

Modellutvalg og evaluering

- ✓ (a) Er oppdeling i trenings-, validerings- og testdata meningsfull og som det skal være for tidsavhengige data? **Ja, dette er gjort.**
- ✓ (b) Er variabelutvinning og imputasjonsstrategiene tilpasset på treningsdata? **Ja, det er sammenlignet mellom imputasjonsmodeller også.**
- ✓ (c) Er alle modelltilpassinger gjort på treningsdata? **Jepp, det er det.**
- ✓ (d) Er modellutvalg gjort korrekt på valideringsdata? **Ja, men kunne ha testet flere modeller ut i fra RMSE.**
- ✗ (e) Er variabelutvalg gjort korrekt på valideringsdata? **Ikke begynt på dette enda.**
- ✗ (f) Er kun den beste modellen analysert med testdata for å angi generaliseringsevnen? **Tror bare ElasticNet, LinearRegression er testet, ville ha testet litt flere modeller. RandomForest f.eks.**

Variabelutvinning og manglende data

- ✓ (a) Gir variablene i datasettet mening? **Rushhour, hverdag/helg, årstid er lagd. De gir mening, selv om det ikke er ønskelig å bruke æøå i variabelnavn.**
- ✓ (b) Brukte studenten både værdata og trafikkdata? **Ja.**
- ✓ (c) Har studenten prøvd å lage flere variabler enn de som allerede er inneholdt i rådata? **Ja, rushhour, hverdag, helg, årstid.**
- ✓ (d) Er alle manglende data godt beskrevet? **Har identifisert manglende data i Relativ luftfuktighet.**
- ✓ (e) Gir strategien for å håndtere manglende data mening? **Jepp, jeg gjør det samme her, 81% mangler, da dropper man.**

Modellering

- ✓ (a) Brukes det riktig type modell (regresjon eller klassifikasjon)? **Ja, det er brukt regresjon.**
- ✗ (b) Er en enkel grunnlinjemodell vurdert? **Nei, det burde ha vært en DummyRegressor her som en baseline.**
- ✗ (c) Er resultater av tre fundamentalt forskjellige modeller rapportert? **LinearRegression og ElasticNet er av samme type.**

Rapport

- ✓ (a) Er rapporten tydelig, velskrevet og forståelig? **Ja!**
- ✓ (b) Er alle data science-stegene forklart? **Jepp.**
- ✓ (c) Er alle valg som studenten gjorde rettfærdiggjort? **Ja.**
- ✓ (d) Har alle figurene en figurlegende som beskriver figuren? **Ja.**
- ✓ (e) Er koden forståelig og godt dokumentert? **Ja, men det hadde vært enda enklere å forstå hvis variabelnavn hadde vært litt enklere og æøå hadde vært unngått.**

Generelle tilbakemeldinger

- (a) Nevn to ting du likte best med denne løsningen av oppgaven. **Jeg likte veldig godt at du fant ut den beste imputasjonsmodellen, og det var mange nyttige figurer med som ga forståelse til dataen.**
- (b) Nevn hva du synes er viktigst å forbedre før neste innleveringen. **Variabelnavn, finne en bedre modell (RandomForest!), også Holidays pakken, <https://pypi.org/project/holidays/>, lar deg lett opprette dummy-vars for ferie i Norge, de ga meg en mye bedre RMSE.**