

Smart monitoring anno 2021

Een nieuwe (r)evolutie?

Door slimme toepassing van machine learning-algoritmes kan met modellen meer informatie uit grote hoeveelheden geluidmonitoring-data gehaald worden.

Door: Robbert-Jan Dikken, Simon de Koning, Wim van der Maarl

Over de auteurs:

dr. ir. R.J. Dikken, ir. S.J. de Koning en ing. W. van der Maarl zijn werkzaam bij Peutz bv, vestiging Zoetermeer

INTRODUCTIE

De toename van relatief goedkope dataopslagcapaciteit resulteert bij geluidmonitoring steeds vaker in grote hoeveelheden data. Daarnaast is de reken capaciteit de laatste jaren ook significant toegenomen. Dit leidt tot nieuwe mogelijkheden voor het gebruik van geluidmonitoringdata.

Tot op heden wordt veelal slechts een klein deel van de geluidmonitoringdata gebruikt binnen de adviespraktijk (bijvoorbeeld bij overschrijding van grenswaarden). Met een combinatie van expertkennis en machine learning-algoritmes kunnen modellen worden ontwikkeld waarmee meer informatie uit de monitoringdata wordt gehaald. In dit artikel worden mogelijkheden van machine learning beschreven en wordt het met twee praktijkvoorbeelden in relatie tot geluidmonitoring toegelicht.

MACHINE LEARNING IN HET KORT

Machine learning is een onderdeel van artificiële intelligentie. Het maakt gebruik van algoritmes die zelfstandig modellen ontwikkelen op basis van de verbanden die het algoritme ontdekt in of leert uit data.¹ Er zijn verschillende soorten toepassingen voor machine learning. Grofweg zijn deze vraagstukken in te delen in:

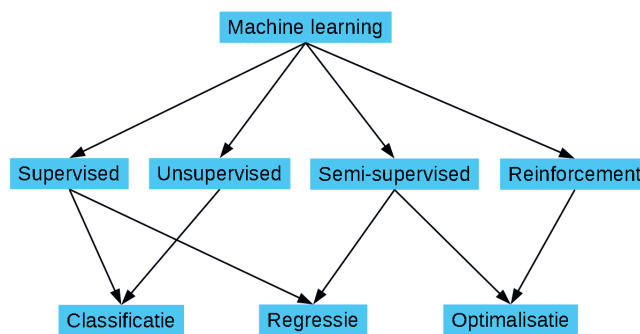
- voorspelling/regressie;
- classificatie;
- optimalisatie.

Bij voorspelling/regressie wordt de samenhang tussen variabelen gebruikt om de waarde van één of meerdere van die variabelen te voorspellen. Een voorbeeld is het voorspellen van de geluidproductie van een turbulente luchtstroming in een luchtkanaal op basis van stromingsparameters.² Classificatie heeft betrekking op het onderscheid maken van verschillende klassen in de data, zoals het klassieke voorbeeld van het onderscheiden van honden en katten in beeldmateriaal. Bij optimalisatie wordt gezocht naar de beste oplossing voor een vraagstuk, zoals bijvoorbeeld de optimale aansturing van klimaatinstallaties in gebouwen.³ Voor deze toepassingen zijn er verschillende leermechanismes:

- supervised;
- unsupervised;
- semi-supervised;
- reinforcement.

Bij supervised learning leert een algoritme relaties tussen input en output in de data. Bij unsupervised learning zoekt het algo-

ritme naar de interne structuur van de data, waarbij het verbanden leert uit ongelabelde data. Semi-supervised learning is een combinatie van supervised en unsupervised learning. Reinforcement learning betreft een mechanisme waarbij een "agent", een besluitnemende entiteit, leert een taak uit te voeren binnen een bepaalde omgeving, waarbij een "beloningssysteem" wordt gehanteerd om een optimum hierin te behalen. Figuur 1 toont een schematische representatie van de verschillende vormen van machine learning in algemene zin.



FIGUUR 1: SCHEMATISCHE WEERGAVE VAN DE VERSCHILLENDE VORMEN VAN MACHINE LEARNING

Voor geluidmonitoringdata zijn het voornamelijk modellen op basis van supervised, unsupervised en semi-supervised learning die kunnen assisteren in de analyse ervan. Hierin bestaat ook weer een grote variëteit aan verschillende algoritmes. Afhankelijk van het vraagstuk dient te worden bepaald wat voor soort algoritme toepasbaar is, en vervolgens welk algoritme naar waarschijnlijkheid het beste model biedt binnen de randvoorwaarden van het vraagstuk. In de in dit artikel beschreven praktijkvoorbeelden worden er twee uitgelicht. Machine learning biedt kansen voor automatisering, maar vereist vak kennis op het gebied van zowel het vraagstuk als van de systematiek van machine learning. Een machine learning-algoritme is namelijk niet bekend met de wetten van de natuurkunde, tenzij deze eenduidig in de dataset zijn verankerd. Het is aan de gebruiker om het algoritme zodanig te ontwerpen/programmeren dat de fysische werkelijkheid hierin op een adequate manier is opgenomen.

ALGORITME-ASSISTENTIE BIJ GELUIDANALYSE

Clustering

Bij de monitoring van geluid gedurende langere tijd op een specifieke locatie zijn vaak meerdere bronnen verantwoordelijk voor de optredende geluidniveaus. Met conventionele analysetechnieken is

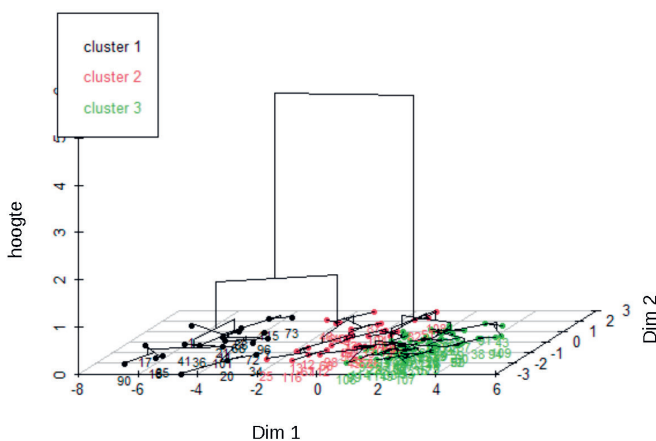
het zeer tijdrovend om de relevante bronnen in grote datasets te duiden en de mate van dominantie van deze bronnen te bepalen. Met machine learning-algoritmes kan via modellen meer informatie op een efficiëntere wijze uit de data worden gehaald. De machine learning-techniek, toegepast bij de analyse van geluiddata, is een unsupervised methode namelijk een clustering-methodiek. Bij clustering analyseert een algoritme de interne structuur en samenhang van een dataset en wordt de dataset ingedeeld in clusters op basis van de onderlinge "afstanden" tussen datapunten. Kleinere afstand tussen datapunten wil zeggen dat er een hogere mate van relatie tussen de datapunten bestaat en de kans dat ze tot eenzelfde cluster behoren groot is. Deze clustering kan vervolgens gebruikt worden om de data te classificeren. Die datapunten zijn als volgt gedefinieerd. In deze toepassing wordt het geluid gerepresenteerd door 1-minuut equivalente, spectrale geluidniveaus. Wanneer er acht octaafbanden in de analyse relevant zijn, zijn de datapunten 8-dimensionaal. Het visualiseren hiervan en het afleiden van verbanden is complex. Een algoritme berekent in een multi-dimensionale ruimte de afstanden op basis van een vooraf gedefinieerde metriek. Zo kan de metriek bijvoorbeeld Euclidisch ($r = \sqrt{\sum_i (x_i - y_i)^2}$) of Manhattan ($r = \sum_i |x_i - y_i|$) zijn.

Clustering kent verschillende algoritmes. Twee veel gebruikte algoritmes zijn k-means clustering en hiërarchische clustering. Bij k-means clustering wordt de data in k clusters ingedeeld zo dat de totale afstand tussen de k middelpunten en de geassocieerde datapunten is geminimaliseerd. Dit betekent dat het algoritme zoekt naar de k middelpunten om de totale afstand te minimaliseren.

Bij het tweede voorbeeld, hiërarchisch clusteren, wordt de data opgedeeld in een hiërarchische structuur van hoofdcluster, subclusters, sub-subclusters, etc. Deze laatste techniek biedt een breder inzicht in de structuur van de data doordat dit gevisualiseerd kan worden met een boomdiagram (dendrogram) die de structuur van de dataset weergeeft.

Geluiddata-analyse met hiërarchisch clusteren

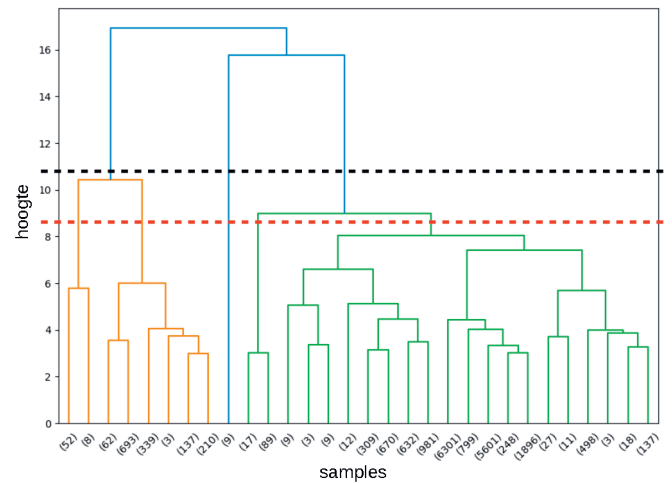
In figuur 2 is een voorbeeld weergegeven van een dendrogram, die de structuur van de data weergeeft, en hoe die zich verhoudt tot een dataset met twee dimensies. In meer dan twee dimensies is dit niet langer zinnig, maar deze representatie geeft goed inzicht in de methodiek.



FIGUUR 2: VOORBEELD VAN EEN DENDROGRAM IN RELATIE TOT EEN 2-DIMENSIONALE DATASET

Bij k-means clusteren moet het aantal clusters iteratief worden bepaald. Bij hiërarchisch clusteren moet het aantal clusters bepaald worden nadat het dendrogram is opgezet. Met behulp van de dendrogramstructuur wordt, in relatie tot het vraagstuk, het aantal clusters bepaald door in het dendrogram een horizontale lijn te trekken die de hoogte, en dus het verschil tussen clus-

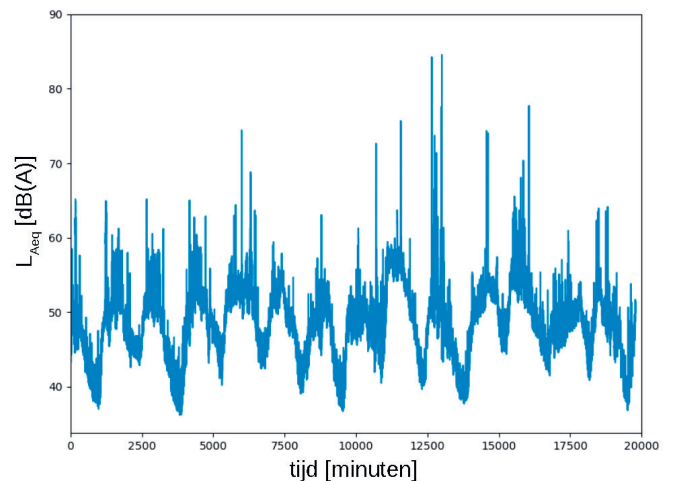
ters, maximaliseert. In figuur 3 is hiervan een voorbeeld gegeven. Er zijn drie hoofdclusters op basis van de data alleen, aangegeven met de zwarte stippellijn. Met de specifieke context in dit voorbeeld (welke hieronder verder toegelicht wordt) kan het detailniveau verhoogd worden naar 5 clusters, weergegeven met de rode stippellijn.



FIGUUR 3: DENDROGRAM OP BASIS VAN GELUIDDATA MET ACHT OCTAAFBANDEN

Praktijkvoorbeeld

Het dendrogram in figuur 3 is samengesteld uit de geluiddata die als equivalent geluidniveau in figuur 4 is weergegeven als een tijdreeks. Deze tijdreeks geeft het fluctuerende karakter van het geluidniveau weer. Deze monitoringsdata betreft een project waar de bijdrage van een bepaalde inrichting op het optredende geluidniveau gezocht wordt.



FIGUUR 4: TIJDREEKS VAN HET EQUIVALENTE GELUIDNIVEAU TER PLAATSE VAN EEN MONITORINGLOCATIE

Ter plaatse van de monitoringlocatie is een nieuw gebouw geïmplementeerd. In de nabijheid ligt een aantal wegen en verkeer draagt dan ook significant bij aan het optredende geluidniveau. Dit bemoeilijkt de bepaling van de bijdrage van geluidbronnen van de nabijgelegen inrichting aan het optredende geluidniveau. Met behulp van een door een algoritme gegenereerd model wordt dit eenvoudiger. Per minuut is het geluidniveau gegeven door de equivalente bijdrage in acht octaafbanden. Een hiërarchisch clustering-algoritme is op de data toegepast, wat resulteert in het eerder getoonde dendrogram (figuur 3). Op basis van de geluidkarakteristieken van de meetlocatie wordt verwacht dat er tenminste drie verschillende bronnen in de meetdata aanwezig zijn:

1. verkeer op de openbare weg;
2. bestemmingsverkeer op de parkeerplaats van de inrichting;
3. installaties van de inrichting.

In het dendrogram is voor drie clusters een horizontale lijn te trekken waar het onderscheid tussen de clusters, gegeven door de totale hoogte in het diagram, maximaal is. Er is echter ook voor vijf clusters een horizontale lijn te trekken waar het onderscheid relatief groot voor is, wat te zien is in verschil in hoogte in het dendrogram. Het aantal snijpunten van de horizontale lijn met de verticale vertakkingen van het dendrogram geeft het aantal clusters. Dit betekent dat er waarschijnlijk twee bronnen in de data te onderscheiden zijn die niet tot de drie verwachte bronnen behoren.

Door de toepassing van het clustering-algoritme kunnen per cluster steekproeven worden genomen in de audiobestanden van de data. In combinatie met de gemiddelde frequentiekenarakteristiek van ieder cluster kunnen de clusters efficiënt gekoppeld worden aan kwalitatieve beschrijvingen van de bron van het geluid in de clusters. Ook kunnen op deze manier eenvoudig de statistieken per cluster (gemiddeld niveau, tijdverdeling enz.) worden bepaald. Daarmee wordt goed inzicht verkregen in de bijdrage van verschillende bronnen in het optredende geluidniveau over langere tijd. In dit praktijkvoorbeeld blijkt dat voor het overgrote deel van de tijd het geluidniveau wordt gedomineerd door verkeer.

TABEL 1: BIJDRAGE AFZONDERLIJKE BRONNEN AAN HET OPTREDENDE GELUIDNIVEAU

	Gemiddeld geluidniveau [dB(A)]			Verdeling [%]		
	dag	avond	nacht	dag	avond	nacht
Verkeer 1	55,1	49,6	46,9	91,0	94,6	91,6
Verkeer 2	55,7	51,1	50,3	7,4	5,2	8,3
Vogels	78,2	-	-	0,1	0,0	0,0
Verkeer naar inrichting	57,4	44,9	45,6	0,5	0,1	0,1
Installaties inrichting	52,0	50,7	43,7	1,0	0,1	0,0

Daarnaast blijkt uit deze analyse dat voor een klein deel van de dag vogels dominant bijdragen aan het geluidniveau. Ook al is dit slechts van korte duur, de geluidbijdrage van de vogels is aanwezig in het frequentiegebied waar praktisch geen A-weging wordt toegepast en kunnen op zeer korte afstand van de monitoringssets komen. Daarbij zijn de pieken die de vogels veroorzaken relatief hoog. Om tot een correcte beoordeling van de bijdrage van de inrichting op het optredende geluidniveau te komen is het gewenst deze pieken buiten de bepaling van het geluidniveau te houden. Zo biedt de toepassing van machine learning dus inzichten in de monitoringdata die met conventionele methoden zeer tijdrovend zouden zijn.

EMISSIE MAPPING

Optimaliseren/zoeken

In het voorgaande voorbeeld wordt de monitoringdata opgedeeld in clusters waarbij verwachte en onverwachte bronnen worden geïdentificeerd. In sommige gevallen is het vooraf duidelijk naar welke bron gezocht wordt in de data. De getalsmatige emissie van deze bron(nen) dient in kaart te worden gebracht. Voor dergelijke vraagstukken kan een optimalisatie/zoek-algoritme een oplossing bieden. Er bestaan verschillende optimalisatie/zoek-algoritmes. Twee voorbeelden zijn genetische algoritmes en simulated annealing algoritmes.

Genetische algoritmes zijn geïnspireerd op biologische evolutie, waarbij een populatie zich langzaam aanpast om zo goed mogelijk met een omgeving compatibel te zijn. Een oplossing voor een vraagstuk is gedefinieerd als een chromosoom. Door verschillende oplossingen binnen een populatie op basis van fitheid voor het specifieke vraagstuk te selecteren en te kruisen, en door wille-

keurige mutaties door te voeren, convergeert de oplossing naar een optimum. Doordat voor de hele populatie de fitheid per oplossing moet worden berekend is het een computationeel zwaar algoritme, terwijl dit tegelijkertijd ook de kracht van het algoritme weergeeft.

Het tweede voorbeeld, simulated annealing, is geïnspireerd op een fysisch proces in metallurgie. Hierbij wordt een metaal met hoge temperatuur heel langzaam afgekoeld. Hierdoor krijgen de metaalatomen de kans hun energetisch optimale positie te vinden en resulteert een metaal met een mooi kristalrooster. Een soortgelijk proces kan worden uitgevoerd om een optimale oplossing te vinden voor een vraagstuk. Beginnend vanuit een willekeurige oplossingsconfiguratie wordt bij iedere iteratie een nieuwe configuratie doorgerekend. Als deze configuratie een betere oplossing voor het vraagstuk geeft wordt de configuratie geaccepteerd als de nieuwe toestand. Als deze configuratie een slechtere oplossing voor het vraagstuk geeft wordt de configuratie geaccepteerd met een waarschijnlijkheid $p = e^{-dQ/T}$, waarin dQ het verschil geeft in de fit voor de huidige configuratie en de nieuwe configuratie en T de rekentemperatuur is. Deze temperatuur is geen werkelijke temperatuur, maar puur een mathematische variabele. Door de temperatuur langzaam af te laten nemen convergeert de oplossing naar een optimum. Doordat in deze methode voor iedere iteratie slechts voor een enkele oplossing de fitheid hoeft te worden bepaald vergt deze methode minder rekenkracht in vergelijking met genetische algoritmes. Om deze reden wordt deze methode verder uitgelicht in dit artikel.

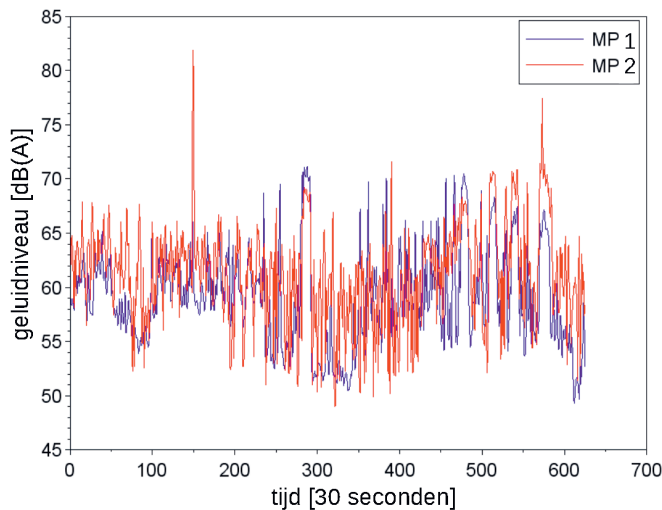
Geluidemissie mapping met simulated annealing

Door op verschillende posities op afstand van een geluidbron het geluidniveau te monitoren kan met behulp van een simulated annealing algoritme de geluidemissie van de betreffende bron in een studiegebied worden bepaald, op een semi-supervised manier. In een vereenvoudigde situatie kan de geluidafname van bron tot meetpunt worden berekend met $L_p = L_{wr} - 20 \log r - 9,1 - 0,005r$ waarbij r de afstand tussen bron en meetpunt is.⁴ In het model wordt de bijdrage van de bron(nen) energetisch gesommeerd met het instantane achtergrondniveau. Het algoritme kan op verschillende manieren worden toegepast. Het kan worden toegepast door voor meerdere bronnen die gelijktijdig in bedrijf zijn de locatie en het bronvermogen te berekenen. Het kan ook worden toegepast door voor bronnen met bekende locatie het bronvermogen dat fluctueert instantaan te bepalen, alsmede een fluctuerend instantaan achtergrondniveau. Deze laatste optie zal hier verder worden toegelicht.

Praktijkvoorbeeld

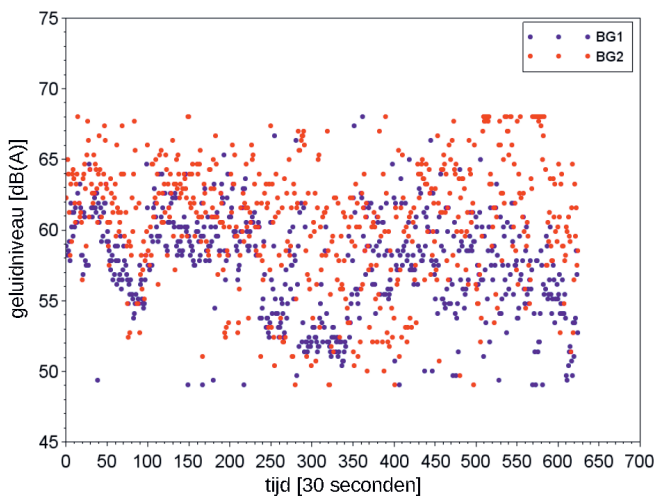
In dit voorbeeld wordt de geluidbelasting van heiwerkzaamheden op een bouwplaats ter plaatse van woningen in de nabije omgeving gemonitord. Door omstandigheden is het niet mogelijk immisiemetingen bij de woningen zelf uit te voeren. De toepassing van een algoritme biedt via extrapolatie de mogelijkheid om inzicht te krijgen in de optredende niveaus nabij de woningen. Aan beide zijden van de bouwplaats is een meetset opgesteld. In figuur 5 is het geluidniveau gedurende een dag weergegeven in 30 seconden equivalenten. Het geluidniveau heeft een sterk fluctuerend karakter en puur visueel is het niet evident wat de bijdrage van heiwerkzaamheden is.

De hei-installaties zijn in dit project uitgerust met een GPS-logger. Hierdoor is de locatie van de hei-installaties - ondanks verplaatsingen - bekend. Door de meetdata van beide meetlocaties te combineren kan een algoritme zoeken naar het fysische model dat de hoogste mate van consistentie geeft. Dit fysisch model bestaat uit de bronvermogens van de hei-installaties (0 als het algoritme geen heiwerkzaamheden constateert) en de fluctuerende achtergrondniveaus ter plaatse van de meetlocaties.

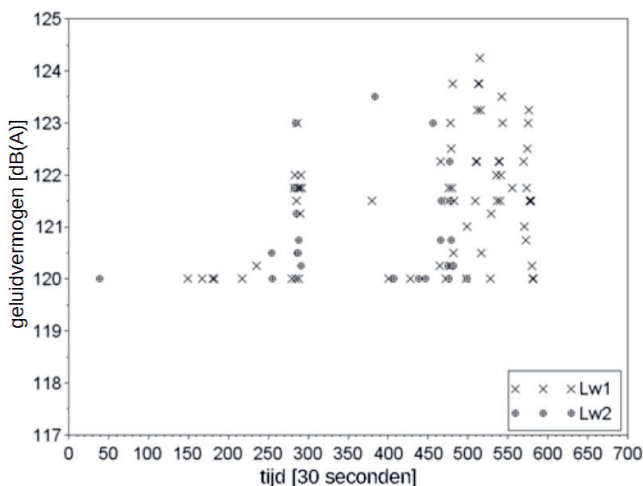


FIGUUR 5: GELUIDNIVEAU TER PLAATSE VAN TWEE MEETLOCATIES

Het simulated annealing algoritme kiest voor iedere iteratie een nieuwe configuratie van bronvermogens en achtergrondniveaus, en op basis van een fitness-check wordt deze configuratie met een probabilistisch karakter geaccepteerd als de nieuwe toestand. Door de kans van acceptatie van "slechte" configuraties af te laten nemen, verkent het algoritme in het begin een groot deel van de oplossingsruimte, maar convergeert de oplossing aan het einde van de rekencyclus. In figuur 6 en 7 zijn de gevonden achtergrondniveaus en bronvermogens weergegeven.

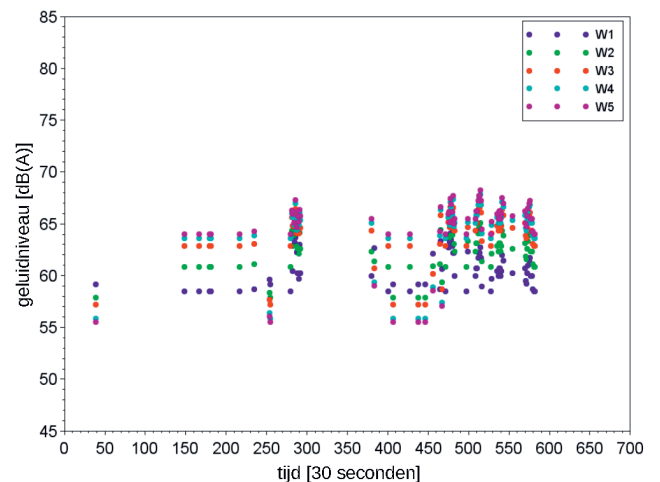


FIGUUR 6: ACHTERGRONDNIVEAU TER PLAATSE VAN BEIDE MEETLOCATIES



FIGUUR 7: BRONVERMOGENS HEI-INSTALLATIES

Met de gevonden activiteit en bronvermogens van de hei-installaties kan vervolgens door extrapolatie berekend worden wat de geluidbelasting specifiek ten gevolge van de heiwerkzaamheden bij de woningen is. In figuur 8 zijn de geluidniveaus ten gevolge van heiwerkzaamheden bij vijf woningen in de nabijheid van de bouwplaats weergegeven.



FIGUUR 8: GELUIDNIVEAUS BIJ WONINGEN SPECIFIEK TEN GEVOLGE VAN HEIWERKZAAMHEDEN

Met deze methodiek, waarbij meetdata, algoritmes en modellen worden gecombineerd, kan voor een groot aantal posities de geluidemissie ten gevolge van een specifieke bron bepaald worden, wat met metingen alleen niet haalbaar is.

CONCLUSIE

Door kennis van geluidmonitoring en machine learning-technieken te combineren is het mogelijk om met modellen meer informatie af te leiden uit geluidmonitoringdata. Er bestaat een grote variëteit aan machine learning-algoritmes. Verschillende soorten vraagstukken vragen om een verschillende aanpak. In dit artikel is een aantal methoden toegelicht. Met praktijkvoorbeelden is de kracht van de toepassing van machine learning-algoritmes geïllustreerd. Naar verwachting zal in de nabije toekomst toepassing van machine learning voor analyse van geluidmonitoringdata een steeds belangrijkere plaats innemen. Door het toenemen van rekenkracht zal het in de toekomst in toenemende mate mogelijk zijn om dergelijke analyses realtime tijdens het monitoringsonderzoek uit te voeren. Op deze manier kan bijvoorbeeld geluidmonitoring worden uitgevoerd over een groter gebied dan - zonder toepassing van machine learning- op basis van de beschikbare meetsets mogelijk is.

REFERENTIES

1. M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc. (1997).
2. R.J. Dikken, P.B. van Luijpen, Voorspellen van luchtstrominggeluid met CFD en machine learning, TVVL Magazine 6, december 2020.
3. R.J. Dikken, Machine learning en data science voor klimaatinstallaties, TVVL Magazine 7, november 2019.
4. "Handleiding meten en rekenen Industrielawaai" (HMRI 1999).