Algoritme

Korte beschrijving van het ETL en AI proces.

Naam *De naam die gebruikt wordt om het algoritme aan te duiden.*	Analyse Woning Data
Organisatie *De volledige naam van de organisatie waar het algoritme ingezet wordt.*	DataFryslân
Korte omschrijving *Een korte beschrijving van het algoritme.*	Een deel van de Fryske woningvoorraad is verouderd. De vraag is hoe groot dat deel is en om wie het gaat. Er zijn datasets gemerged en via een beslisboom is de definitie van duurzame woningvoorraad in de data berekend.
ETL (extract, transform, load) proces? *Betreft het een toepassing om de data te transformeren ten behoeve van extractie (denk aan aantallen of percentages), dan kan dat hier kort worden toegelicht.*	In de Remote Access omgeving van het CBS is data van de Woonbase en Energiearmoede gemerged. Vervolgens zijn, na het maken van de definitie variabelen (zie later), aantallen geexporteerd op basis van locatie (provincie, gemeente of buurt) naar variabelen.
Type algoritme *Is het algoritme zelflerend? In een niet-zelflerend algoritme specificeert de mens de regels die de computer moet volgen. Als het een zelflerend algoritme is, leert de machine over de patronen in de data.*	Simpele beslisboom om categorieën in variabelen te genereren
Methoden en modellen *Standaardmethoden of modellen die het algoritme gebruikt.*	Een standaardmethode in R om variabelen te genereren met “if-else” statements

Beleidsterrein *Trefwoorden over het beleidsterrein waarin het algoritme wordt ingezet.*	• Sociaal • Wonen • Leefomgeving • Duurzaamheid
Status *De status van het algoritme: in ontwikkeling, in gebruik, of buiten gebruik.*	Gebruikt voordat het ETL proces begon.
Doel *Het doel waarvoor het algoritme ontwikkeld is en/of hoe de inzet ervan bijdraagt aan het behalen van die doelen.*	Doel van de beslisboom was om te bepalen of een woning aanpasbaar/kansrijk of energetisch duurzaam is. Dit is gedaan aan de hand van woning technische gegevens: oppervlakte, bouwjaar, energie labels en energetische kwaliteit.
Impact *De impact van het algoritme op burgers en bedrijven. Bijvoorbeeld: hoe werkt het algoritme en wat zijn de verwachte consequenties daarvan voor het individu of bedrijf?*	Alle data die vanuit de beschermde omgeving van het CBS worden geëxporteerd worden gecontroleerd door het CBS. Door de waarborging van het CBS op de data output, is de impact van het algoritme op burgers daarom zeer klein, zeker gezien dat de gebruikte gegevens niet tot personen te herleiden zijn.
Proportionaliteit *Een afweging van de voor- en nadelen van de inzet van het algoritme en waarom dit redelijk gerechtvaardigd is.*	Omdat de gepresenteerde waarden niet tot personen of objecten te herleiden vallen, zijn er geen nadelen op dit vlak. Het voordeel is dat beleidsmedewerkers op een laag geografisch niveau informatie krijgen de toekomstbestendigheid van woningen in de provincie, gemeente en buurt.
Menselijke tussenkomst *Een omschrijving van hoe uitkomsten van het algoritme door een mens gecontroleerd en bijgesteld (kunnen) worden.* *Op een schaal van 1 (volledig geautomatiseerd menselijk toezicht) t/m 5 (volledig menselijk toezicht).*	De cutoff points van de beslisboom zijn bepaald door de onderzoekers en gebaseerd op bestaande literatuur.

Monitoring *Een overzicht van hoe de inzet van het algoritme wordt gemonitord.* *Op een schaal van 1 (geen monitoring) t/m 5 (volledige monitoring).*	De geautoriseerde onderzoekers van DataFryslân hebben zicht op de uitkomsten, het script en wat ermee gedaan wordt. Daarnaast heeft het CBS ook zicht op het model en de resultaten doordat alle data en scripts door de output controle moet.
Risico *Een overzicht van de voorziene risico's bij de inzet van het algoritme.* *Op een schaal van 1 (zeer laag privacy risico) t/m 5 (zeer hoog privacy risico).*	Het risiconiveau van de inzet van dit algoritme zijn beoordeeld als 'laag' omdat er een onafhankelijke organisatie tussen zit die expliciet de data en scripts controleert.
Prestatienormen *Een omschrijving van de verwachte prestaties van het algoritme en hoe die worden gemeten.*	Woningen zijn gedefinieerd op basis van beslisregels. Dit is een deterministisch proces. De beslisboom werkt derhalve zoals verwacht goed gezien de definities van toekomstbestendige woningen.
Databronnen *Een overzicht van de databronnen die op dit moment gebruikt worden door het algoritme en/of in het begin gebruikt zijn bij het maken van het algoritme.*	CBS microdata: Woonbase, Energiearmoede maatwerkbestand

Heeft u vragen? U kunt altijd mailen naar info@datafryslan.nl over de gebruikte AI of ETL scripts.