

Obligatorisk læringsaktivitet i Flow 2, Dataanalyse



Afleveringsfrist:

Onsdag den 6. november 2023, kl. 22

Udarbejdet af:

Thorbjørn Baum, lektor Cphbusiness

Thorbjørn Wulf, adjunkt Cphbusiness

Indhold

Opgave 1 – Den bedste forbrugertillidsindikator	3
Opgave 1.1 – Kombinationsalgoritme i R	3
Opgave 1.2 – R^2 og forbrugertillidsindikatorer	3
Opgave 1.3 – Spørgsmål i indikatoren	3
Opgave 1.4 – Forudsigelser med afsæt i jeres indikatorer	3
Opgave 1.5 – Sammenlign med en mikroøkonomisk indikator	3
Opgave 2 – Principal Component Analysis	4
Opgave 2.1 – PCA regression	4
Opgave 2.2 – De vigtigste indikatorer	4
Opgave 2.3 – Forudsig forbruget	4
Opgave 2.4 – Forudsigelser fra DI	4
Opgave 3 – Julehandel i 2024	5
Opgave 3.1 – Forudsigelse	5
Opgave 3.2 – Validering af model	5
Opgave 4 – Stabilitet i jeres forbrugertillidsindikator	6
Opgave 4.1 – Test af model fra opgave 1	6
Opgave 5 – Data Science modellen og DMI's API	7
Opgave 1.1 – Det første skridt	7
Opgave 1.2 – Undersøg de data API'et returnerer	7
Opgave 1.3 – Andre API'er	7
Opgave 1.4 – En graf med mangler	7
Opgave 1.5 – Eksempler på anvendelse?	8

Opgave 1 – Den bedste forbrugertillidsindikator

Opgave 1.1 – Kombinationsalgoritme i R

Lav alle kombinationer af de 12 spørgsmål i forbrugerundersøgelsen af DST. I skal bruge data fra 1. kvartal 2000 til og med 4. kvartal 2024.

Opgave 1.2 – R^2 og forbrugertillidsindikatorer

Hvad er R^2 for den indikatorer, der bedst forklarer variationen i den kvartalsvise årlige realvækst i husholdningernes forbrugsudgift? Diskutér om denne indikator er bedre end DI's forbrugertillidsindikator. (Hint: se jeres besvarelse af opgave 2 i 2. OLA i flow 1 og opstil en tabel, der sammenligner de to indikatorer¹)

Opgave 1.3 – Spørgsmål i indikatoren

Hvilke spørgsmål indgår i den indikator, som er bedst til at forklarer variationen i forbruget? Giver kombinationen af spørgsmål analytisk mening? I bedes overveje alternative indikatorer fra jeres oversigt.

Opgave 1.4 – Forudsigelser med afsæt i jeres indikatorer

Forudsig udviklingen i husholdningernes forbrugsudgift i 3. og 4. kvartal 2024. (Hint: brug jeres svar i opgave 2 fra OLA 2 sammen med jeres svar fra de forrige opgaver. Vær opmærksom på, om tallene fra forbrugerundersøgelsen for oktober er udgivet og I har den første måned i 4. kvartal)

Opgave 1.5 – Sammenlign med en mikroøkonomisk indikator

Find den bedste indikator, der alene består af mikroøkonomiske spørgsmål i forbrugertillidsundersøgelsen og sammenlign indikatoren med jeres tidligere svar i opgave 1.

¹ Se tabel i artiklen: "Forbruget fortsætter fremgangen i 2016"

Opgave 2 – Principal Component Analysis

Opgave 2.1 – PCA regression

Lav en PCA regression på jeres data fra opgave 1, hvor Y er jeres årlige realvækst i husholdningernes forbrugsudgift og X er alle de 12 spørgsmål fra DST's forbrugerforventningsundersøgelsen.

Opgave 2.2 – De vigtigste indikatorer

Hvilke 3 spørgsmål i forbrugerforventningsundersøgelsen har de højeste vægte? Diskutér om spørgsmålene giver analytisk mening.

Opgave 2.3 – Forudsig forbruget

Med afsæt i jeres vægte fra opgave 2.1 skal I beregne den årlige realvækst i husholdningernes forbrugsudgift for 3. kvartal 2023. Hvad er årsagen til I ikke helt tror på jeres beregnede årlige realvækst i husholdningernes forbrugsudgift for 3. kvartal 2023? Hvad skyldes det underlige resultat?

Opgave 2.4 – Forudsigelser fra DI

Hvad forventede DI den årlige realvækst i husholdningernes forbrugsudgift er i 2016? (se artiklen: "Forbruget fortsætter fremgangen i 2016") Hvad endte den årlige realvækst i husholdningernes forbrugsudgift med reelt at være?

Opgave 3 – Julehandel i 2024

Opgave 3.1 – Forudsigelse

Lav en Machine Learning model, der med afsæt i DST's forbrugertillidsindikator, kan forudsige om julehandlen i 2024 er større end i 2023.

Opgave 3.2 – Validering af model

Lav en vurdering af validiteten af jeres model fra opgave 3.1.

Opgave 4 – Stabilitet i jeres forbrugertillidsindikator

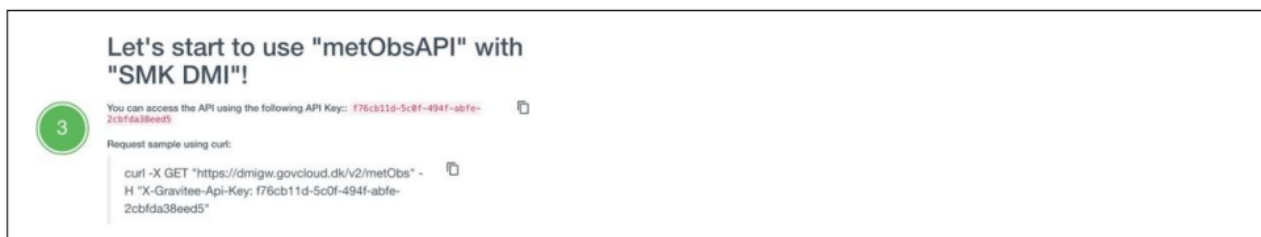
Opgave 4.1 – Test af model fra opgave 1

Undersøg stabiliteten af jeres fundne indikator fra opgave 1. Giv en grundig forklaring på opsætning til at undersøge stabiliteten.

Opgave 5 – Data Science modellen og DMI's API

Opgave 1.1 – Det første skridt

Sørg for at registrere dig på metObsAPI'et og tag et skærmdump svarende til nedenstående, men med din nøgle og ikke den, som er på billedet. Desuden skal applikationen hedde "My CPH Weather app" i stedet for "SMK DMI".



Opgave 1.2 – Undersøg de data API'et returnerer.

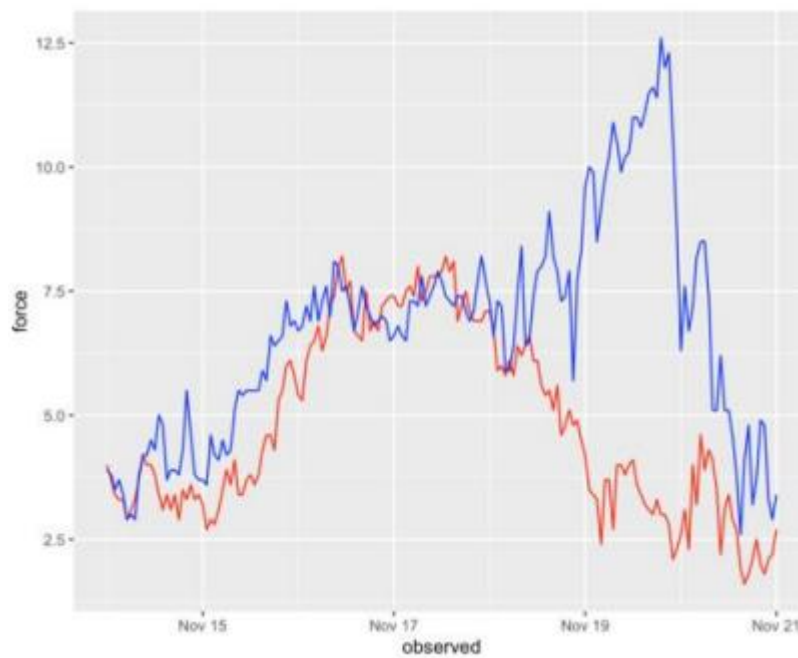
Vedhæft et link til den side hvor de forskellige observationer er beskrevet. Hvilken enhed angiver man "visibility" i? Hvor ofte opdateres "wind_max"? Hvilken by gemmer sig bag stationen med id'et 05272?

Opgave 1.3 – Andre API'er

DMI har også et andet API, hvor man kan trække en vejrudsigt. Angiv linket til API'ets "endpoint". Hvor mange "collections" kan man spørge til? Hvilket ID har Lille Bælt? Hvilket filformat ender man med at få "forecasten" i?

Opgave 1.4 – En graf med mangler

Her er en graf over vindobservationer fra to stationer - Århus og Anholt - fra stormen i november '22. Hvilken retning kom vinden fra? Hvordan ville I illustrere retningen på grafen? Lav samme graf (gerne meget bedre) for stormen, der i oktober 2023 hærgede Danmark.



Opgave 1.5 – Eksempler på anvendelse?

Kan I finde links til steder, hvor virksomheder eller personer har brugt DMI's API? Forklar business casen i jeres eksempler på anvendelse af API'et.