

Forecasting

1 Populariteit van een app

Een bedrijf heeft vorige maand een nieuwe app gelanceerd. Ze hebben iedere dag genoteerd hoeveel mensen deze gedownload hebben. Je vindt deze gegevens in "populariteitApp.csv". Er zijn in dit geval geen verkeerde of ontbrekende gegevens. Je zou het bestand dus gemakkelijk moeten kunnen lezen. Zet wel de datums om naar datetime objecten.

1. Maak een plot van de gegevens zodat je al wat inzicht krijgt in de data.
2. Kan je de seizoensgrootte vinden?
3. Voorspel nu, voor iedere voorspellingsmethode, de drie volgende dagen. Bereken telkens de MAE, RMSE en de MAPE. Vervolledig volgende tabel met alle resultaten:

techniek	dag 1	dag 2	dag 3	MAE	RMSE	MAPE
naïef						
gemiddelde						
voortschrijdend (m=5)						
lin. combinatie (m=13)						
trend						
additieve decompositie (m=7)						
multiplicatieve decompositie (m=7)						

Geef nog antwoord op deze vragen:

4. lineaire combinatie: welke gewichten vind je? Welke waarde speelt dus de grootste rol in het voorspellen van de volgende waarde?
5. trend: wat is de formule van de regressielijn (gebruik lineaire regressie)?
6. additieve decompositie: wat is de formule voor de trendlijn (lineaire regressie)?
7. additieve decompositie: wat zijn de waarden voor het weerkerend patroon?
8. multiplicatieve decompositie: wat is de formule voor de trendlijn (lineaire regressie)?
9. multiplicatieve decompositie: wat zijn de waarden voor het weerkerend patroon?
10. welke techniek geeft de beste voorspelling en waarom?

2 Bezoekers in een pretpark

Een pretpark wil kunnen voorspellen hoeveel bezoekers er zullen komen in een bepaalde maand. Daarvoor hebben ze gedurende 4 jaar opgeschreven hoeveel bezoekers er waren, voor iedere maand. Je vindt deze data in "pretpark.csv". Deze data is ook direct leesbaar in Python (geen foute of ontbrekende data).

1. Maak een grafiek van de data. Wat merk je visueel op? Stijgt of daalt het aantal bezoekers? Is er een patroon herkenbaar?
2. Zou je het additieve of multiplicatieve model gebruiken?
3. Wat is de grootte van 1 "seizoen"?
4. Doe nu een decompositie. Plot deze in een diagram.
5. We gaan nu een model maken om voorspellingen te doen naar de toekomst. Doe een lineaire regressie op de trend en bepaal de slope en de intercept. Welke vergelijking vind je voor de trend?
6. Bereken de waarden van de trend voor het komende jaar. Welke waarden kom je uit?
7. Wat zijn de waarden van het seizoen?
8. Vermenigvuldig de trend-waarden nu met het patroon. Je hebt nu een schatting van het aantal bezoekers voor het volgende jaar. Welke waarden kom je uit (rond de waarden af op een geheel getal)?
9. wat is de MAE, RMSE en de MAPE? Wat is de betekenis van de MAPE?