

# Business Intelligence II: Eksamenscase 2025

## HD 1. del, ordinær eksamen

### Data

Datafilen, `rentalbikes.csv`, indeholder oplysninger vedrørende leje af cykler i en (amerikansk) storby gennem en toårig periode. Cyklerne udlejes både til “registrerede brugere” og til “lejlighedsvis brugere”. Data er opgjort på timebasis, dvs der er 24 datalinjer per dag.

Casen fokuserer *andelen* af cykler lejet af lejlighedsvis brugere: Der ønskes opstillet modeller, som viser, hvordan denne andel afhænger af tidsmæssige og vejrmæssige variable. Analysen skal gøre brug af mindst to *relevante* analysemetoder/modeller blandt dem, der er gennemgået på kurset.

### Variabelforklaring

Andelen af cykler lejet af lejlighedsvis brugere findes i variabelen `fraction`.

### Tidsmæssige variable

Datafilen indeholder en række oplysninger vedrørende tidspunktet for målingerne:

- `season` indeholder årstiden
- `year`, `month` og `hour` indeholder året, måneden og klokkeslættet
- `weekday` indeholder ugedagen
- `holiday` indeholder oplysninger om dagen var en helligdag

### Vejrmæssige variable

- `temperature` angiver temperaturen (i grader celsius)
- `feelslike` angiver den “oplevede temperatur” (i grader celsius), altså temperaturen korrigeret for “chill factor” mv
- `windspeed` angiver vindhastigheden (i m/s)
- `humidity` angiver luftfugtigheden (i procent)
- `weather` giver en overordnet beskrivelse af vejret

## Eksamen

Eksaminationstiden på 20 minutter fordeles på følgende måde:

- De første 10 minutter bruges på din gennemgang af den forberedte analyse af datasættet (jf. ovenfor). Du må medbringe slides af analyseresultater, softwareoutput og andet til brug for gennemgangen af den forberedte analyse (i udprintet form).
- De næste 5 minutter bruges på gennemgang af ét af følgende spørgsmål, som trækkes ved eksamen:
  1. Gør kort rede for ideen bag logistisk regression og diskuter kort metodens fordele og ulemper.
  2. Gør kort rede for ideen bag regulariseret og splinebaseret regression og diskuter kort metodens fordele og ulemper.
  3. Gør kort rede for ideen bag nearest neighbour-metoder og diskuter kort metodens fordele og ulemper.
  4. Gør kort rede for ideen bag træ-baserede metoder og diskuter kort metodens fordele og ulemper.
  5. Gør kort rede for ideen bag neurale netværk og diskuter kort metodens fordele og ulemper.
  6. Gør kort rede for ideen bag support vector machines og diskuter kort metodens fordele og ulemper.
- De sidste 5 minutter bruges på votering.

## Bemærk

- Det er dit eget ansvar at nå at gennemgå de væsentlige konklusioner i den forberedte analyse indenfor de første 10 minutter af eksamenstiden.
- Gyldigt studiekort (eller anden officielt udstedt billed-id) *skal* fremvises ved eksaminationens start.

## Bedømmelse

- Der lægges i bedømmelsen vægt på din evne til at udvælge relevante dele af pensum til brug for den gennemførte analyse, og ligeledes vægt på din evne til at formidle analysens væsentligste konklusioner indenfor den afsatte tidsramme (de første 10 min.).
- Fastsættelsen af eksamenskarakteren baseres på kursets læringsmål, som du finder her: <https://kursus.katalog.cbs.dk/2024-2025/DIP-D1FMV5012U.aspx>.
- Der kan stilles korte spørgsmål fra eksaminator og bi-eksaminator undervejs i gennemgangen.