



Taak "Machine Learning"

Project "Solar"

**HO
GENT**

Project “Solar”: situering

De eigenaar van een huishoudelijke zonnepaneelinstallatie zou graag continu een voorspelling hebben van de opbrengst van zijn panelen gedurende de komende uren om het eigen verbruik te optimaliseren: bij een verwachte hoge opbrengst kan hij dan bijv. beslissen om de wasmachine aan te zetten.

Project “Solar”: situering (vervolg)

- Hij beschikt over de meterstand per uur sedert ongeveer één jaar.
- Daarnaast zijn ook de gegevens van de waarnemingen van het weer en de uren van zonsopgang en –ondergang in dezelfde periode periode.

Waarnemingen van het weer en zon-op en -onder

- Deze gegevens zijn afkomstig van <https://opendata.meteo.be/downloadPage.php>
- Het is de zg. “synoptische data”.
- De gegeven dataset bevat gegevens van vier automatische weerstations in de buurt van de woonplaats van de zonnepaneleneigenaar.
- Elk weerstation is geïdentificeerd door de geografische coördinaten in de kolom “the_geom”.
- Lees zeer aandachtig de bijhorende documentatie:
 - https://opendata.meteo.be/geonetwork/srv/eng/catalog.search#/metadata/RMI_DATASET_SYNOP
 - <https://opendata.meteo.be/documentation/?dataset=synop>
- De gegevens voor zonsopgang en –ondergang zijn afkomstig van: <https://robinfo.oma.be/nl/astro-info/zon/zon-opkomst-ondergang-2024>

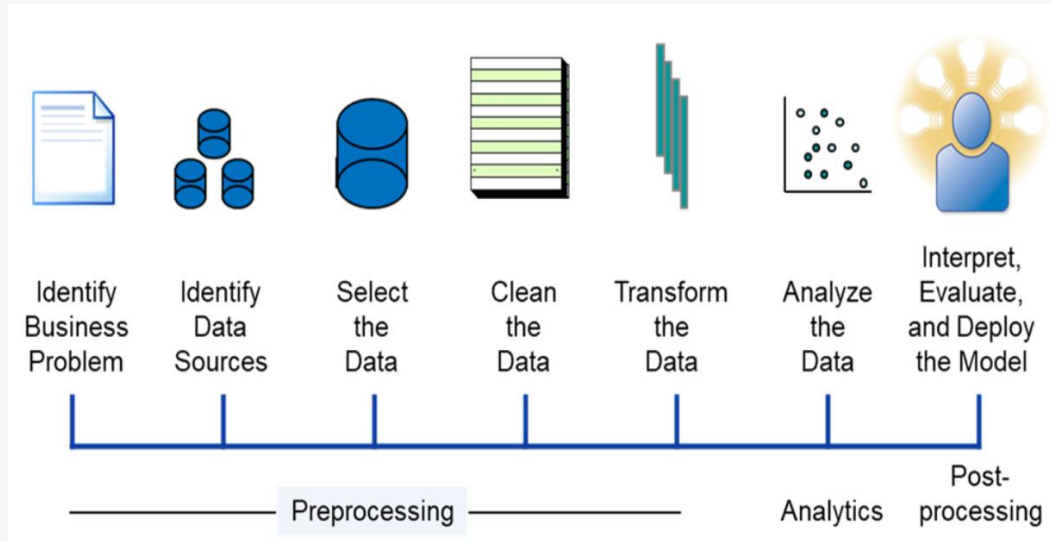
Datasets

- <https://github.com/HOGENT-ML/course/blob/main/datasets/solar.csv>
- <https://github.com/HOGENT-ML/course/blob/main/datasets/weather.csv>
- <https://github.com/HOGENT-ML/course/blob/main/datasets/sunrise-sunset.xlsx>

Taak

- Verken, zowel visueel als cijfermatig, de data met de bedoeling inzicht te verwerven in de natuur van de data.
- Probeer de data op te schonen (data cleaning!). Let vooral op mogelijke outliers in de registratie van de zonnepanelen.
- Stel een regressiemodel op om de opbrengst per uur te voorspellen.
 - Kies een optimaal regressiemodel door alle in de les geziene modellen uit te proberen en te vergelijken volgens de "best practices".
 - Kies als maatstaf de gemiddelde procentuele afwijking op uurbasis.
- Mogelijkheid tot vraagstelling en werken aan het project wekelijks tijdens het vierde lesuur (vanaf week 2).
- Op het einde van het semester zal je de resultaten individueel presenteren aan de lector. Bij die gelegenheid krijg je een bestand met nieuwe waarnemingen van het weer (bijv. voor mei 2024). Aan de hand van een zelfgeschreven Python-applicatie maak je dan, op basis van je model voorspellingen voor de opbrengst per uur voor een opgegeven periode.
- Je ontvangt dan ook een bestand met de “echte” waarden, zodat je meteen de nauwkeurigheid kan berekenen. Je zal op voorhand testbestanden krijgen om je applicatie te testen.

Houd rekening met: het data mining proces



± 80-90 % van de totale projecttijd gaat naar preprocessing wegens:

- Bad veracity!
- Data en business probleem begrijpen is meestal niet eenvoudig.
- Algoritmes zelf zijn off-the-shelf beschikbaar (meestal geen PhD nodig)
- 80 % transpiratie, 20 % inspiratie

Evaluatie

- U werkt alleen.
- Presentatie van project in week 13 gedurende een korte mondeling toelichting.
- 25 % van de evaluatie in eerste zittijd.
- Geen tweede examenkans → punten project worden overgenomen.