken metode som er mest presis når vi predikerer boligpriser mellom Byparken og Vågsbotn. Likevel sørger kryssvalidering i Super Learner for at problemer tilknyttet overfitting minimeres, noe som medfører at det ikke skal være et problem å inkludere et bredt modellspekter når metoden benyttes. Sett i lys av historisk boligprisutvikling kan det likevel virke som at modellen i gjennomsnitt resulterer i mer volatile prediksjoner enn hva som er naturlig i boligmarkedet.

Tabell 5.9: Koeffisienter for de inkluderte metodene i Super Learner

| | SL.mean | SL.lm | SL.glmnet | SL.nnet | SL.kernelKnn | SL.polymars |
|-------|---------|--------|-----------|---------|--------------|-------------|
| År 1 | 0 | 0.4309 | 0 | 0 | 0.0597 | 0.5094 |
| År 2 | 0 | 0.1982 | 0.4362 | 0 | 0.2171 | 0.1485 |
| År 3 | 0 | 0.4318 | 0.0526 | 0 | 0.1201 | 0.3955 |
| År 4 | 0 | 0.3692 | 0.4396 | 0 | 0.0700 | 0.1212 |
| År 5 | 0 | 0.6642 | 0 | 0 | 0.0348 | 0.3010 |
| År 6 | 0 | 0.2358 | 0.3363 | 0 | 0.0472 | 0.3807 |
| År 7 | 0 | 0.5371 | 0 | 0 | 0.1230 | 0.3400 |
| År 8 | 0 | 0.4862 | 0 | 0 | 0.2155 | 0.2983 |
| År 9 | 0 | 0.1456 | 0.1709 | 0 | 0.1415 | 0.5421 |
| År 10 | 0 | 0.3151 | 0 | 0 | 0.0959 | 0.5890 |

Tabell 5.9 viser den endelige vektingen for hver av de inkluderte metodene i Super Learner, hvor vektingen av metodene hvert år summeres til 1. Hvor stor andel hver av de respektive metodene skal utgjøre av den helhetlige modellen bestemmes gjennom kryssvalidering in-sample, hvor formålet er å minimere valideringssettets RMSE. Fra tabellen fremkommer det at Sample Mean (SL.mean) og Nevrale Nettverk (SL.nnet) gjennomgående ekskluderes i utarbeidelsen av prediksjonene. Vi legger videre merke til at ordinær lineær regresjon (SL.lm) og Multivariate Adaptive Polynomial Spline (SL.polymars) utgjør størsteparten av den helhetlige modellen. Elastic Net (SL.glmnet) er også sentral for enkelte perioder. Felles for ordinær lineær regresjon og Elastic Net er at begge metodene antar lineære sammenhenger i datagrunnlaget. MARS (Multivariate Adaptive Polynomial Spline) kan derimot tilpasse sine prediksjoner etter ikke-lineære forhold gjennom både en optimal inndeling av treningsdata i knuter³⁰ og polynomisk regresjon (Friedman, 1991).

Videre i analysen vil tabellene som viser årlig utvikling for henholdsvis predikert boligprisvekt, historisk priser og treningssett presenteres i appendiks, og refereres til i hovedteksten. Fra Tabell A3.2 i appendiks fremgår det at gjennomsnittlig årlig prosentvis boligprisvekst predikert med Super Learner er 1.73%. Ettersom historisk årlig prisutvikling i gjennomsnitt har vært 1.18% er vekstdifferanse predikert til å være 0.55 prosentpoeng mellom predikerte kvadratmeterpriser med Bybanen og historiske boligpriser (uten Bybanen). Dette impliserer at boligprisveksten i gjennomsnitt vil være 46.6% større fra og med året Bybanen ferdigstilles og ni år frem i tid.

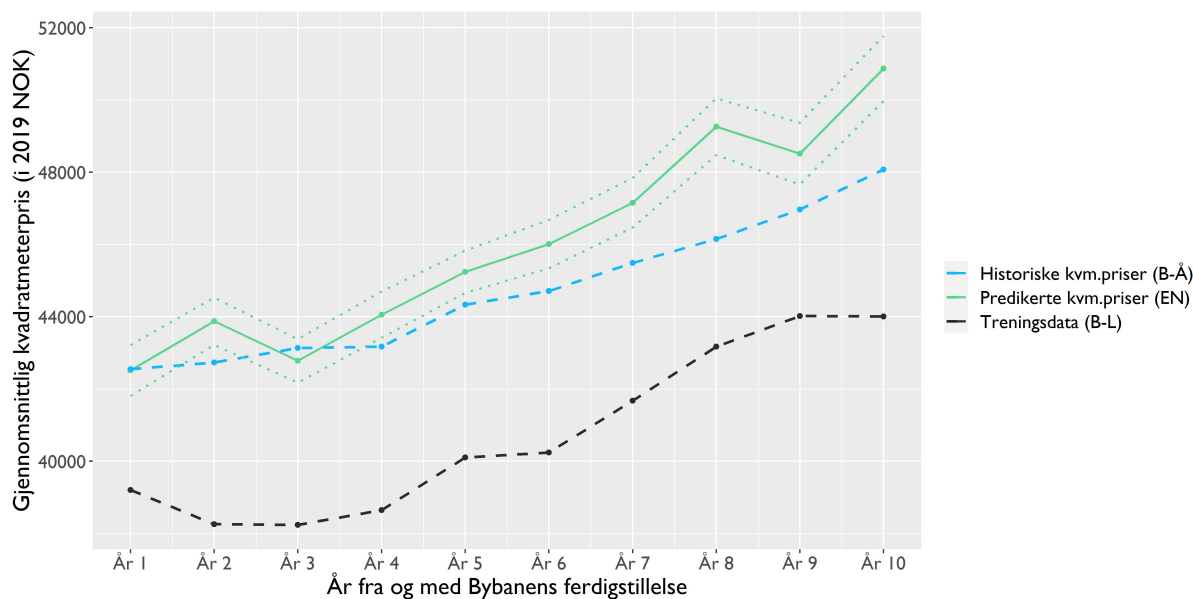
³⁰Denne prosessen kalles “pruning”, hvor kryssvalidering benyttes for å finne den optimale tilpasningen av antall knuter for å oppnå en modell som hverken er for fleksibel eller for statistisk.

Oppsummert predikerer henholdsvis Elastic Net og Super Learner en gjennomsnittlig årlig vekst i kvadratmeterpris på 1.82% og 1.73% fra og med Bybanens ferdigstillelse og ni år frem i tid for boliger lokalisert nærmere enn 3000 meter fra nærmeste bybanestopp rundt den kommende Åsane-traseen. På et aggregert nivå, foreligger det derfor muligheter for en økt avkastning rundt bybanetraseen til Åsane.

5.4.2 Alle boliger innen 1000 meter fra nærmeste bybanestopp

For å analysere hvorvidt avstand til Bybanen predikeres å være avgjørende for boligprisveksten rundt Åsane i tiden etter Bybanens inntreden, ønsker vi videre å se nærmere på prisutviklingen for alle boliger innen 1000 meter fra nærmeste bybanestopp.

Figur 5.9: Gjennomsnittet av predikerte boligpriser for alle boliger innen 1000 meter fra nærmeste Bybanestopp - Elastic Net

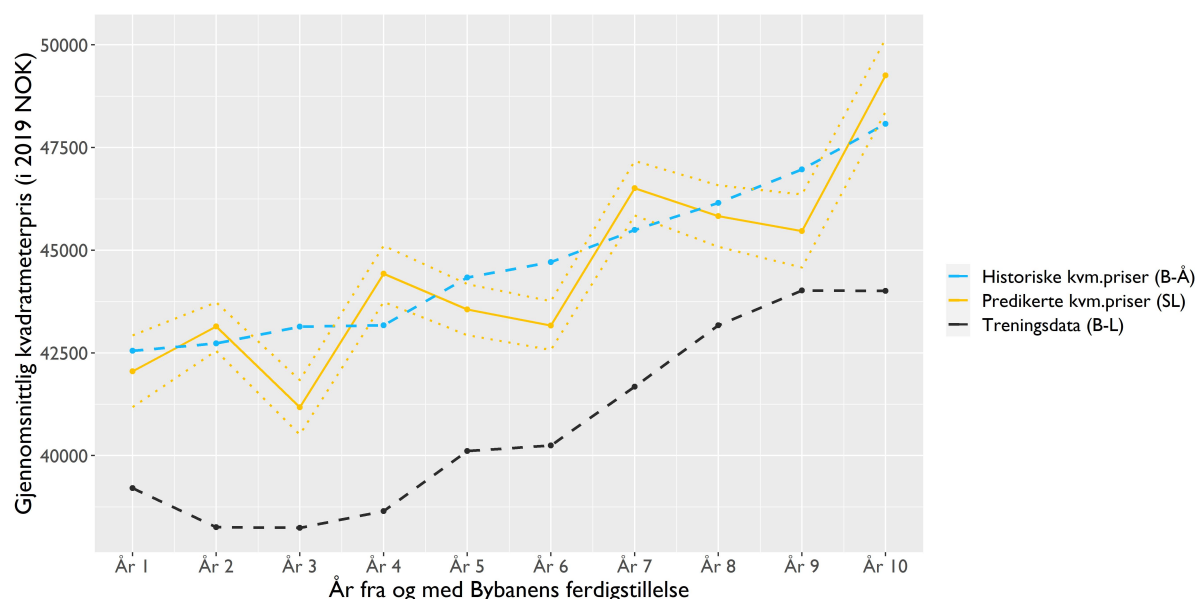


Den gjennomsnittlige årlige boligprisveksten som predikeres med Elastic Net, gitt Bybanens ferdigstillelse i 2010, er 2.04%. Historiske priser for boliger lokalisert innen 1000 meter fra nærmeste bybanestopp mellom Byparken og Åsane, har i midlertid økt med i gjennomsnitt 1.37% årlig i perioden 2010-2019. Disse tallene fremgår fra Tabell A3.6 i appendiks.

Når vi sammenligner historiske priser med predikert prisvekst, finner vi at Bybanen

vil forårsake en økt prisvekst på 0.67 prosentpoeng³¹ for boliger innen 1000 meter fra nærmeste bybanestopp. Gjennomsnittlige årlig boligprisvekst for boliger innen 1000 meter fra nærmeste bybanestopp predikeres videre å være 0.22 prosentpoeng³² høyere enn veksten som predikeres for alle boliger innen 3000 meter fra nærmeste Bybanestopp (1.82%) når Elastic Net benyttes for å predikere.

Figur 5.10: Gjennomsnittet av predikerte boligpriser for alle boliger innen 1000 meter fra nærmeste Bybanestopp - Super Learner



I likhet med hva vi observerte i kapittel 5.4.1 resulterer Super Learner også her i en mer volatile gjennomsnittlige predikerte boligpriser enn Elastic Net. For Super Learner er predikert vekst i gjennomsnitt 1.88% med Bybanen, noe som fremgår av Tabell A3.7. Dette indikerer at gjennomsnittet av predikerte kvadratmeterpriser med Bybanen vil være 0.51 prosentpoeng høyere enn historiske priser (1.37%) når Super Learner benyttes til å predikere. Prediksjonene fra Super Learner impliserer videre at boliger innen 1000 meter fra Bybanen vil oppleve en prisvekst som er 0.15 prosentpoeng høyere enn boliger innen 3000 meter fra Bybanen, gitt at Bybanen til Åsane var et faktum mellom 2010 og 2019. Oppsummert viser foreløpige resultater at boliger innen 1000 meter fra nærmeste bybanestopp vil oppleve en høyere prisvekst enn alle boliger innen 3000 meter fra nærmeste

³¹Dette er differansen mellom predikerte priser med Bybanen hvor Elastic Net benyttes og historiske boligpriser mellom Byparken og Åsane, derav $2.04\% - 1.37\% = 0.67$ prosentpoeng.

³² $2.04\% - 1.82\% = 0.22$ prosentpoeng.

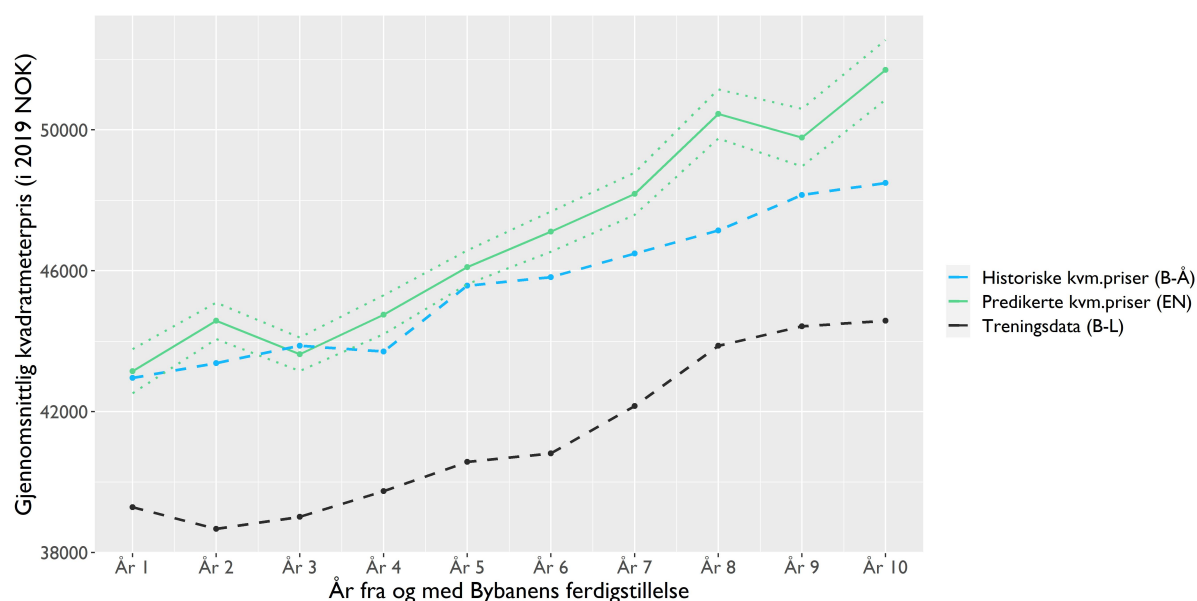
bybanestopp. Selv om Fredriksen (2013) analyserer prisdifferansen mellom boliger som henholdsvis ligger innen 100 og 1000 meter fra Bybanen, peker våre resultater i samme retning - at det potensielt eksisterer en høyere prisvekst jo nærmere boligen ligger Bybanen.

5.4.3 Leiligheter og eneboliger innen 1000 meter fra nærmeste bybanestopp

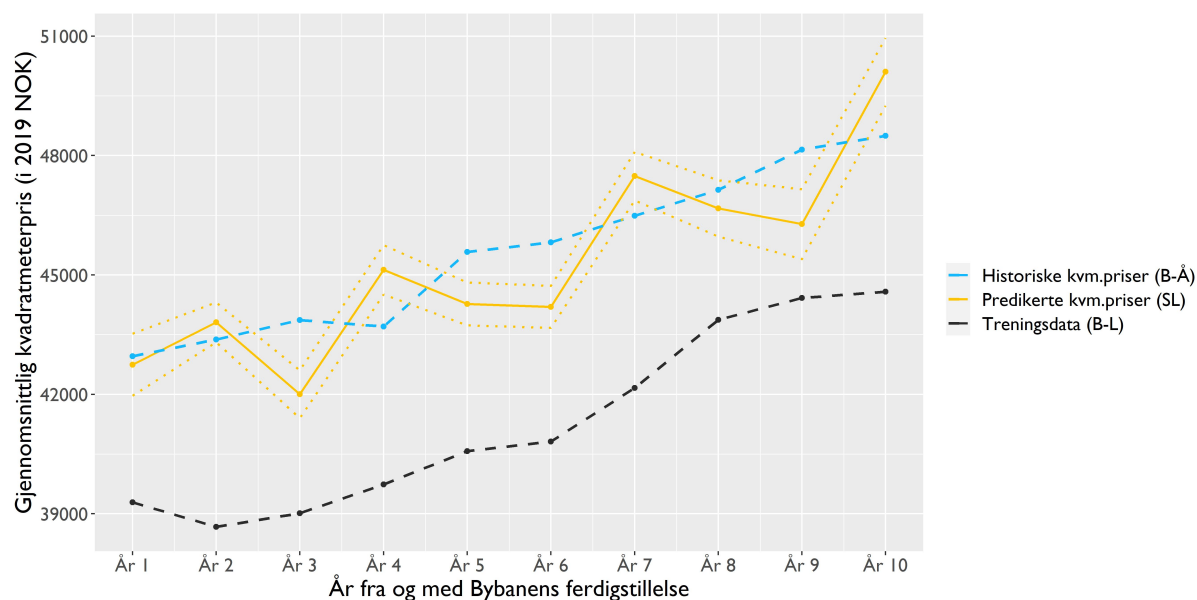
For investor står det sentralt å foreta risikovurderinger, samt utarbeide kostnadskalkyler for hver boligutbygging/ombygging som vurderes. Risikoen tilknyttet hver spesifikke boligutbygging/ombygging preges i stor grad av hvilke føringer bymiljøetaten og gjeldende KPA legger til rette for i det aktuelle område (Bergen Kommune, 2019b). For enkelte områder ilegger kommunen sterke føringer for hvilken bystruktur og bebyggelsestyper som ønskes. Fra samtaler med medlem i utvalg for miljø- og utvikling i SV, Jarle Brattespe, og medlem i utvalg for miljø- og byutvikling i Høyre, Charlotte Spurkeland, fremkommer det at Åsane bydel frem til nå har hatt en uheldig byutvikling, hvor handlesentre og boligbebyggelse ligger uønsket spredt. Planen er at Bybanen vil føre til en økt sentrering av bebyggelsen rundt den fremtidige trasseen, hvor en kombinasjon av ulike boligtyper er ønskelig.

Ettersom kommunen ønsker en kombinasjon av ulike boligtyper rundt den fremtidige Bybanen til Åsane, er det videre interessant å analysere hvorvidt det predikeres å foreligge systematiske forskjeller i prisvekst for ulike boligtyper i umiddelbar nærhet av Bybanen. Dette fordi potensiell prisvekst for ulike boligtyper må veies opp mot hvor krevende det er å få godkjent utbygging/ombygging av de ulike boligtypene. Derfor vil i det følgende predikere boligpriser for leiligheter og eneboliger innen 1000 meter fra nærmeste bybanestopp hver for seg.

Figur 5.11: Gjennomsnittet av predikerte boligpriser for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Elastic Net



Figur 5.12: Gjennomsnittet av predikerte boligpriser for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Super Learner



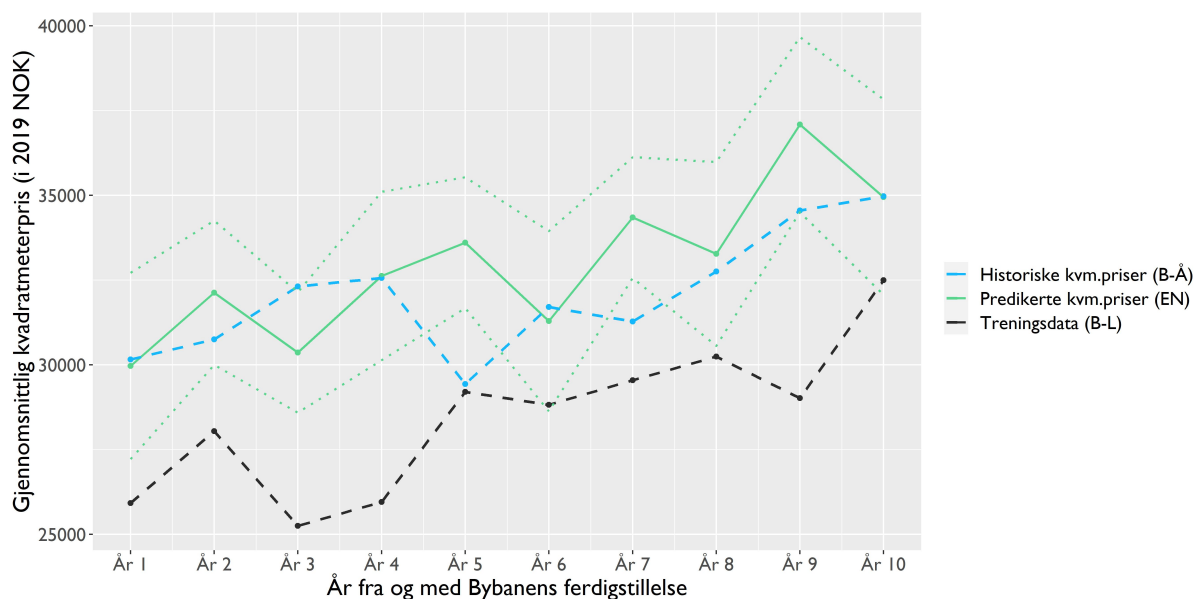
Dersom vi sammenligner prediksjonene presentert i Figur 5.9 med 5.11 og Figur 5.10 med 5.12, ser vi at utviklingen i prisnivå for alle boliger innen 1000 meter fra nærmeste bybanestopp er veldig lik prisveksten som predikeres for leiligheter innen 1000 meter fra nærmeste bybanestopp. At veksten er relativt lik kan sees tydeligere dersom vi

sammenligner tabellene som tallfester prosentvis vekst for de ulike prediksjonene. Dette innebærer en sammenstilling av Tabell A3.6 mot A3.10 og A3.7 mot A3.11. Grunnen til at vi får såpass like prediksjoner er trolig et resultat av at størsteparten av datagrunnlaget nært Bybanen består av leiligheter, noe vi ser fra Figur 3.4.

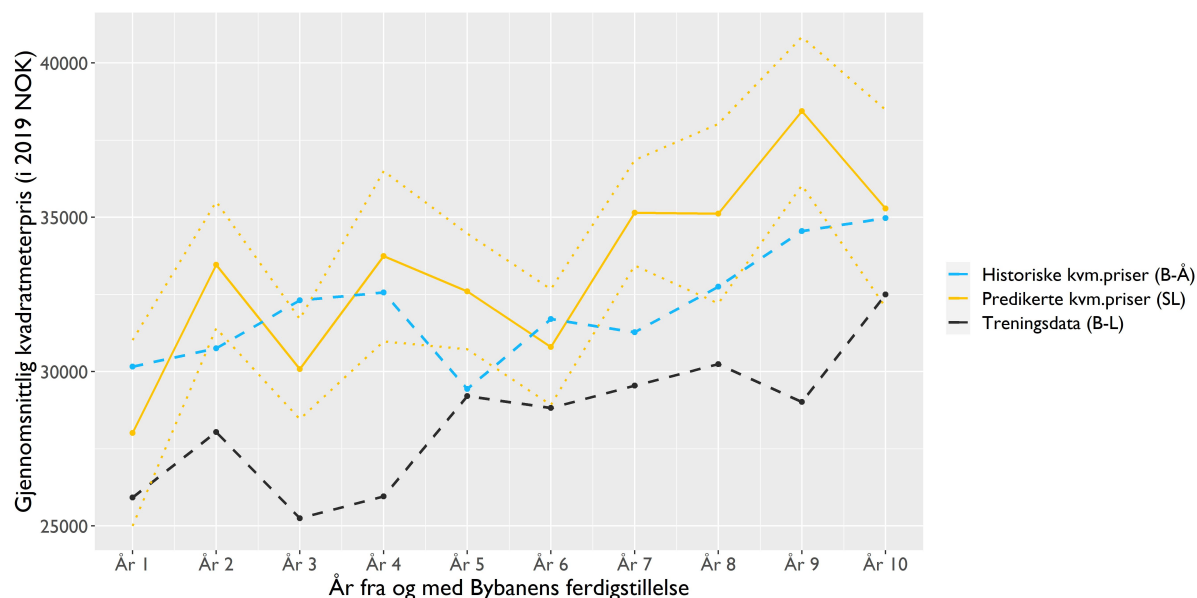
Elastic Net predikerer gjennomsnittlig prisveksten for leiligheter innen 1000 meter fra nærmeste bybanestopp til 2.05% med Bybanen, hvor Super Learner predikerer tilsvarende prisvekst til 1.88%. Historiske gjennomsnittspriser for de samme observasjonene er derimot 1.36% for perioden mellom 2010 og 2019. Dette antyder at Bybanens inntreden til Åsane predikeres å medføre en økt boligprisvekst på henholdsvis 0.69 og 0.52 prosentpoeng når Elastic Net og Super Learner benyttes for å predikere.

Figur 5.13 og 5.14 viser derimot at gjennomsnittlige prediksjoner for eneboliger innen 1000 meter fra nærmeste bybanestopp skiller seg ut fra det foregående prediksjonsutvalget.

Figur 5.13: Gjennomsnittet av predikerte boligpriser for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net



Figur 5.14: Gjennomsnittet av predikerte boligpriser for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Super Learner



Fra Tabell A3.14 og A3.15 i appendiks fremgår det at gjennomsnittlig predikert boligprisvekst med Bybanen for eneboliger innen 1000 meter fra Bybanen er henholdsvis 1.96% og 3.10% når Elastic Net og Super Learner benyttes til å predikere. Historisk gjennomsnittlig prisøkning for de samme observasjonene har derimot vært 1.78%. Studerer vi utviklingen i årlige gjennomsnittlige prediksjoner fremgår det at både Elastic Net og Super Learner virker til å produsere relativt volatile resultater. Dette illustreres også tydelig gjennom de grafiske fremstillingene over. Når Super Learner benyttes, svinger predikerte gjennomsnittlige boligpriser på det meste med 29.5% i løpet av to år. Usikkerheten i prediksjonene gjenspeiles i brede konfidensintervaller.

Likevel indikerer resultatene at prisnivået, gitt at Bybanen hadde eksistert i Åsane, gjennomgående predikeres til å vokse mer enn det historiske prisnivået. Årlig predikert gjennomsnittsvekst med Bybanen er henholdsvis 0.18 prosentpoeng³³ og 1.32 prosentpoeng³⁴ høyere enn historiske priser når Elastic Net og Super Learner benyttes. Disse resultatene fremgår fra Tabell A3.14 og A3.15 i appendiks. Ettersom det ble solgt relativt få eneboliger mellom Byparken og Lagunen i tidsrommet 2010-2019, er

³³Differansen mellom predikert årlig gjennomsnittlig vekst (1.96%) og gjennomsnittlig vekst i historiske priser (1.78%)

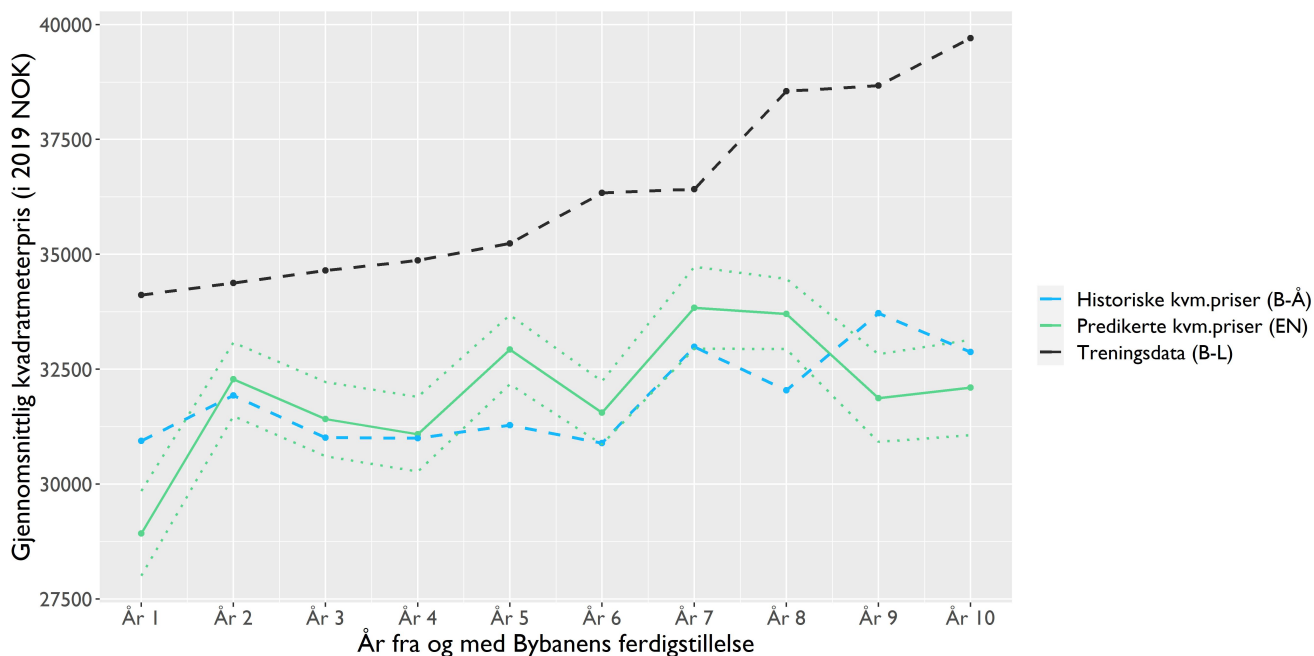
³⁴Differansen mellom predikert årlig gjennomsnittlig vekst (3.10%) og gjennomsnittlig vekst i historiske priser (1.78%)

prediksjonene som fremgår i Figur 5.13 og 5.14 mindre troverdige. Vi bør derfor være forsiktige med å trekke en slutning om predikert utvikling (Kim, 2009). Denne usikkerheten underbygges av at Elastic Net og Super Learner, som ellers har ganske sammenfallende gjennomsnittlig prisvekst, nå predikerer ganske ulikt.

5.4.4 Alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp

Frem til nå har vi sett at det predikeres å foreligge en høyere prisvekst for boliger som ligger innen 1000 meter fra Bybanen sammenlignet med boliger innen 3000 meter fra Bybanen, gitt Bybanens inntreden til Åsane. For å videre undersøke hvordan avstand til Bybanen påvirker boligprisvekst, ønsker vi avslutningsvis å predikere boligpriser for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp. De aktuelle prediksjonene illustreres i Figur 5.15 og 5.16.

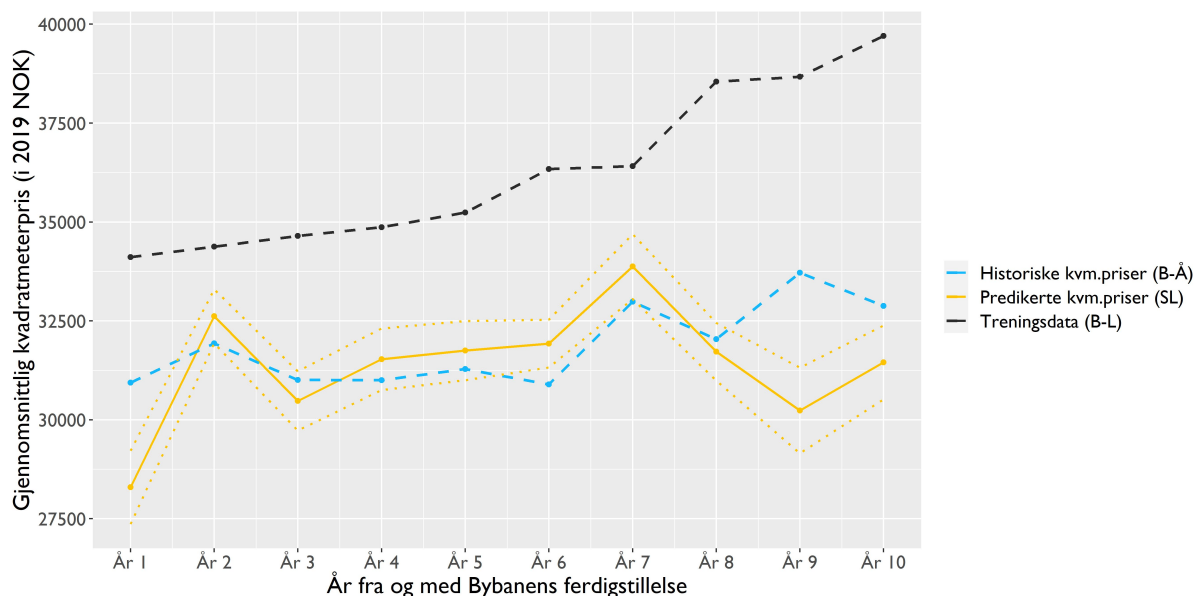
Figur 5.15: Gjennomsnittet av predikerte boligpriser for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net



Årlig gjennomsnittlig prisvekst med Bybanen predikeres å være 1.30% for alle boligtyper mellom 1000 og 3000 meter fra nærmeste bybanestopp. Historisk prisutvikling er derimot

0.73% i gjennomsnitt for de samme observasjonene³⁵. Dette impliserer at veksten i kvadratmeterpris med Bybanen vil være 0.57 prosentpoeng³⁶ høyere enn veksten uten.

Figur 5.16: Gjennomsnittet av predikerte boligpriser for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner



Gjennomsnittlig predikert boligprisvekst som predikeres med Super Learner er ganske lik veksten som predikeres når Elastic Net benyttes som prediksjonsmetode. Årlig vekst fra og med Bybanens inntreden til Åsane og ni år frem i tid er predikert til å være 1.39% i snitt, mot en historisk vekst på 0.73%. Dette tilsvarer en økt vekst på 0.66 prosentpoeng³⁷.

Dersom vi sammenligner predikert vekst i kvadratmeterpriser for boliger innen 1000 meter fra Bybanen med boligprisveksten som predikeres for boliger som er plassert mellom 1000 og 3000 meter fra nærmeste bybanestopp, ser vi at veksten er klart høyere for boliger innen 1000 meter fra Bybanen. Denne veksten predikeres å være henholdsvis 0.66 og 0.49 prosentpoeng³⁸ høyere for boliger innen 1000 meter fra Bybanen enn boliger mellom 1000 og 3000 meter fra Bybanen, når vi benytter Elastic Net og Super Learner.

For oversikten sin del er gjennomsnittlig årlig vekst sammenstilt i tabellen under for de ulike prediksjonene gjennomført i analysen.

³⁵Disse tallene er hentet fra Tabell A3.18 i appendiks

³⁶ $1.30\% - 0.73\% = 0.57$ prosentpoeng

³⁷ $1.39\% - 0.73\% = 0.66$ prosentpoeng

³⁸ $2.04\% - 1.38\% = 0.66$ prosentpoeng og $1.88\% - 1.39\% = 0.49$ prosentpoeng

Tabell 5.10: Oversikt over gjennomsnittlig boligprisvekst for vårt utvalg av prediksjoner - Elastic Net

| | Historisk vekst (Før Bybanen) | Predikert vekst (Etter Bybanen) | Differanse |
|----------------------------------|----------------------------------|------------------------------------|---------------|
| Innen 3000 meter (alle boliger) | 1.18% | 1.82% | +0.64 %-poeng |
| Innen 1000 meter (alle boliger) | 1.37% | 2.04% | +0.67 %-poeng |
| Innen 3000 meter (leiligheter) | 1.36% | 2.05% | +0.69 %-poeng |
| Innen 1000 meter (eneboliger) | 1.78% | 1.96% | +0.18 %-poeng |
| 1000 - 3000 meter (alle boliger) | 0.73% | 1.38% | +0.57 %-poeng |

Tabell 5.11: Oversikt over gjennomsnittlig boligprisvekst for vårt utvalg av prediksjoner - Super Learner

| | Historisk vekst (Før Bybanen) | Predikert vekst (Etter Bybanen) | Differanse |
|----------------------------------|----------------------------------|------------------------------------|---------------|
| Innen 3000 meter (alle boliger) | 1.18% | 1.73% | +0.55 %-poeng |
| Innen 1000 meter (alle boliger) | 1.37% | 1.88% | +0.51 %-poeng |
| Innen 1000 meter (leiligheter) | 1.36% | 1.88% | +0.52 %-poeng |
| Innen 1000 meter (eneboliger) | 1.78% | 3.10% | +1.32 %-poeng |
| 1000 - 3000 meter (alle boliger) | 0.73% | 1.39% | +0.66 %-poeng |

5.4.5 Kvantifisering av bybaneeffekten

For å gjennomføre en ytterligere vurdering av hvordan avstand til Bybanen predikeres å påvirke gjennomsnittlig kvadratmeterpris, ønsker vi å fremstille absoluttverdien for koeffisientene tilhørende dummyvariablene for *gåavstand til nærmeste bybanestopp*. Koeffisientene fremkommer fra prediksjonene av boligprisene for alle boligtyper innen 3000 meter fra nærmeste bybanestopp, modellert med Elastic Net. R-pakken *vip* (Greenwell, 2017), som står for Variable Importance Plots, muliggjør en grafisk representasjon av koeffisientene gjennom funksjonen `vi_model`³⁹.

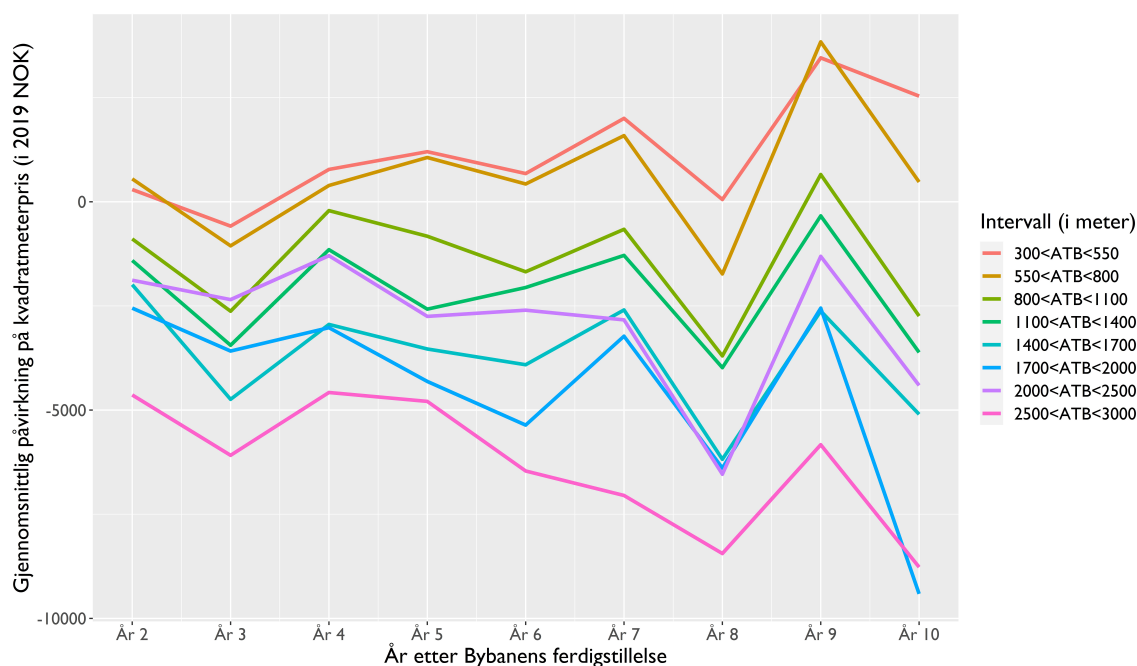
Fordelen med Elastic Net er at metoden muliggjør både regularisering og variabelreduksjon. Som tidligere forklart, medfører dette at modellen enten reduserer størrelsen på koeffisientestimatene tilhørende korrelerte variabler, eller utelukker variablene helt. Størrelsen på koeffisientene tolkes som endringen i avhengig variabel som skyldes én enhets endring i uavhengig variabel, gitt at alle andre variabler holdes konstant. Forutsetningen om at alle andre variabler holdes konstant er urimelig i vårt tilfelle. Dette er fordi vi

³⁹`vi_model` står for Model-specific variable importance

benytter en modell med relativt mange variabler. Fra korrelasjonsmatrisen (Figur 3.1) fremgår det at variabelen *gåavstand til nærmeste bybanestopp* er korrelert med flere inkluderte uavhengige variabler. Dette medfører at det er vanskelig å tro på at en endring i antall meter fra Bybanen er upåvirket av endringer i andre variabler. For eksempel ser vi at avstand til sentrum og avstand til nærmeste bybanestopp er relativt sterkt korrelert. Vi ser videre at ingen av de respektive variablene ekskluderes i prediksjonsmodellen som danner grunnlaget for å utlede variablenes viktighet med funksjonen *vi_model*. Dersom vi eksempelvis studerer en leilighet som ligger 500 meter nærmere et Bybanestopp enn en ellers identisk leilighet, vil det være vanskelig å vite om en eventuell prisdifferanse skyldes det faktum at leiligheten ligger nærmere Bybanen, eller om prisdifferansen fremgår som et resultat av at den både ligger nært Bybanen og sentrum.

Det er viktig å påpeke at analysens hovedformål er å bygge presise prediksjonsmodeller. Statistiske inferens er derfor ikke i fokus. Gjennom statistiske beregningsmetoder er det mulig å oppnå en tilnærmet rendyrket effekt av nærhet til Bybanen. Dette er imidlertid en omfattende oppgave med et annet fokus enn hva som vektlegges i denne utredningen. Likevel kan vi fremstille verdiene for koeffisientene for avstand til nærmeste bybanestopp grafisk for å få en indikasjon på hvor viktig modellen mener at variabelen er når boligpriser predikeres.

Figur 5.17: Utvikling av variable importance for alle dummy-avstander til Bybanen modellert med Elastic Net



Referansepunktet for verdiene som fremgår fra Figur 5.17 er boliger som ligger innen 300 meter fra nærmeste bybanestopp. Dette betyr at linjen som representerer boliger mellom 300 og 550 meter fra nærmeste bybanestopp i gjennomsnitt predikeres til å være høyere priset enn boliger som ligger nærmere enn 300 meter fra nærmeste bybanestopp. Resultatene tyder på at det er mer ettertraktet å bo mellom 300 og 800 meter fra Bybanen, enn å bosette seg i umiddelbar nærhet til Bybanen ($ATB > 300$ meter). Med unntak av boliger som ligger mellom 300 og 800 meter fra Bybanen, vil økt avstand til Bybanen henge sammen med fallende boligpriser, alt annet like. Koeffisientene for år åtte, ni og 10 er videre relativt volatile. Dette kan være forårsaket av at vi ikke har observasjoner fra Nesttun-Lagunen fra år åtte og utover. Dette kan påvirke prediksjonene i Åsane ettersom fant fant indikasjoner på at Byparken-Lagunen ville utgjøre det mest presise treningssettet når boligprisene i Åsane predikeres.

6 Diskusjon

I analysedelen utarbeidet vi et prediksjonsrammeverk som ble benyttet for å predikere boligprisene rundt den planlagte Åsane-trasseen for perioden 2010-2019, i et tenkt tilfelle hvor Bybanen til Åsane var ferdigstilt i 2010. Fra analysen fremgår det at boligens avstand til nærmeste bybanestopp predikeres å være en viktig faktor å hensynta ved investeringer i boligmarkedet rundt Åsane-traseen. I dette kapitlet vil vi først diskutere om funnene som fremgår fra prediksjonsrammeverkets fire steg også er relevant for andre modeller som søker å estimere eller predikere boligpriser. Deretter vil vi diskutere hvilken verdi det har for privatpersoner og investorer å ha informasjon vedrørende den predikerte boligprisveksten i Åsane med Bybanen. For å gi investorer og privatpersoner et mer helhetlig bilde av investeringsmulighetene rundt Åsane, ønsker vi videre å gjennomgå utvalgte politiske føringer tilknyttet utbygging og bystruktur for Åsane-området. Vi vil i korthet også diskutere relevante krav til boligutbygging rundt Bybanen og hvordan prediksjonsmodeller kan brukes som et supplerende verktøy når politiske beslutninger tilknyttet boligutbygging skal tas. Avslutningsvis vil vi diskutere prediksjonsmodellens svakheter, samt relevante emner for fremtidig forskning.

6.1 Prediksjonsrammeverkets relevans

Ifølge eiendomsmegler i Krogsvæn, Fredrik Søreide, bruker mange eiendomsmeglere i dag en estimeringsmodell utarbeidet av Eiendomsverdi AS som et hjelpemiddel for å verdivurdere boliger som legges ut for salg (F. Søreide, 20. mars 2020). Videre fremgår det fra samtaler med eiendomsmegler i W Eiendomsmegling, Christina Erstad, at meglere med mindre erfaring i stor grad støtter seg på slike modeller. Dette medfører at upresise estimater kan stå i fare for å resultere i mer unøyaktige verdivurderinger (C. Erstad, 15. mars 2020).

I boligmarkedet, hvor selger ikke ønsker å avsløre sin reservasjonspris og kjøper står overfor en usikkerhet tilknyttet boligens faktiske verdi, vil kjøper justere sin opplevde verdi gjennom et referansepunkt. Referansepunktet til kjøper avgjøres gjennom forventninger, som i sin tur påvirkes av verdivurderingen. Dette medfører derfor at verdivurderingen ikke bare er et verktøy for å vurdere uobserverbare karakteristikk ved en bolig. Den påvirker

også kjøpers forventede verdi og dermed prisen som faktisk blir betalt for boligen. (Kheyr et al., 2018) Det fremgår fra vår analyse, supplert med empiriske funn, at det er tydelig at Bybanens påvirkning på boligprisene er positiv. Dersom en slik effekt enten underestimeres, eller ikke medregnes i verdivurderingen i det hele tatt, kan selger derfor stå i fare for å selge boligen til en for lav pris.

Fra hva vi vet finnes det ingen estimeringsmodeller som inkorporerer hverken en bybanespesifikk variabel, eller en variabel som representerer den estimerte påvirkningen av en fremtidig utbygging av et offentlig gode, som Bybanen, når boliger verdivurderes. Modellen til Eiendomsverdi estimerer eksempelvis boligverdi basert på historiske boligtransaksjoner fra nærområde. Dette medfører høyere presisjon jo flere boligtransaksjoner det foreligger i estimeringsgrunnlaget. Dersom en bolig, som er lokalisert i nærheten av et bybanestopp, verdivurderes basert på 20 nærliggende boligtransaksjoner hvor kun én av transaksjonene er registrert i tiden etter Bybanens ferdigstillelse, vil modellen stå i fare for å underestimere betydningen av bybaneeffekten. Dette fordi modellen, ifølge modell- og analysedirektør fra Eiendomsverdi AS, Anders Lund, tillegger den ene observasjonen som er påvirket av Bybanen relativt liten vekt. Ettersom estimeringsmodellen til Eiendomsverdi hverken inneholder en bybanespesifikk variabel eller en variabel som hensyntar boligens avstand til sentrum, vil effekten av nærhet til både Bybanen og til sentrum stå i fare for å feilestimeres dersom det ikke foreligger et tilstrekkelig antall sammenlignbare objekter i estimeringsgrunnlaget, sier Lund.

På tross av at bybaneeffekten kan fanges opp indirekte gjennom andre inkluderte variabler, oppnådde vi i kapittel 5.2.4 en økt presisjon på 5.2% og 7.8% med henholdsvis Elastic Net og Super Learner når vi inkorporerte variabelen *gåavstand til nærmeste bybanestopp*, alt annet like. Dette indikerer at selv i de tilfellene hvor det foreligger et tilstrekkelig antall sammenlignbare observasjoner for boligen som verdivurderes, vil estimatene til Eiendomsverdi likevel stå i fare for å være mer upresise enn dersom en bybanespesifikk variabel hadde blitt inkludert.

Vi besitter derimot ikke det samme datagrunnlaget som Eiendomsverdi, og kan således ikke trekke en slutning om at en bybanespesifikk variabel vil føre til økt presisjon også i deres estimater. Det kan tenkes at deres modeller bygger på et utvalg av variabler som fanger opp variasjonen som skyldes Bybanen, og at bybaneeffekten derfor forklares

gjennom andre inkluderte variabler. Fra samtaler med Anders Lund fremkommer det at variabler som representerer avstand til både Bybanen og sentrum kan være av interesse for deres modeller, men at det ikke finnes en garanti for at variablene fører til høyere presisjon. I tilfeller hvor det er mangel på sammenlignbare observasjoner, er det derimot mer sannsynlig at det er fordelaktig å inkludere en bybanespesifikk variabel. I tillegg kan det også være verdifullt å se på muligheten for å inkorporere en variabel som hensyntar en *fremtidig* utbygging av et offentlig gode som Bybanen sin påvirkning på boligprisene. Som nevnt i kapittel 2 er informasjon om fremtidig inntjening og verdi også sentralt for prisen av et aktivum i dag.

Prediksjonsrammeverkets fremgangsmåten kan også være nyttig som et utgangspunkt når formålet er å predikere observasjoner i et område hvor det kun foreligger relevant treningsdata fra referanseprosjekter fra et annet geografiske områder. Dette kan være tilfelle hvor en ny bro, en ny tunnel eller annen sentral infrastruktur skal introduseres.

6.2 Prediksjonenes informasjonsverdi

Predikert boligprisvekst fra og med Bybanens ferdigstillelse til Åsane og ni år frem i tid, kan tilføre verdifull informasjon både for investorer og privatpersoner. Fra kapittel 5 predikeres det å foreligge muligheter for god avkastning ved å eie bolig i nærheten av bybanetraseen til Åsane. Resultatene indikerer at boliger som er plassert nært Bybanen vil oppleve en høyere prisvekst enn boliger lengre borte. Gjennom et illustrativt eksempel i kapittel 5.4.1 observerte vi at en tilsynelatende liten vekstdifferanse i boligpriser kan medføre høy avkastning på innskutt egenkapital over tid.

For privatpersoner og investorer som søker avkastning gjennom boligprisvekst alene, er informasjon om vekst i prisnivå sentralt - dersom det investeres i et område med høy boligprisvekst vil også avkastningen bli høy. For mange investorer er målet imidlertid å gjennomføre sine utbyggingsplaner så fort som mulig for så å reinvestere/ta utbytte. Selv om målet er en effektiv gjennomføring kan det i praksis ofte gå lang tid fra investor erverver en tomt, til et boligkompleks står klart til salg. Først skal tomten anskaffes, deretter skal utbyggingen planlegges og søkes om, så skal boligene bygges ut. For store prosjekter kan hele prosessen ta mange år. Fra mailkorrespondanse med en av salgslederne

i eiendomsselskapet Bonava⁴⁰, Jan Kåre Aurdal (J.K. Aurdal, 19.12.2020), fremgår det at selskapet ervervet en stor tomt like ved det kommende bybanestoppet Mindemyren i 2017. Det er nå godkjent utbygging av omtrent 700 leiligheter på tomten, hvor de siste boligene er estimert å være solgt i 2028. Dette er et eksempel på et utbyggingsprosjekt hvor det er relevant å ha informasjon vedrørende boligprisvekst fra tomten erverves til prosjektet avsluttes. Dersom boligprisene vil stige i perioden fra ervervelse til salg, vil ikke alternativkostnaden ved å binde opp kapital være like stor.

En kvantitativ vurdering av hvordan boligprisveksten vil endre seg i områder hvor Bybanen skal komme, kan dermed hjelpe investorer og privatpersoner til å fatte bedre beslutninger.

6.3 Politisk perspektiv og byutvikling rundt Åsane-traseen

For å formidle et mer helhetlig bilde av investeringsmulighetene rundt Åsane, ønsker vi videre å diskutere politiske aspekter rundt Bybanen til Åsane, samt hvilke føringer som vil gjelde for byutvikling rundt Åsane-traseen.

6.3.1 Usikkerhet rundt Bybanen og boligutbygging i Åsane-området

Siden Bybanen til Åsane ble vedtatt i 2016 har trasévalget rundt Bryggen vært et sentralt politisk ankepunkt. Opprinnelig skulle utbyggingen av Bybanen til Åsane vært påbegynt før utbyggingen av bybanetraseen til Fyllingsdalen. Grunnet sterke politiske uenigheter rundt trasévalg og finansiering, er ferdigstillelsen estimert forskjøvet fra 2024 til 2031 (Miljøløftet, 2020). Selv om Bybanen til Åsane er vedtatt, kan vedtaket endres av kommende bystyret, og ifølge medlem i utvalg for miljø og byutvikling i SV, Jarle Brattespe, kan politiske uenigheter i verste fall føre til at prosjektet skrinlegges. Usikkerheten tilknyttet prosjektet påfører både utbygger og investor økt risiko.

I tillegg til at det foreligger risiko tilknyttet Bybanens utbygging, foreligger det også interessekonflikter rundt byutviklingen i Åsane bydel. Området har ifølge Brattespe hatt

⁴⁰Bonava er et ledende selskap innen eiendomsutvikling i Nord-Europa, tungt investert i åtte ulike land (Bonava, 2020)

en uheldig utvikling hvor handlesentre og boligbebyggelse ligger mer spredt enn ønskelig. Dette medfører i neste omgang økt biltrafikk, noe som ifølge Brattespe er uønsket fra et byutviklingsperspektiv. Den uheldige utviklingen er delvis forårsaket av at grunneierne av de største eiendommene i kjernen av Åsane er bedre tjent med å drive handlesentre i stedet for boligutbygging. Dette vanskeliggjør en sentrering av boligbebyggelsen. Politikerne ønsker derimot at området skal preges av handelsvirksomhet på gateplan og et større offentlig rom, hvor bebyggelsen sentreres rundt den fremtidige Bybanen. (J. Brattespe, 15. august 2020)

Et annen utfordring med bystrukturen i Åsane er at det foreligger et overskudd av boliger sammenlignet med arbeidsplasser. Dette medfører at mange som bor i Åsane pendler til Ytrebygda for å jobbe. Brattespe påpeker at denne skjevfordelingen vil gjøre det lettere å få godkjent en utbygging i Fana og Ytrebygda enn i Åsane. Samtidig er et av målene med Bybanen å sentrere bebyggelsen rundt traseen, noe som underbygger viktigheten av å regulere for utbygging også i Åsane, sier tidligere byråd for finans, innovasjon og eiendom i Bergen, Håkon Pettersen.

Det er viktig at investor er klar over risikoen tilknyttet både utbyggingen av selve Bybanen til Åsane, men også det politiske perspektivet rundt boligutbygging tilknyttet Åsane-området.

6.3.2 Krav for utbygging rundt Bybanen

Fra samtaler med medlem i utvalg for miljø- og byutvikling i Høyre, Charlotte Spurkeland, fremkommer det at byrådets pålegging av rekkefølgekrav⁴¹ og anleggsbidrag⁴² for utbygger, gjennom områdets reguleringsbestemmelser, blitt mer brukt som et finansieringsverktøy i senere tid. Kravene kan kun stilles for prosjekter med behov for finansiering av offentlige inngrep innen samme planområde, og dets økonomiske implikasjon skal være av samme størrelsesorden som det økonomiske potensialet i utbyggingen.

Ettersom utbyggingen av Bybanen er kostbar, i tillegg til at den økonomiske verdien av å få lov til å bygge nært Bybanen er stor, kan kommunen stille større rekkefølgekrav og pålegge høyere anleggsbidrag for boliginvesteringer nært Bybanen, påpeker Spurkeland.

⁴¹Rekkefølgekrav er krav utbygger må innfri for å få godkjent sine prosjekter (Harborg, 2017)

⁴²Anleggsbidrag er økonomiske godtgjørelse utbygger må betale for å bidra finansielt til offentlige utbyggingstiltak innen det aktuelle planområdet (Regjeringen, 2014)

Eksempelvis ble utbygger Kanalveien 60 AS belastet et anleggsbidrag på 14.7 millioner kroner for å føre opp et boligkompleks like ved det kommende bybanestoppet Mindemyren. Dette tilsvarer omtrent 2000 kroner per kvadratmeter P-areal, noe som utgjør en betydelig sum sett i lys av gjennomsnittlige kvadratmeterpriser i Bergen bydel. Pengene var ment som et bidrag tilknyttet utbygging av infrastruktur i det aktuelle planområdet (Bergen Kommune, 2018).

Fra et politisk ståsted er det viktig at anleggsbidragenes størrelse og rekkefølgekravenes implikasjoner ikke forhindrer utbygging i områder hvor byutvikling er ønskelig. I en vurdering av kravenes størrelse, står et godt informasjonsgrunnlag om potensielle inntekter og kostnader sentralt. Hvis Bybanen kommer til Åsane vil predikerte boligpriser være til god hjelp for å minke usikkerheten tilknyttet inntektssiden. Dersom både politikerne og utbyggerne begge har en mer presis forståelse for hvordan boligprisene vil utvikle seg for det aktuelle området, vil det bli lettere å finne et krysningsspunkt som både fører til at utbygger er tjent med utbygging, samtidig som kommunen får dekket sitt finansielle behov.

Som nevnt kan anleggsbidragene og rekkefølgekravene kun stilles for prosjekter med behov for finansiering av offentlige inngrep innen samme planområdet. Dersom en utbygger eksempelvis planlegger å bygge et leilighetskompleks i et område hvor kommunen ikke har behov for offentlige inngrep, kan ikke utbygger kravsettes. Vårt prediksjonsrammeverk kan således være verdifult dersom utbygger står overfor et valg om enten å satse på utbygging i et område nært Bybanen, hvor et stort anleggsbidrag må medregnes, eller i et område lengre unna, hvor utbygger ikke vil bli belastet ekstra for utbyggingen. Det blir da lettere for investor å utarbeide mer presise lønnsomhetsvurderinger for ulike prosjekter. Det må påpekes at informasjon vedrørende rekkefølgekravenes og anleggsbidragenes størrelse ikke er tilgjengelig før en utbygging søkes om. Ifølge Spurkeland er det likevel ikke urimelig å anta at kravene vil være mer betydelige nært Bybanen.

6.4 Prediksjonsmodellens svakheter

En sentral svakhet med analysen vår er at vi ikke har et testsett for å vurdere prediksjonsmodellenes presisjon når vi predikerer boligprisene mellom Byparken og Vågsbotn fra og med Bybanens ferdigstillelse og ni år frem i tid. Selv om det foreligger

indikasjoner på modellpresisjon fra prediksjonsrammeverkets oppbygging, kan vi ikke tallfeste hvor presist modellene våre vil predikere boligpriser i Åsane.

Videre er modellene som predikerer boligpriser i Åsane med Bybanen trent og validert på transaksjonsverdier for boliger mellom Byparken og Lagunen. Dette fører til at modellen ikke fanger opp Åsane-spesifikke effekter når vi predikerer boligprisene i Åsane. I kapittel 5 søkte vi å løse problemet ved å finne et trenings- og valideringssett som i samsvarte godt med observasjonene vi ønsket å predikere i Åsane. Det er likevel viktig å påpeke at observasjoner mellom Byparken og Lagunen ikke vil replikere eksakt hva som observeres mellom Byparken og Åsane. Fremgangsmåten vil trolig føre til noe ustabile prediksjoner.

Modellutvalget vi benyttet i utredningen er videre. Selv om Super Learner tester for en rekke modeller simultant, kan det ikke utelukkes at det finnes modeller som vil oppnå høyere presisjon. Datasettet som benyttes for å trene og validere modellene er også begrenset. Vi har inkludert en del sentrale variabler fra den hedonistiske prisingsmodellen. Likevel vil variabler som etasje, antall rom, antall bad og nærhet til nærmeste hovedvei trolig medføre økt presisjon når boligpriser predikeres. Vi forsøkte å få tak i flere slike variabler, men denne informasjonen var ikke tilgjengelig fra hverken Eiendomsverdi, Finn eller andre aktuelle aktører sine databaser. For enkelte perioder foreligger det også relativt få observasjoner, spesielt for eneboliger, tomannsboliger og rekkehus. Dette kan svekke presisjonen. Det er også viktig å påpeke at utredningen vår kun presenterer predikerte boligpriser på et aggregert nivå. Hva som vil gjelde for hver enkelt bolig vil variere. Alle prediksjonsmodellene våre er imidlertid bygget opp på en måte som muliggjør prediksjoner på individnivå, men av hensyn til plass ikke fremlagt i denne utredningen.

I kapittel 5.2.2, hvor vi valgte ut trenings- og valideringssett, nevnte vi at boligprisene rundt Byparken-Nesttun i liten grad var påvirket av Bybaneeffekten i perioden 2001-2007. Dette på tross av at boligprisene ifølge hypotesen om effisiente markeder (Malkiel og Fama, 1970) vil påvirkes av bybaneeffekten allerede idet første byggetrinn ble vedtatt i år 2000. Dersom teorien stemmer også i vårt tilfellet, har vi valgt et datasett hvor observasjonene mellom Byparken og Nesttun som ikke kan antas å være upåvirket av bybaneeffekten. Det er flere momenter som taler for at vår avgrensning likevel er fornuftig. For det første foreligger det indikasjoner på at boligmarkedet på et generelt nivå ikke er effisient (Herath og Maier, 2015). I tillegg fremkommer det fra en samtale med J. Brattespe fra Venstre at

da Bybanen ble vedtatt i år 2000 var det stor usikkerhet tilknyttet hvor linjen skulle gå, og om den i det hele tatt ville komme. Dette understrekes av at Søreide fra Eiendomsmegler Krogsveen påpeker at det er først når selve utbyggingen av Bybanen starter at prisene potensielt vil endre seg. Vi anser det derfor ikke som urimelig å anta at boligprisene var mer eller mindre upåvirket av bybaneeffekten mellom 2001 og 2007. Ved å fjerne observasjoner fra år 2000 og 2008 unngår vi at treningsdata fanger opp den potensielle bybaneeffekten fra vedtaksåret og fra det året utbyggingen startet.

Sist, men ikke minst, er det tilnærmet umulig å vite hvordan Bybanens inntreden i Åsane vil påvirke boligprisene i området tilknyttet traseen. Vi argumenterte i kapittel 5.2.2 for at Byparken-Lagunen ville være det treningssettet som trolig ville fange opp den fremtidige bybaneeffekten i Åsane mest presist. Likevel er dette kun bygget på noen kvalifiserte antakelser. Tiden vil vise hvordan boligprisene faktisk vil utvikle seg.

6.5 Forslag til videre forskning

Valg av prediksjonsmetode vil avhenge av formålet med prediksjonene. I noen situasjoner er målet å oppnå høyest mulig «accuracy» (presisjon), hvor det andre ganger er sentralt å ha en modell med høy «interpretability» (transparens). Ved valg av metode vil det ofte foreligge en avveining mellom presisjon og transparens, hvor en modell som oppnår høy presisjon ofte vil være lite tolkbar (Shmueli et al., 2010). I vår utredning har fokuset imidlertid vært oppnåelse av høyest mulig presisjon. I kjølevannet av denne avveiningen finner vi at Super Learner utgjør den mest presise modellen når vi tester modellpresisjon gjennom de ulike stegene i prediksjonsrammeverkets oppbygging.

Vi har imidlertid liten kunnskap om hvordan algoritmen til Super Learner lærer, og derfor også lite informasjon om hvordan modellens uavhengige variabler predikeres å påvirke boligprisene. Dette kommer av at Super Learner faller under kategorien “black-box” algoritmer. Black-box algoritmer er en beskrivelse av modeller der input og output er kjent, men hvor virkemåten er ukjent (Nayak, 2019). Pakkene lime (Pedersen og Benesty, 2019) og shap (Molnar, 2020) muliggjør derimot at sluttbruker kan gå i dybden bak avanserte black-box algoritmer. Shapley Additive Explanations (SHAP) er et rammeverk av metoder som muliggjør en mer eksplisitt tolkning av hver enkelt boligprediksjon gjennom å tildele individuelle prediksjoner en verdi for deres forklaringskraft mot den

avhengige variabelen (Lundberg og Lee, 2017). Det finnes også metoder som muliggjør lokale forklaringer. Lokale forklaringer prøver å identifisere hvordan ulike uavhengige variabler påvirker en spesifikk prediksjon i modellen (Aas et al., 2019). Ved å ta i bruk en slik fremgangsmåte kan vi tallfeste hvordan de inkluderte variablene påvirker verdien av hvert spesifikke hus i Åsane-området. For eksempel kan vi få informasjon om hvordan modellene modellerer at avstand til nærmeste bybanestopp vil påvirke boligprisen av huset. Mer konkret informasjon om variablenes påvirkning på prisdannelsen av boliger er nyttig informasjon for investorer ettersom det gir dem et økt informasjonsgrunnlag.

En utfordring er imidlertid avhengighet mellom uavhengige variabler. Dette diskuterte vi i kapittel 5.4.5 hvor vi studerte hvordan boligprisene varierer med avstand til nærmeste bybanestopp gjennom et variable importance plot. Det finnes imidlertid fremgangsmåter som tar hånd om utfordringer tilknyttet avhengige variabler når vi ønsker å fremlegge forklaringer bak lokale prediksjoner. Dette gjøres gjennom en metode som heter Kernel SHAP (Aas et al., 2019). Metoden er imidlertid enda ikke utviklet hverken for Elastic Net eller Super Learner, men finnes for andre metoder som OLS og XGboost. En analyse av hvordan ulike modeller predikerer variablenes påvirkning på boligpriser, enten på et generelt nivå, eller i Åsane, ville vært et interessant tema for videre forskning.

I tillegg hadde det vært interessant å studert om modellen Eiendomsverdi benytter for å verdivurdere boliger ville estimert bedre dersom en bybanespesifikk variabel hadde vært inkludert i deres variabelgrunnlag. Vi tok derfor kontakt med Anders Lund fra Eiendomsverdi for å høre hvorfor ikke variablene var inkludert i deres modell som estimerer boligpriser for meglere. Han responderte med at dette var noe de ville se på i fremtiden.

7 Konklusjon

Gjennom denne utredningen har vi, ved å benytte maskinlæringsmetodene Elastic Net og Super Learner, analysert følgende problemstilling:

Hvordan ville utviklingen i boligprisene mellom Byparken og Åsane vært i perioden 2010-2019, i et tenkt scenario hvor Bybanen til Åsane hadde blitt ferdigstilt i 2010?

En sentral utfordring ved denne problemstillingen er at det ikke foreligger observasjoner mellom Byparken og Åsane som er påvirket av bybaneeffekten. Vi utarbeidet derfor et prediksjonsrammeverk, som gjennom fire steg, innhentet nok informasjon til å ta kvalifiserte avgjørelser om hvilken modell som trolig vil prestere best når vi ønsker å predikere boligprisene i Åsane med Bybanen. Innsikten fra prediksjonsrammeverket ble videre benyttet til å predikere boligprisene i Åsane gitt Bybanens ferdigstillelse i 2010 og ni år frem i tid. Disse sammenhengene ble testet ved å bruke transaksjonsdata fra Byparken til Lagunen, hvor Bybanen eksisterer, og mellom Byparken-Åsane, hvor Bybanen ikke eksisterer.

Fra prediksjonsrammeverket fremgår det at variablene som inkluderes i både Elastic Net og Super Learner ikke skal transformeres når vi predikerer boligprisene rundt Åsane-trasseen med Bybanen. Videre konkluderte vi med at boligtransaksjonene mellom Byparken og Lagunen vil utgjøre prediksjonsmodellens trenings- og valideringssett. Vi fant også indikasjoner på at bybaneeffekten er overførbart mellom to ulike geografiske områder. Gjennom det siste av prediksjonsrammeverkets fire steg finner vi indikasjoner på at det kan være sentralt å inkludere en bybanespesifikk variabel, som *avstand til nærmeste bybanestopp*, for å oppnå høyere presisjon når formålet er å predikere boligpriser i områder hvor et offentlig gode som Bybanen skal introduseres. For våre prediksjoner oppnår Elastic Net og Super Learner henholdsvis 5.2% og 7.8% laverer RMSE når variabelen inkluderes, alt annet like. Gjennom prediksjonsmodellens oppbygging fremgår det også som sentralt at Super Learner trolig vil utgjøre den mest presise modellen når vi predikerer boligprisene i Åsane fra og med Bybanens ferdigstillelse og ni år frem i tid.

Grunnlaget for denne antakelsen er for det første at Super Learner oppnår høyest presisjon når vi trener og validerer prediksjonsmodellene på et område hvor Bybanen foreligger til å predikere boligprisene i et annet område hvor Bybanen også foreligger. Dette impliserer

at metoden er mer presis enn Elastic Net når observasjoner som er påvirket av Bybanen benyttes for å predikere andre observasjoner som også er påvirket av Bybanen. For det andre er presisjonen også høyere når vi predikerer boligprisene rundt Åsane-traseen før Bybanens inntreden (2001-2007), trent og validert på observasjoner fra Byparken-Lagunen for perioden før Bybanen (2001-2007). Disse momentene indikerer at Super Learner er en mer presis metode enn Elastic Net når vi bruker et trenings- og valideringssett fra Byparken-Lagunen til å predikere boligprisene i Åsane med Bybanen.

Med bakgrunn i funnene fra prediksjonsrammeverkets fire steg, brukte vi Elastic Net og Super Learner, begge uten transformerte variabler, til å predikere utviklingen i boligprisene rundt bybanetraseen mellom Byparken og Vågsbotn fra 2010 til 2019, forutsatt at Bybanen til Åsane hadde vært ferdigstilt i 2010. Fra prediksjonsutvalget som blir presentert gjennom utredningen, fremkommer det at en gjennomsnittlig bolig som ligger innen 1000 meter fra nærmeste bybanestopp predikeres å oppleve en høyere prosentvis prisvekst enn boliger som ligger mellom 1000 og 3000 meter unna. Når vi sammenligner predikert prisvekst med Bybanen mot historisk prisvekst uten Bybanen, begge for Åsane-området, fremgår det at prediksjonene tyder på at introduksjonen av Bybanen til Åsane vil føre til en meravkastning for boliger mellom Byparken og Åsane. Ved første øyekast, kan imidlertid denne meravkastningen virke ubetydelig. Dersom vi hensyntar at veksten er inflasjonsjustert både etter konsumprisindeksen og en boligprisindeks for Bergensområdet, og legger til grunn at det ikke er uvanlig med en høy gjeldsgrad på, illustrerer vi at den predikerte veksten med Bybanen kan ha stor påvirkning på avkastning på innskutt egenkapital.

Ved å studere absoluttverdien for koeffisientene tilhørende dummyvariablene for gåavstand til nærmeste bybanestopp, modellert med Elastic Net, finner vi at *avstand til nærmeste bybanestopp* er sental når boligpriser tilknyttet Bybanen predikeres. Med unntak av boliger som ligger mellom 300 og 800 meter fra Bybanen, vil økt avstand til Bybanen henge sammen med fallende boligpriser, alt annet like. Det er viktig å påpeke at grunnlaget for å hevde *inferens* er svak i denne sammenheng. Funnene bevitner derfor kun *indikasjoner*.

Utrednings funn kan ha verdi i ulike sammenhenger. For det første kan fremgangsmåten fra prediksjonsrammeverket være til hjelp når formålet er å predikere observasjoner i et område hvor det kun foreligger relevant informasjon fra referanseprosjekter i et

annet geografiske områder. Dette kan være tilfelle hvor en ny bro, en ny tunnel eller annen sentral infrastruktur skal introduseres. Funnene våre tyder også på at det en bybanespesifikk variabel, og/eller en variabel som representerer den estimerte påvirkningen av en fremtidig utbygging av et offentlig gode, som Bybanen, kan medføre økt presisjon for andre estimeringsmodeller hvor formålet er verdivurdering av bolig. Dette er spesielt aktuelt i de tilfellene hvor estimeringsmodellen bygger på et datagrunnlag hvor få av observasjonene er påvirket av bybaneeffekten. Sist, men ikke minst er det aktuelt for både privatpersoner, investorer og politikere å ha kunnskap om hvordan Bybanen vil påvirke boligprisene i Åsane etter dets inntreden. Et økt informasjonsgrunnlag kan medføre bedre investeringer og mer korrekt politiske avgjørelser. Selv om vi predikerer historiske boligpriser med en bybaneeffekt, er endring i prisnivå fremdeles relevant også for fremtiden.

Referanser

- Aas, K., Jullum, M., og Løland, A. (2019). Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *arXiv preprint arXiv:1903.10464*.
- Aliyu, A. A., Kemiki, O. A., og Bello, M. U. (2018). Analysis of current empirical studies on transport value-added effect and proximate housing price capture. *Traektoriâ Nauki= Path of Science*, 4(12).
- Bergen Kommune (2016). Levekår og helse i bergen. *Levekårsrapport 2016*. Hentet 10. mars 2020 fra: <https://www.bergen.kommune.no/publisering/api/filer/T537114832>.
- Bergen Kommune (2018). Utbyggingsavtale mindemyren. Hentet 01. desember 2020 fra: <https://www.bergen.kommune.no/hvaskjer/kunngjoringer/mindemyren>.
- Bergen Kommune (2019a). Bybanens fem byggetrinn. Hentet 12. september 2020 fra: <https://www.bergen.kommune.no/hvaskjer/tema/reguleringsplaner-for-bybanen/bybanens-fem-byggetrinn>.
- Bergen Kommune (2019b). Kommuneplanens arealdel. Hentet 12. september 2020 fra: <https://www.bergen.kommune.no/publisering/api/filer/T537387153>.
- Bergen Kommune: Byutvikling/Plan- og bygningsetaten (2020). Bybanen fra sentrum til Åsane. orientering om planarbeidet. saksnr.: 201807496/206 (ds1). sider 12–13. Hentet 10. desember 2020 fra: <https://www.bergen.kommune.no/publisering/api/filer/T543359068>.
- Bonava (2020). This is bonava. Hentet 15. mars 2020 fra: <https://www.bonava.com/en/about-us/bonava>.
- Boorse, J. W. (2001). This is light rail transit. Technical Report (No. E-C033).
- Diamond, P., Shafir, E., og Tversky, A. (1997). Money illusion. *Quarterly Journal of Economics*, 112(2):341–374.
- Diaz, R. B. og Mclean, V. (1999). Impacts of rail transit on property values. I *American Public Transit Association Rapid Transit Conference Proceedings*, sider 1–8.
- Donovan, S., Genter, J., Petrenas, B., Mumby, N., Hazledine, T., Litman, T., Hewison, G., Guidera, T., O'Reilly, L., Green, A., et al. (2008). Managing transport challenges when oil prices rise. *New Zealand Transport Agency Research Report 04*, 8:148pp.
- Dorantes, L. M., Paez, A., og Vassallo, J. M. (2011). Analysis of house prices to assess economic impacts of new public transport infrastructure: Madrid metro line 12. *Transportation research record*, 2245(1):131–139.
- Eiendomsverdi (2020a). Eiendomsinformasjon satt i system? Hentet 2. februar 2020 fra: <https://eiendomsverdi.no/Default.aspx>.
- Eiendomsverdi (2020b). Hvordan påvirker avstand til stasjonen boligprisene i pendlerområder? Hentet 2. februar 2020 fra: <https://eiendomnorge.no/aktuelt/blogg/hvordan-pavirker-avstand-til-stasjonen-boligprisene-i-pendleromrader>.
- Finanstilsynet (2020). Vurdering av boliglånsforskriften og forbrukslånsforskriften.

- Hentet 8. desember 2020 fra: <https://www.regjeringen.no/contentassets/ce0df804bb2e4948a83d95476822428a/horingsnotat-2338194.pdf>.
- Fortmann-Roe, S. (2012). Accurately measuring model prediction error. *Online: <http://scott.fortmann-roe.com/docs/MeasuringError.html>*.
- Fredriksen, K. M. R. (2013). Bybanens påvirkning på boligpriser i bergen. Master's thesis, Universitetet i Agder; University of Agder.
- Friedman, J., Hastie, T., og Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1(4). Hentet 10. april 2020 fra: <https://cran.r-project.org/web/packages/glmnet/index.html>.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, sider 1–67.
- Gallo, M. (2018). The impact of urban transit systems on property values: A model and some evidences from the city of naples. *Journal of Advanced Transportation*, 2018.
- Golub, A., Guhathakurta, S., og Sollaapuram, B. (2012). Spatial and temporal capitalization effects of light rail in phoenix: From conception, planning, and construction to operation. *Journal of Planning Education and Research*, 32(4):415–429.
- Greenwell, B. (2017). vip: Variable importance plots. Hentet 10. april 2020 fra: <https://cran.r-project.org/web/packages/vip/index.html>.
- Harborg, E. (2017). Kommuners praksis rundt bruk av utbyggingsavtaler-en studie om kostnadsfordeling av rekkefølgebestemmelser.
- Hastie, T., Tibshirani, R., og Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Herath, S. og Maier, G. (2015). Informational efficiency of the real estate market: A meta-analysis.
- Hoerl, A. E. og Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Hyndman, R. (2013). The difference between prediction intervals and confidence intervals. Hentet 10.april 2020 fra: <https://robjhyndman.com/hyndsight/intervals/>.
- Hyndman, R. J. og Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Iversen, K. (2019). Hvordan setter eiendomsmegler riktig prisantydning på din bolig? Hentet 2. februar 2020 fra: <https://dnbeiendom.no/altombolig/kjop-og-salg/tips-til-selgere/hvordan-fa-verdivurdering-av-bolig>.
- James, G., Witten, D., Hastie, T., og Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Jones, W. (2015). Transport infrastructure and value uplift. Technical report, Tech. rept. Department of Infrastructure and Regional Development, Australian Government.
- Jung, J.-K., Patnam, M., og Ter-Martirosyan, A. (2018). *An Algorithmic Crystal Ball: Forecasts-based on Machine Learning*. International Monetary Fund.

- Khezr, P., Ahmad, S., et al. (2018). Anchoring in the housing market: Evidence from sydney. Technical report.
- Kim, S.-Y. (2009). Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC bioinformatics*, 10(1):147.
- Lekve, M. (2013). Tilgang til infrastruktur øker boligprisene: Nærhet til bybanen «verdt» 300.000 kroner boligprisene. Hentet 10. desember 2020 fra: <https://e24.no/privatoekonomi/i/EWBE83/tilgang-til-infrastruktur-oeker-boligprisene-naerhet-til-bybanen-verdt-300000-kroner>.
- Lundberg, S. M. og Lee, S.-I. (2017). A unified approach to interpreting model predictions. sider 4765–4774.
- Malkiel, B. G. og Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417.
- Matre, E. L. (2019). Se film som viser vedtatt bybanetrasé til Åsane. Hentet 1. februar 2020 fra: <https://www.bergen.kommune.no/hvaskjer/bymiljo/se-film-som-viser-vedtatt-bybanetras-til-asane>.
- Miljøløftet (2020). Bakgrunn, vedtak og fremdrift. Hentet 5. mai 2020 fra: <https://xn--miljloftet-o8ab.no/bybanen-til-asane2/bybanen-til-asane/>.
- Molnar, C. (2020). Interpretable machine learning. *The Comprehensive R Archive Network*. Hentet 12. september 2020 fra: <https://cran.r-project.org/web/packages/iml/vignettes/intro.html>.
- Moody, J. (2019). Accurately measuring model prediction error. *Towards Data Science*. Hentet 25. februar 2020 fra: <https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e#:~:text=Sep>.
- Murray, C. (2016). Land value uplift from light rail. *Available at SSRN 2834855*.
- Nayak, A. (2019). Idea behind lime and shap. *Towards Data Science*. Hentet 12. september 2020 fra: <https://towardsdatascience.com/idea-behind-lime-and-shap-b603d35d34eb>.
- Norconsult og Viak, A. (2020). Oppsummering av skissefasen. sider 10–15.
- Norges Bank (2020). Rentestatistikk. Hentet 15. mars 2020 fra: <https://www.norges-bank.no/tema/Statistikk/Rentestatistikk/>.
- Oleszak, M. (2019). Regularization: Ridge, lasso and elastic net. Hentet 3. mars 2020 fra: <https://www.datacamp.com/community/tutorials/tutorial-ridge-lasso-elastic-net>.
- Osland, L. (2001). Den hedonistiske metoden og estimering av attributtpriser. *Norsk økonomisk tidsskrift*, 115(1):1–22.
- Pedersen, T. L. og Benesty, M. (2019). Understanding lime. *The Comprehensive R Archive Network*. Hentet 12. september 2020 fra: https://cran.r-project.org/web/packages/lime/vignettes/Understanding_lime.html.
- Polley, E. og van der Laan, M. S. (2017). Super learner prediction: R package, version 2.0-21. Hentet 10. april 2020 fra: <https://cran.r-project.org/web/packages/SuperLearner/index.html>.

- R core team (2020). The r project for statistical computing. Hentet 15. mars 2020 fra: <https://www.r-project.org/>.
- Regjeringen (2014). §§ 17-3 og 11-9 nr. 4 - lovligheten av bruk av utbyggingsavtaler (§ 17-3) - rekkefølgekrav i kommuneplan og plan- og bygningsloven (§ 11-9 nr. 4).
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229.
- Seo, K. (2016). *Impacts of transportation investment on real property values: An analysis with spatial hedonic price models*. Arizona State University.
- Shmueli, G. et al. (2010). To explain or to predict? *Statistical science*, 25(3):289–310.
- Skyss (2020a). Reiseplanlegger. Hentet 10. desember 2020 fra: <https://reise.skyss.no/planner/travel-plans>.
- Skyss (2020b). Årsrapport 2019. Hentet 15. mars 2020 fra: <https://www.skyss.no/globalassets/strategiar-og-fagstoff/strategiar-og-handlingsprogram/arsrapport/arsrapport-kollektivstrategi-2019.pdf>.
- Srour, I. M., Kockelman, K. M., og Dunn, T. P. (2002). Accessibility indices: Connection to residential land prices and location choices. *Transportation Research Record*, 1805(1):25–34.
- SSB (2017). Hva er egentlig BNP? Hentet fra: <https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/hva-er-egentlig-bnp>.
- SSB (2020a). Konsumprisindeksen. Hentet 22. februar 2020 fra: <https://www.ssb.no/kpi>.
- SSB (2020b). Prisindeks for brukte boliger. Hentet 22. februar 2020 fra: <https://www.ssb.no/statbank/table/07221/>.
- Steinert, M. og Crowe, S. (2001). Global real estate investment: Characteristics, optimal portfolio allocation and future trends. *Pacific Rim Property Research Journal*, 7(4):223–239.
- Sutton, G. D., Mihaljek, D., og Subelyte, A. (2017). Interest rates and house prices in the united states and around the world.
- Tayo, B. O. (2019). Simplicity vs complexity in machine learning — finding the right balance. *Towards Data Science*. Hentet 2. mars 2020 fra: <https://towardsdatascience.com/simplicity-vs-complexity-in-machine-learning-finding-the-right-balance-c9000d1726fb>.
- Tibshirani, R. (1996). Regression selection and shrinkage via the lasso. *Journal of the Royal Statistical Society Series B*, 58(1):267–288.
- Van der Laan, M. J., Polley, E. C., og Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, 6(1).
- Vegard, K. E. (2018). Dette bruker nordmenn penger på. Hentet 10. juni 2020 fra: <https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/dette-bruker-nordmenn-penger-pa>.

Zou, H. og Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Appendiks

A1 Bakgrunnsinformasjon

Figur A1.1: Fremstilling av omtrentlig gåavstand fra Byparken.



| Sone | Sonegrense (i meter) | Sone | Sonegrense (i meter) |
|------|----------------------|------|----------------------|
| 1 | 900 | 6 | 5600 |
| 2 | 1500 | 7 | 6800 |
| 3 | 2400 | 8 | 8300 |
| 4 | 3400 | 9 | 10000 |
| 5 | 4500 | 10 | 12000 |

Tabell A1.1: Bybanenstráséenes egenskaper (Skyss, 2020a)

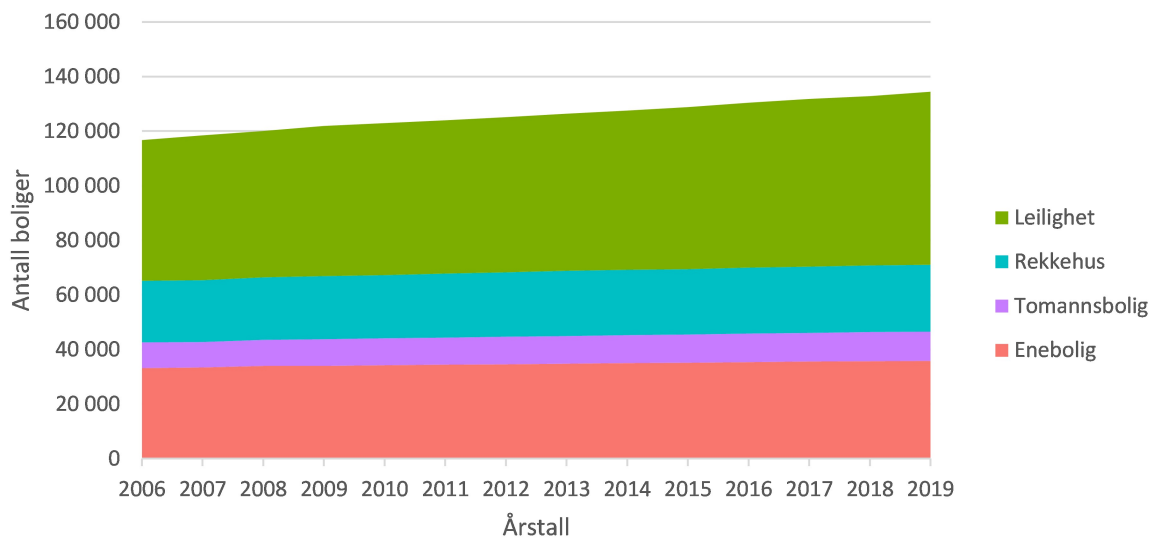
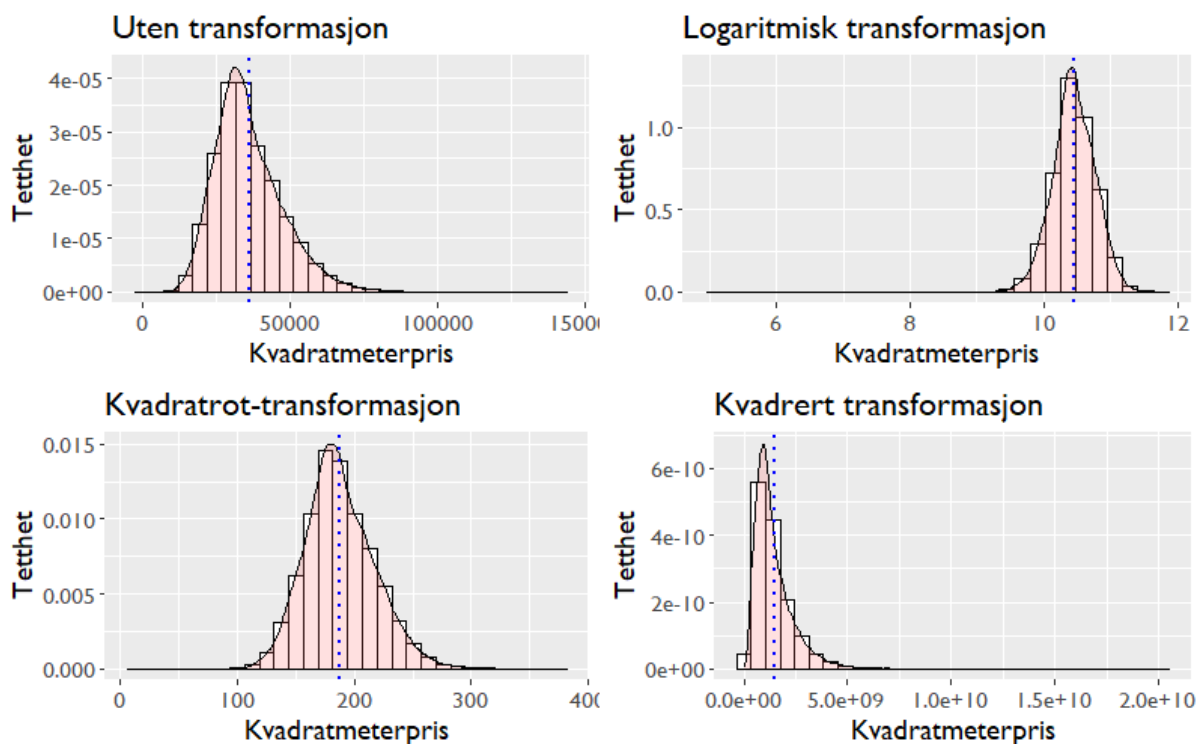
| Bybanetrásé | Lengde | Antall stopp | Reisetid |
|--------------------|---------------|---------------------|-----------------|
| Byparken-Nesttun | 9.8 km | 15 stopp | 23 min |
| Nesttun-Lagunen | 3.6 km | 5 stopp | 8 min |
| Lagunen-Flesland | 7.0 km | 7 stopp | 20 min |
| Byparken-Vågsbotn | 12.8 km | 14 stopp | ca 30 min |

Informasjonen vedrørende lengde, antall bybanestopp og reisetid for traseen til Åsane er ikke fastsatt. Tallene er hentet fra fagrappport utarbeidet av Bergen Kommune (Bergen Kommune: Byutvikling/Plan- og bygningsetaten, 2020)

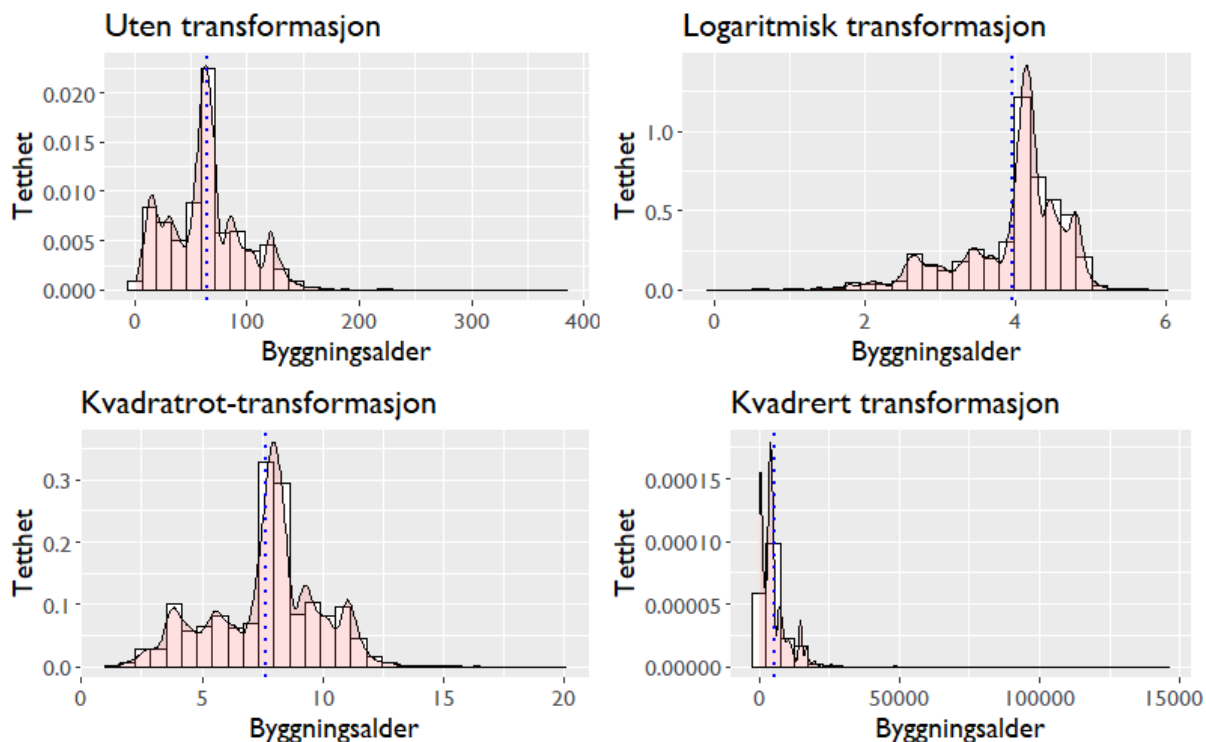
A2 Variabler

Tabell A2.1: Variabelliste

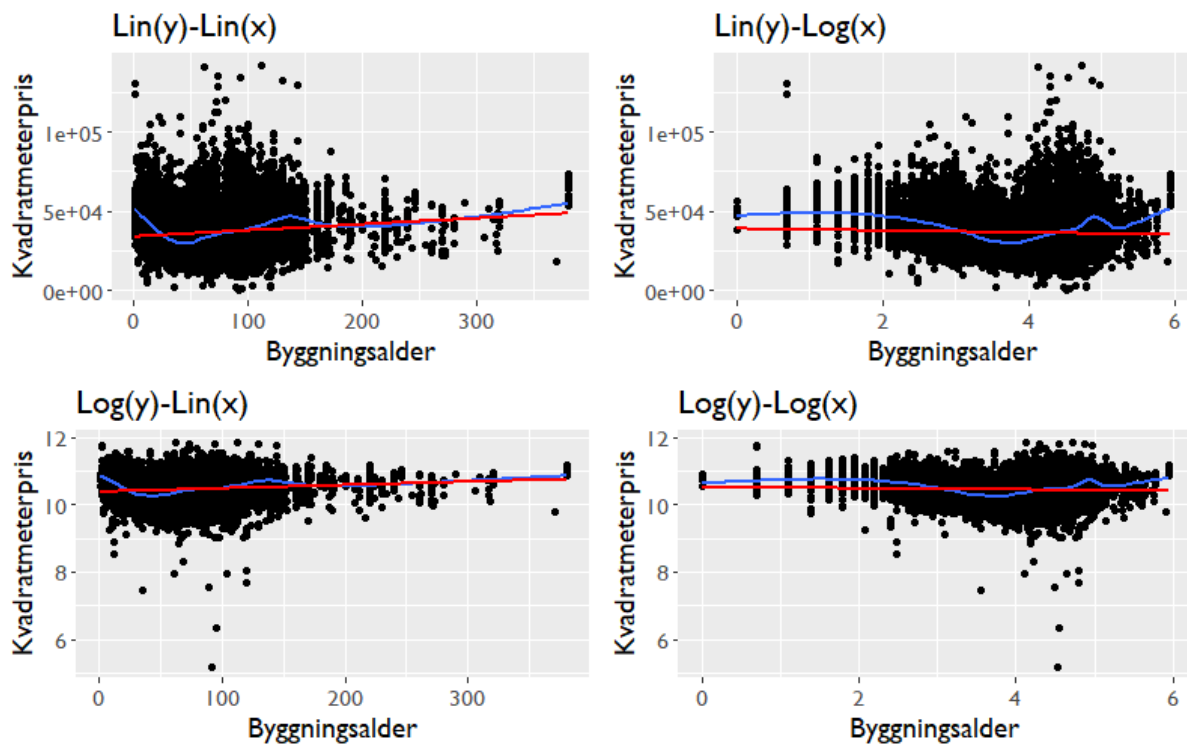
| Responsvariabel | Datatype | Enhet | Variabeltype | Kilde |
|--|------------------------|----------------------|---------------------|------------------|
| Kvadraterpris | Numerisk kontinuertlig | NOK | Pris | Eiendomsverdi AS |
| Forklaringsvariabler | Datatype | Enhet | Variabeltype | Kilde |
| Bygningsalder | Numerisk diskret | Ar | Strukturell | Eiendomsverdi AS |
| Boligtype | Kategorisk nominal | Faktor med fire nivå | Strukturell | Eiendomsverdi AS |
| Bruttoareal (BTA) | Numerisk diskret | Kvadratmeter | Strukturell | Eiendomsverdi AS |
| Salgstid | Numerisk diskret | Døgn | Tidsmessig | Eiendomsverdi AS |
| Nærhet til interessepunkt | Kategorisk nominal | Ja/Nei | Lokasjon | Google Maps |
| Obligasjonsrenter | Numerisk diskret | Prosent | Økonomisk | SSB |
| Styringsrenter | Numerisk diskret | Prosent | Økonomisk | SSB |
| Endring i BNP | Numerisk diskret | Prosent | Økonomisk | SSB |
| Invbyggere med innvandrerbakgrunn (Lk ⁴³ 1) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Barneutflyttinger 0-5 år (Lk 2) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Barn med enslige foreldre (Lk 3) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Median inntekt | Numerisk kontinuertlig | NOK | Miljømessig | Bergen Kommune |
| Barn i lavinntektsusholdninger (Lk 4) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Ung uførepensjonister (Lk 5) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Barn med barnevernstiltak (Lk 6) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Andel med dårlig norsk språk (Lk 7) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Kommunale utleieboliger (Lk 8) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Anmeldelser (Lk 9) | Numerisk kontinuertlig | Per 1000 innbyggere | Miljømessig | Bergen Kommune |
| Avstand til Bybanen (ATB) | Kategorisk ordinal | Faktor med 13 nivåer | Lokasjon | Google Maps |
| Avstand til sentrum (ATS) | Kategorisk ordinal | Faktor med 9 nivåer | Lokasjon | Google Maps |

Figur A2.1: Vekst av antall boliger per boligtype (2006-2019)**Figur A2.2:** Fordeling - kvadratmeterpris med og uten transformasjoner

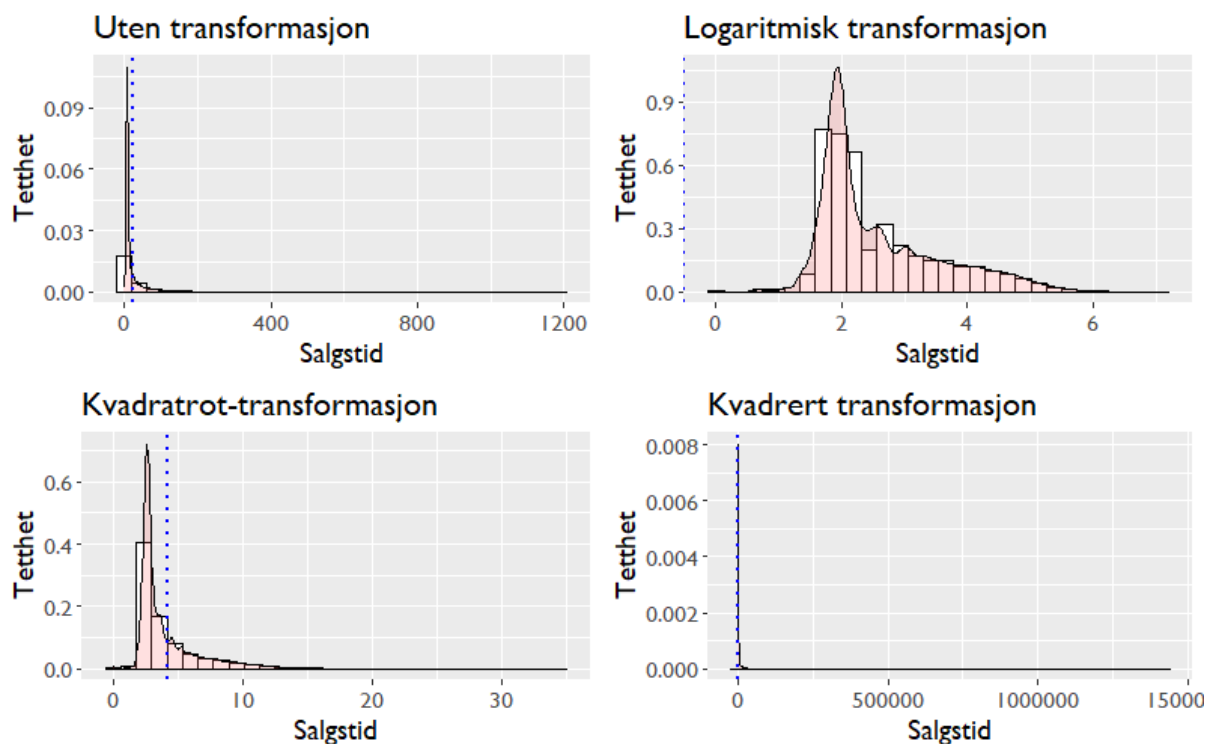
Figur A2.3: Fordeling - bygningsalder med og uten transformasjoner



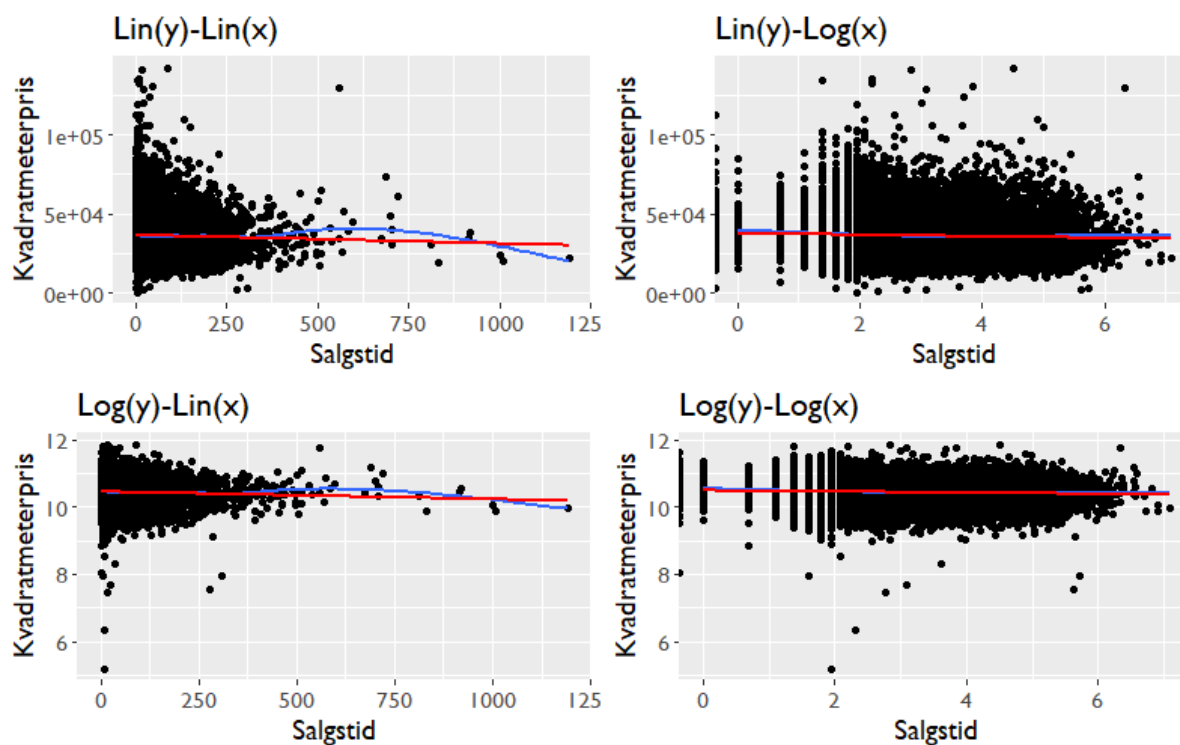
Figur A2.4: Scatter plot - bygningsalder med og uten transformasjoner



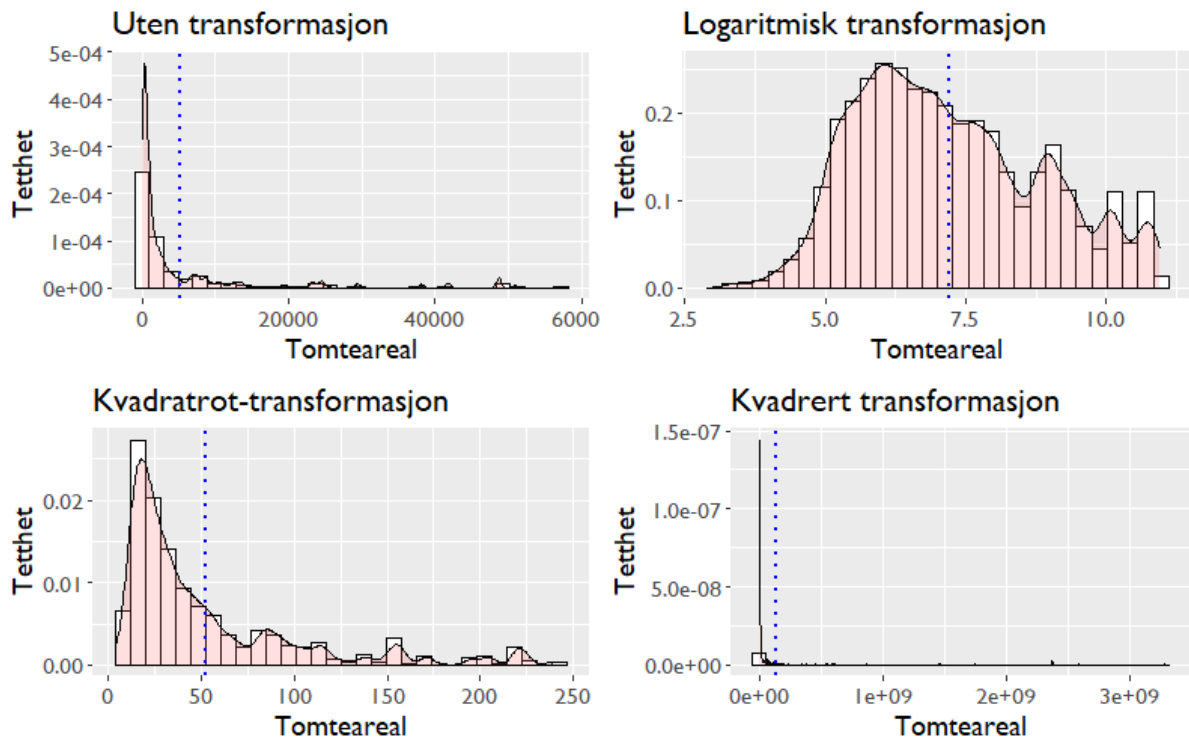
Figur A2.5: Fordeling - salgstid med og uten transformasjoner



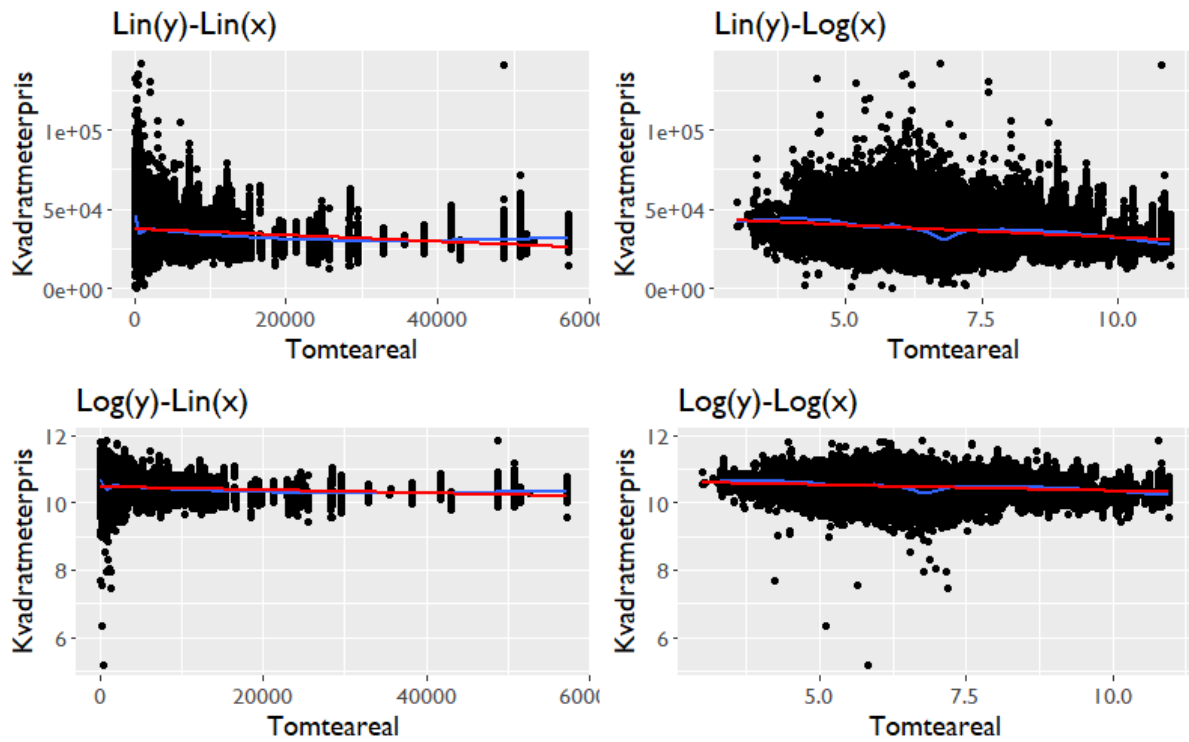
Figur A2.6: Scatter plot - salgstid med og uten transformasjoner



Figur A2.7: Fordeling - tomteareal med og uten transformasjoner



Figur A2.8: Scatter plot - tomteareal med og uten transformasjoner



A3 Prediksjoner av boligpriser rundt Åsane-trasseen

A3.1 Årlige prediksjoner for alle boliger innen 3000 meter fra nærmeste bybanestopp

Tabell A3.1: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 4.61 % | 0.34 % | -1.07 % |
| År 2-3 | -2.24 % | 0.00 % | 0.49 % |
| År 3-4 | 1.41 % | -0.21 % | 0.77 % |
| År 4-5 | 3.58 % | 1.95 % | 2.64 % |
| År 5-6 | -1.89 % | -1.44 % | 1.65 % |
| År 6-7 | 6.48 % | 5.76 % | 1.99 % |
| År 7-8 | 2.14 % | -0.67 % | 4.66 % |
| År 8-9 | -2.32 % | 3.28 % | 1.55 % |
| År 9-10 | 4.63 % | 1.63 % | 1.29 % |
| Snitt | 1.82 % | 1.18 % | 1.55 % |

Tabell A3.2: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 3000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett

| | Predikert vekst - SL (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 5.29 % | 0.34 % | -1.07 % |
| År 2-3 | -4.94 % | 0.00 % | 0.49 % |
| År 3-4 | 6.18 % | -0.21 % | 0.77 % |
| År 4-5 | -1.24 % | 1.95 % | 2.64 % |
| År 5-6 | -1.73 % | -1.44 % | 1.65 % |
| År 6-7 | 9.35 % | 5.76 % | 1.99 % |
| År 7-8 | -3.74 % | -0.67 % | 4.66 % |
| År 8-9 | -1.60 % | 3.28 % | 1.55 % |
| År 9-10 | 8.05 % | 1.63 % | 1.29 % |
| Snitt | 1.73 % | 1.18 % | 1.55 % |

Tabell A3.3: Koeffisientene for prediksjonene av alle boliger innen 3000 meter fra nærmeste bybanestopp - Super Learner

| | SL.mean | SL.lm | SL.glmnet | SL.nnet | SL.kernelKnn | SL.polymars |
|-------|---------|--------|-----------|---------|--------------|-------------|
| År 1 | 0 | 0.4309 | 0 | 0 | 0.0597 | 0.5094 |
| År 2 | 0 | 0.1982 | 0.4362 | 0 | 0.2171 | 0.1485 |
| År 3 | 0 | 0.4318 | 0.0526 | 0 | 0.1201 | 0.3955 |
| År 4 | 0 | 0.3692 | 0.4396 | 0 | 0.0700 | 0.1212 |
| År 5 | 0 | 0.6642 | 0 | 0 | 0.0348 | 0.3010 |
| År 6 | 0 | 0.2358 | 0.3363 | 0 | 0.0472 | 0.3807 |
| År 7 | 0 | 0.5371 | 0 | 0 | 0.1230 | 0.3400 |
| År 8 | 0 | 0.4862 | 0 | 0 | 0.2155 | 0.2983 |
| År 9 | 0 | 0.1456 | 0.1709 | 0 | 0.1415 | 0.5421 |
| År 10 | 0 | 0.3151 | 0 | 0 | 0.0959 | 0.5890 |

Tabell A3.4: Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 3000 meter nærmeste bybanestopp - Elastic Net

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|---------|---------|---------|---------|---------|
| Øvre grense | 38159.6 | 39797.8 | 38886.7 | 39485.2 | 40829.0 |
| Nedre grense | 36733.1 | 38550.4 | 37705.8 | 38188.6 | 39628.5 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 40101.9 | 42699.0 | 43650.4 | 42732.4 | 44758.3 |
| Nedre grense | 38831.0 | 41346.9 | 42190.3 | 41120.3 | 42979.2 |
| Snitt | | | | | |
| Øvre grense | 41110.0 | | | | |
| Nedre grense | 39727.4 | | | | |

Tabell A3.5: Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 3000 meter fra nærmeste bybanestopp - Super Learner

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|---------|---------|---------|---------|---------|
| Øvre grense | 37707.3 | 39426.9 | 37545.3 | 39892.4 | 39349.8 |
| Nedre grense | 36133.4 | 38317.9 | 36360.9 | 38577.7 | 38143.5 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 38611.5 | 42281.1 | 40763.5 | 40258.3 | 43461.2 |
| Nedre grense | 37543.8 | 40998.2 | 39399.3 | 38618.0 | 41763.3 |
| Snitt | | | | | |
| Øvre grense | 39929.7 | | | | |
| Nedre grense | 38585.6 | | | | |

A3.2 Årlige prediksjoner for alle boliger innen 1000 meter fra nærmeste bybanestopp

Tabell A3.6: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 3.21 % | 0.43 % | -2.42 % |
| År 2-3 | -2.50 % | 0.95 % | -0.05 % |
| År 3-4 | 2.99 % | 0.08 % | 1.07 % |
| År 4-5 | 2.69 % | 2.70 % | 3.78 % |
| År 5-6 | 1.69 % | 0.85 % | 0.34 % |
| År 6-7 | 2.49 % | 1.75 % | 3.56 % |
| År 7-8 | 4.46 % | 1.45 % | 3.58 % |
| År 8-9 | -1.52 % | 1.77 % | 1.97 % |
| År 9-10 | 4.86 % | 2.35 % | -0.03 % |
| Snitt | 2.04 % | 1.37 % | 1.31 % |

Tabell A3.7: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett

| | Predikert vekst - SL (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1 | 2.59 % | 0.43 % | -2.42 % |
| År 2 | -4.56 % | 0.95 % | -0.05 % |
| År 3 | 7.90 % | 0.08 % | 1.07 % |
| År 4 | -1.96 % | 2.70 % | 3.78 % |
| År 5 | -0.89 % | 0.85 % | 0.34 % |
| År 6 | 7.75 % | 1.75 % | 3.56 % |
| År 7 | -1.46 % | 1.45 % | 3.58 % |
| År 8 | -0.80 % | 1.77 % | 1.97 % |
| År 9 | 8.33 % | 2.35 % | -0.03 % |
| Snitt | 1.88 % | 1.37 % | 1.31 % |

Tabell A3.8: Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net

| | År 1 | År 2 | År 3 | År 4 | År 5 | |
|--------------|---------|---------|---------|---------|---------|---------|
| Øvre grense | 43219.5 | 44533.5 | 43383.7 | 44698.9 | 45830.2 | |
| Nedre grense | 41811.5 | 43224.8 | 42179.9 | 43422.3 | 44662.8 | |
| | År 6 | År 7 | År 8 | År 9 | År 10 | |
| Øvre grense | 46682.1 | 47840.8 | 50050.5 | 49370.6 | 51772.3 | |
| Nedre grense | 45340.5 | 46475.4 | 48475.3 | 47659.2 | 49972.1 | |
| Snitt | | | | | | |
| Øvre grense | | | | | | 46738.2 |
| Nedre grense | | | | | | 45322.4 |

Tabell A3.9: Konfidensintervall for gjennomsnittet av predikerte boligpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner

| | År 1 | År 2 | År 3 | År 4 | År 5 | |
|--------------|---------|---------|---------|---------|---------|---------|
| Øvre grense | 42925.7 | 43731.1 | 41839.3 | 45112.2 | 44178.5 | |
| Nedre grense | 41182.0 | 42553.5 | 40510.3 | 43741.6 | 42931.4 | |
| | År 6 | År 7 | År 8 | År 9 | År 10 | |
| Øvre grense | 43761.2 | 47171.7 | 46582.4 | 46357.0 | 50141.5 | |
| Nedre grense | 42570.9 | 45849.1 | 45083.4 | 44578.3 | 48369.2 | |
| Snitt | | | | | | |
| Øvre grense | | | | | | 45180.1 |
| Nedre grense | | | | | | 43737.0 |

A3.3 Årlige prediksjoner for alle leiligheter innen 1000 meter fra nærmeste bybanestopp

Tabell A3.10: Gjennomsnittet av årlig prosentvis prisvekst for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---------------------------------------|-------------------------------------|---------------|
| År 1-2 | 3.31 % | 0.98 % | -1.57 % |
| År 2-3 | -2.13 % | 1.14 % | 0.89 % |
| År 3-4 | 2.58 % | -0.37 % | 1.86 % |
| År 4-5 | 3.02 % | 4.28 % | 2.10 % |
| År 5-6 | 2.19 % | 0.53 % | 0.60 % |
| År 6-7 | 2.28 % | 1.46 % | 3.29 % |
| År 7-8 | 4.71 % | 1.40 % | 4.06 % |
| År 8-9 | -1.33 % | 2.14 % | 1.25 % |
| År 9-10 | 3.87 % | 0.72 % | 0.36 % |
| Snitt | 2.05 % | 1.36 % | 1.43 % |

Tabell A3.11: Gjennomsnittet av årlig prosentvis prisvekst for alle leiligheter innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett

| | Predikert vekst - SL (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 2.49 % | 0.98 % | -1.57 % |
| År 2-3 | -4.12 % | 1.14 % | 0.89 % |
| År 3-4 | 7.45 % | -0.37 % | 1.86 % |
| År 4-5 | -1.91 % | 4.28 % | 2.10 % |
| År 5-6 | -0.16 % | 0.53 % | 0.60 % |
| År 6-7 | 7.43 % | 1.46 % | 3.29 % |
| År 7-8 | -1.71 % | 1.40 % | 4.06 % |
| År 8-9 | -0.84 % | 2.14 % | 1.25 % |
| År 9-10 | 8.27 % | 0.72 % | 0.36 % |
| Snitt | 1.88 % | 1.36 % | 1.43 % |

Tabell A3.12: Konfidensintervall for gjennomsnittet av predikerte leilighetpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|-------------|-------------|-------------|-------------|--------------|
| Øvre grense | 43777.1 | 45096.2 | 44102.2 | 45301.4 | 46579.1 |
| Nedre grense | 42520.4 | 44060.7 | 43153.1 | 44201.1 | 45622.1 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 47678.7 | 48781.6 | 51146.9 | 50595.9 | 52549.6 |
| Nedre grense | 46538.5 | 47587.2 | 49756.4 | 48961.6 | 50857.2 |
| Snitt | | | | | |
| Øvre grense | 47560.9 | | | | |
| Nedre grense | 46325.8 | | | | |

Tabell A3.13: Konfidensintervall for gjennomsnittet av predikerte leilighetpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|-------------|-------------|-------------|-------------|--------------|
| Øvre grense | 43524.8 | 44311.0 | 42605.1 | 45750.8 | 44809.3 |
| Nedre grense | 41965.0 | 43311.4 | 41402.9 | 44513.0 | 43731.6 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 44724.6 | 48091.2 | 47379.0 | 47157.8 | 50953.0 |
| Nedre grense | 43671.3 | 46876.5 | 45963.3 | 45396.7 | 49256.4 |
| Snitt | | | | | |
| Øvre grense | 45930.7 | | | | |
| Nedre grense | 44608.8 | | | | |

A3.4 Årlige prediksjoner for alle eneboliger innen 1000 meter fra nærmeste bybanestopp

Tabell A3.14: Gjennomsnittet av årlig prosentvis prisvekst vekst for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 7.20 % | 1.97 % | 8.19 % |
| År 2-3 | -5.48 % | 5.07 % | -9.96 % |
| År 3-4 | 7.43 % | 0.77 % | 2.79 % |
| År 4-5 | 3.00 % | -9.60 % | 12.52 % |
| År 5-6 | -6.88 % | 7.71 % | -1.30 % |
| År 6-7 | 9.77 % | -1.35 % | 2.50 % |
| År 7-8 | -3.13 % | 4.72 % | 2.36 % |
| År 8-9 | 11.47 % | 5.49 % | -4.05 % |
| År 9-10 | -5.76 % | 1.21 % | 11.99 % |
| Snitt | 1.96 % | 1.78 % | 2.78 % |

Tabell A3.15: Gjennomsnittet av årlig prosentvis prisvekst for alle eneboliger innen 1000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett

| | Predikert vekst - SL (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 19.43 % | 1.97 % | 8.19 % |
| År 2-3 | -10.09 % | 5.07 % | -9.96 % |
| År 3-4 | 12.16 % | 0.77 % | 2.79 % |
| År 4-5 | -3.38 % | -9.60 % | 12.52 % |
| År 5-6 | -5.52 % | 7.71 % | -1.30 % |
| År 6-7 | 14.11 % | -1.35 % | 2.50 % |
| År 7-8 | -0.08 % | 4.72 % | 2.36 % |
| År 8-9 | 9.46 % | 5.49 % | -4.05 % |
| År 9-10 | -8.20 % | 1.21 % | 11.99 % |
| Snitt | 3.10 % | 1.78 % | 2.78 % |

Tabell A3.16: Konfidensintervall for gjennomsnittet av predikerte eneboligpriser innen 1000 meter fra nærmeste bybanestopp - Elastic Net

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|---------|---------|---------|---------|---------|
| Øvre grense | 32716.1 | 34254.5 | 32138.7 | 35102.7 | 35530.0 |
| Nedre grense | 27221.0 | 29996.1 | 28588.6 | 30139.4 | 31670.3 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 33944.1 | 36127.0 | 35983.3 | 39668.7 | 37830.6 |
| Nedre grense | 28631.0 | 32561.6 | 30558.7 | 34503.5 | 32067.1 |
| Snitt | | | | | |
| Øvre grense | | | | | |
| Nedre grense | | | | | |

Tabell A3.17: Konfidensintervall for gjennomsnittet av predikerte eneboligpriser innen 1000 meter fra nærmeste bybanestopp - Super Learner

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|---------|---------|---------|---------|---------|
| Øvre grense | 31025.6 | 35507.0 | 31702.8 | 36497.3 | 34473.7 |
| Nedre grense | 25002.9 | 31408.0 | 28458.6 | 30980.9 | 30724.0 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 32674.0 | 36852.1 | 38022.7 | 40838.3 | 38478.3 |
| Nedre grense | 28925.2 | 33437.2 | 32207.9 | 36033.9 | 32086.9 |
| Snitt | | | | | |
| Øvre grense | | | | | |
| Nedre grense | | | | | |

A3.5 Årlige prediksjoner for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp

Tabell A3.18: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net, Historiske priser, Treningssett

| | Predikert vekst - EN (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---------------------------------------|-------------------------------------|---------------|
| År 1-2 | 11.59 % | 3.20 % | 0.77 % |
| År 2-3 | -2.69 % | -2.89 % | 0.79 % |
| År 3-4 | -1.05 % | -0.03 % | 0.64 % |
| År 4-5 | 5.92 % | 0.91 % | 1.06 % |
| År 5-6 | -4.17 % | -1.24 % | 3.12 % |
| År 6-7 | 7.24 % | 6.78 % | 0.22 % |
| År 7-8 | -0.40 % | -2.88 % | 5.86 % |
| År 8-9 | -5.43 % | 5.24 % | 0.33 % |
| År 9-10 | 0.72 % | -2.50 % | 2.66 % |
| Snitt | 1.30 % | 0.73 % | 1.72 % |

Tabell A3.19: Gjennomsnittet av årlig prosentvis prisvekst for alle boliger mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner, Historiske priser, Treningssett

| | Predikert vekst - SL (Med Bybanen) | Historiske priser (Uten Bybanen) | Treningssett |
|--------------|---|---|---------------------|
| År 1-2 | 15.28 % | 3.20 % | 0.77 % |
| År 2-3 | -6.57 % | -2.89 % | 0.79 % |
| År 3-4 | 3.46 % | -0.03 % | 0.64 % |
| År 4-5 | 0.69 % | 0.91 % | 1.06 % |
| År 5-6 | 0.55 % | -1.24 % | 3.12 % |
| År 6-7 | 6.11 % | 6.78 % | 0.22 % |
| År 7-8 | -6.36 % | -2.88 % | 5.86 % |
| År 8-9 | -4.68 % | 5.24 % | 0.33 % |
| År 9-10 | 4.01 % | -2.50 % | 2.66 % |
| Snitt | 1.39 % | 0.73 % | 1.72 % |

Tabell A3.20: Konfidensintervall for gjennomsnittet av predikerte boligpriser mellom 1000 og 3000 meter fra nærmeste bybanestopp - Elastic Net

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|-------------|-------------|-------------|-------------|--------------|
| Øvre grense | 29849.8 | 33084.8 | 32220.5 | 31896.7 | 33679.8 |
| Nedre grense | 28006.8 | 31480.0 | 30607.6 | 30274.5 | 32173.8 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 32237.4 | 34728.1 | 34465.4 | 32825.5 | 33139.5 |
| Nedre grense | 30867.5 | 32944.5 | 32938.1 | 30916.5 | 31061.7 |
| Snitt | | | | | |
| Øvre grense | 32812.8 | | | | |
| Nedre grense | 31127.1 | | | | |

Tabell A3.21: Konfidensintervall for gjennomsnittet av predikerte boligpriser mellom 1000 og 3000 meter fra nærmeste bybanestopp - Super Learner

| | År 1 | År 2 | År 3 | År 4 | År 5 |
|--------------|-------------|-------------|-------------|-------------|--------------|
| Øvre grense | 29222.5 | 33299.5 | 31225.0 | 32310.0 | 32496.9 |
| Nedre grense | 27365.8 | 31936.4 | 29726.3 | 30752.3 | 31002.1 |
| | År 6 | År 7 | År 8 | År 9 | År 10 |
| Øvre grense | 32527.4 | 34694.2 | 32451.6 | 31317.5 | 32385.0 |
| Nedre grense | 31321.8 | 33055.1 | 30991.9 | 29154.5 | 30514.4 |
| Snitt | | | | | |
| Øvre grense | 32193.0 | | | | |
| Nedre grense | 30582.1 | | | | |