



/HUB

enerzjyearmoed

ûnderzoeksopzet

voorspelmodel



INLEIDING

Voor deze /HUB heeft DataFryslân een model ontwikkeld om de energiearmoede in Friesland te kunnen voorspellen. Hierbij wordt gebruikgemaakt van de definitie van energiearmoede zoals die eerder door TNO¹ is opgesteld. Deze definitie is als volgt: Laag Inkomen en Hoge Energierekening (LIHK) en/of Laag inkomen, Lage Energetische Kwaliteit (LILEK) – met Laag Energetische Kwaliteit wordt bedoeld de mate waarin de betreffende woning adequaat is geïsoleerd. Hierbij volgt DataFryslân het kader van bestaand onderzoek over energiearmoede. In de rest van dit document zal aan deze definitie gerefereerd worden als (*LILEK en/of LIHK*).

Het doel van het model is om op buurtniveau een voorspelling te kunnen geven van de kans in een bepaalde buurt op energiearmoede. Aan de hand van deze gegevens krijgen beleidsmedewerkers en energiecoaches een fijnmaziger beeld van waar energiearmoede zich voordoet en welke factoren van invloed zijn op het ontstaan ervan.

Dat er gebruik wordt gemaakt van een voorspelmodel in plaats van de daadwerkelijke *LILEK en/of LIHK* heeft te maken met de privacywetgeving; lage aantallen van microdata mogen niet geëxporteerd worden uit de CBS-omgeving, omdat de data hiermee te herleiden zouden kunnen zijn tot individuele personen. Aangezien veel buurten in Friesland, zeker in het buitengebied, dunbevolkt zijn, zou dit betekenen dat er voor het merendeel van deze buurten geen data kunnen worden weergegeven. Een voorspelmodel vangt dit af, omdat er met een schatting gewerkt wordt in plaats van met de feitelijke data, en er dus geen privacy risico bestaat.

MODEL

Het voorspelmodel is getraind op basis van CBS-microdata van heel Friesland. Dit zijn data op huishoudniveau van alle Friese huishoudens. *LILEK en/of LIHK* is door het CBS gecodeerd in de microdata als 0 (geen energiearmoede) of 1 (wel energiearmoede). DataFryslân heeft een model getraind dat *LILEK en/of LIHK* voorspelt aan de hand van 23 andere variabelen in de microdata². De 10 variabelen die tot de beste voorspellingen leiden zijn:

¹ <https://www.tno.nl/nl/over-tno/nieuws/2021/9/tno-brengt-energiearmoede-gedetailleerd-in-kaart/>

² Nb. Maatregelen gericht op het isoleren van woningen zijn ook van belang voor het tegengaan van energiearmoede. Met name huishoudens die niet in staat zijn om isolatie maatregelen te nemen zijn meer geneigd om energiearmoede te ervaren.

1. oppervlakte van de woning in m ²
2. gemeentes in Fryslân, grenzen van 2022
3. woningtype (vb. hoek- of tussenwoning)
4. type eigenaar (vb. huur of koop)
5. bouwjaar van de woning
6. WOZ waarde
7. stedelijkheid van de buurt
8. huishoudsamenstelling (vb. éénouder- of meerouder gezin)
9. energielabel
10. aantal personen in het huishouden

Dit voorspelmodel geeft de kans op energiearmoede per huishouden op in procenten, geaggregeerd op buurtniveau. Zo kan deze informatie niet alleen gebruikt worden om in kaart te brengen in welke buurten energiearmoede waarschijnlijk het grootste probleem is, maar is ook te zien welke factoren de grootste bijdrage leveren aan dit probleem. Daardoor kan effectiever worden ingezoomd op die huishoudens waar de nood het hoogst is.

MODEL PERFORMANCE

De accuratesse van het model is niet zozeer bepalend voor het beoordelen van de kwaliteit van het model, aangezien we een ongebalanceerde dataset hebben met relatief weinig huishoudens die energiearmoede ervaren. Ongeveer 6% procent van de Friese huishoudens in onze dataset hebben energiearmoede. Belangrijk is dat het model correct energiearmoede voorspelt (een '1') en niet wanneer huishoudens geen energiearmoede ervaren (een '0'). Een zeer accuraat model dat elk huishouden classificeert zonder energiearmoede zal een hoge accuratesse hebben. Een voorbeeld: als 97 huishoudens geen energiearmoede hebben en 3 wel dan zal een model die 100 huishoudens classificeert als niet-energiearm een accuratesse hebben van 97%. De kracht van ons model is met name te zien bij de sensitiviteitswaarde. Het model heeft een sensitiviteit van 95%, wat betekent dat 95% van de huishoudens met energiearmoede door het model correct zijn geïdentificeerd. Het model voorspelt dus in 95% van de gevallen correct een '1'.