

Data Requirements



Digital Twin 3.0

Handpicked Agencies

Breda

Door: Koen Pijnenburg

Introductie

Het doel van het Twindle 3.0 project is om de bestaande oplossing uit te breiden zodat de luchtkwaliteit voorspelt kan worden. Om dit te bereiken zal de volgende vraag gesteld worden

Welke data is benodigd om luchtkwaliteit te voorspellen?

Naast de data die Twindle genereert kan het zijn dat hier externe databronnen voor benodigd zijn. In dit document zal dit onderzocht worden welke bronnen dit zijn door de huidige situatie te analyseren en te bekijken hoe de opkomende uitdagingen door andere zijn opgelost.

Inhoudsopgave

Twindle Data	4
Externe bronnen	5
Temperatuur	5
Luchtvochtigheid	5
TVOC & CO2	6
Aantal personen	6
Conclusie	7
Bronnen	8

Twindle Data

De Twindle applicatie genereert zelf data. Een overzicht van deze data kan in de onderstaande tabel gevonden worden. Per meetwaarde zal een korte toelichting gegeven worden.

```
TVOC          float64 # 'Total volatile organic compounds' in PPB
Pressure      float64 # Luchtdruk in Pa.
CO2           float64 # CO2 waarde in PPM.
Illumination  float64 # Verlichtingsniveau in flux.
Activity      float64 # ?
TimeStamp     datetime64[ns] # Tijd van meting.
TempInt       float64 # Temperatuur in celcius.
Humidity      float64 # Luchtvochtigheidspercentage.
DOOR_OPEN_STATUS float64 # Boolean of de deur open of dicht is.
DOOR_OPEN_TIMES float64 # Aantal keer dat de deur geopend is.
room          object # Ruimte waar de meting is verricht.
```

De bovenstaande waarden bepalen samen de classificatie van de luchtkwaliteit, energierendement en veiligheid in een ruimte. Scores worden berekend door middel van de onderstaande tabel.

Metric	Min (warning)	Max (warning)	Min (danger)	Max (danger)
Temperature	17	22,5	16,5	23,5
Humidity	40	60	30	70
CO2	400	800	300	1200
TVOC	-1	200	-1	250

Per meetwaarde wordt bekeken of het binnen bereik van goed, 3 punten, acceptabel, 2 punten, of slecht, 1 punt is. Deze punten worden gebruikt in de scoresystemen om classificaties te berekenen.

1. Air Quality: Temperature, humidity, CO2 & TVOC -> 12 punten maximaal.
2. Rendement: Temperature, windows open/closed -> 6 punten maximaal.
3. Safety: passive infrared sensor (PIR), CO2 -> 6 punten maximaal.

Externe bronnen

Het doel is om de in het voorgaande hoofdstuk genoemde classificaties te voorspellen. Om dit accuraat te kunnen doen zullen externe indicatoren voor temperatuur, luchtvochtigheid, CO₂ & TVOC, en het aantal personen in een ruimte gevonden moeten worden.

Temperatuur

Uit een onderzoek (Aguilera et al., 2019, p. 13) blijkt dat binnentemperatuur met ~92% accuracy te voorspellen valt. Hiervoor was een decision tree model getraind met de onderstaande data.

<u>Variabel</u>	<u>Informatie toevoeging</u>	<u>Correlation coefficient</u>
Thermostaatkraan instelling	0.55	0.14
Oppervlakte ruimte	0.23	0.13
Gemiddelde buitentemperatuur (uurlijk)	0.20	0.14
Bouwjaar	0.15	0.03
Zonnestraling	0.12	0.05
Aantal personen	0.10	0.10
Luchtvochtigheid buiten	0.09	0.11
Raamstand	0.03	0.11

Opmerkelijk is dat de hoeveelheid informatie die een variabel toevoegt aan het model niet perse gelijk staat met de correlatie tot de binnentemperatuur.

Luchtvochtigheid

Luchtvochtigheid kan erg accuraat voorspelt worden met een NNARX model (Lu & Viljanen, 2008, p. 347). Dit model voorspelt de verwachte luchtvochtigheid over 1 tijdsinterval met ~99% nauwkeurigheid. Het maakt gebruik van de onderstaande features.

- Tijdsinterval (15 min.);
- Buitentemperatuur;
- Binnentemperatuur;
- Luchtvochtigheid buiten;
- Luchtvochtigheid binnen;

TVOC & CO2

Uit onderzoek (Chen et al., 2018, p. 245) is gebleken dat TVOC waarden veranderen losstaand van de activiteit in een ruimte. CO2 waarden volgen dezelfde trend maar worden ook beïnvloed door het aantal personen in een ruimte.

Dit betekent dat time-series prediction toegepast kan worden om op een bepaalde dag en tijdstip de TVOC & CO2 waarde in te schatten. De volgende data was gebruikt om een support vector machine model te trainen.

- AM; Uur waarin de meting is genomen.
- Dag van de week.
- Actuele tijd.
- Lag; tijdsinterval tussen 1 en 12 minuten.

Voor CO2 kunnen dezelfde features gebruikt worden. Het aantal personen, of een deur openstaat en voor hoe lang moeten hier aan toegevoegd worden.

Aantal personen

Het aantal personen in een ruimte kan voorspelt worden aan de hand van de trend van de CO2 waarde (Brennan et al., 2018, p. 254). De nauwkeurigheid van de voorspelling was ~48% met een KNN model door gebruik te maken van de onderstaande data:

- Tijdsvenster (minuten); Aantal minuten aan sensor data in deze interval.
- Geduld; Aantal minuten delay voor volgende datapunt.
- CO2; CO2 waarde in parts per million
- Aantal personen; Het aantal aanwezige personen in een ruimte

Aangezien de nauwkeurigheid van dit model niet erg hoog is zullen er waarschijnlijk aanpassingen aan het model gemaakt moeten worden.

Conclusie

Ter conclusie kan er antwoord gegeven worden op de vraag welke data is benodigd om luchtkwaliteit te voorspellen? Dit zal gedaan worden door een lijst op te stellen met te verzamelen data en mogelijke bronnen. Deze lijst kan opgedeeld worden in gebouw en weer categorieën.

Gebouw

Meerdere modellen hebben constante en/of variable informatie van het gebouw nodig. Deze gegevens kunnen opgevraagd worden of er moeten sensoren voor geïnstalleerd worden.

Constant

- Oppervlakte van de ruimtes
- Bouwjaar

Variable

- Thermostaatkraan instellingen
- Raamstand

Weer

Gegevens rondom het weer kunnen waarschijnlijk verzameld worden vanaf het koninklijk nederlandse meteorologisch instituut (KNMI). De volgende gegevens moeten verzameld worden:

- Buitentemperatuur
- Luchtvochtigheid
- Zonnestraling

Bronnen

Aguilera, Andersen, & Toftum. (2019). Prediction of Indoor Air Temperature Using Weather Data and Simple Building Descriptors. *International Journal of Environmental Research and Public Health*, 16(22), 1–20. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6888563/>

Lu, T., & Viljanen, M. (2008). Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural Computing and Applications*, 18(4), 345–357. <https://doi.org/10.1007/s00521-008-0185-3>

Chen, S., Mihara, K., & Wen, J. (2018). Time series prediction of CO₂, TVOC and HCHO based on machine learning at different sampling points. *Building and Environment*, 146, 238–246. <https://doi.org/10.1016/j.buildenv.2018.09.054>

Brennan, C., Taylor, G. W., & Spachos, P. (2018). Designing learned CO₂ -based occupancy estimation in smart buildings. *IET Wireless Sensor Systems*, 8(6), 249–255. <https://doi.org/10.1049/iet-wss.2018.5027>