Auteur	[redacted]
Datum	11-12-2020
Versienummer	1.0
Status	Definitief
Product	20112 – Voorspellend model problematische schulden Kinderopvangtoeslag
Onderdeel van verwerking	M7842
Onderdeel van GEB	GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'. Deze deel GEB is een uitwerking van de selectie van ouders voor thema B Persoonlijke begeleiding op basis van data-analyse.

A1. Doel product	1. <u>Selecteren van ouders</u> Hoofddoel van het product is het selecteren van ouders met een verhoogd risico op problematische toeslagschulden KOT. Voor burgers met: - een grote openstaande schuld KOT - en met een lopende beschikking KOT - en waarbij nog geen sprake van dwanginvordering voor de openstaande schulden KOT. Hiervoor is een model opgesteld waarmee wordt voorspeld welke ouders over een half jaar de grootste kans hebben om in betalingsproblemen te komen als gevolg van hoge terugvorderingen bij de KOT. 2. <u>Trainen van het model</u> Om een goede voorspellende werking van het model te kunnen blijven garanderen, wordt het model periodiek opnieuw getraind.
A3. Voorgenomen gegevensverwerkingen	Er is een model gemaakt m.b.v. training sets. Elke training set bevat gegevens (zie onder A2) van de burger op een bepaald tijdstip, en het feit of deze burger een half jaar later een dwanginvordering voor de openstaande schuld KOT opgelegd heeft gekregen. Op basis van deze trainingsets wordt een model gemaakt met een aantal **finale indicatoren** die voorspellen of een burger een half jaar later in betalingsproblemen komt voor KOT. Vervolgens wordt bij het draaien van het model de huidige waarden van de **finale indicatoren** van het model berekend en wordt het model toegepast zodat voorspeld wordt wie de grootste kans heeft om over een half jaar in

	betalingsproblemen te komen voor KOT. De X aantal ouders met de grootste kans hierop worden geleverd aan Handhavingsregie. Bij de ouders met de grootste kans op problematische toeslagschulden KOT wordt vervolgens Persoonlijke begeleiding aangeboden. Deze verwerking staat beschreven in GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'. Omdat de indicatoren die goed voorspellen of een burger een half jaar later in betalingsproblemen komt voor KOT kunnen wijzigen in de loop van de tijd, is het noodzakelijk om het model regelmatig opnieuw te trainen met nieuwe trainingssets. De uitkomst hiervan kan zijn dat de selectie van de **finale indicatoren** in de toekomst zal wijzigen. Om een goede voorspellende werking van het model te kunnen blijven garanderen, is daarom de volledige set aan hieronder genoemde indicatoren nodig voor het opnieuw trainen van het model. Let op: wijzigingen t.o.v. de vorige versie van het model: - Bij het draaien van het nieuwe model zijn de 8 finale indicatoren hard geprogrammeerd en worden niet uit de datapreparatie van het Risicoselectiemodel gehaald. - De velden van de trainingssets komen deels uit een extract van de datapreparatie van het Risicoselectiemodel en deels uit het datafundament. Zie voor een verdere uitwerking hiervan 'Trainen van het model' in sectie A8.		
A2. Persoonsgegevens	**Gewoon**	**INDICATOREN** Let op: **blauwe velden** zijn de **finale indicatoren**, de overige velden zijn potentiele indicatoren die nu niet in het voorspellend model worden gebruikt.	**Functionele vertaling indicatoren**
		aant_kinderen_in_opvang	**Aantal kinderen waarvoor KOT is aangevraagd per BSN van de ouder.**
		aant_nihilstelling	Aantal keer dat BSN voor KOT op nul gezet is.
		afst_adr_opvang	Afstand in kilometers tussen huisadres aanvrager en adres van de kinderopvang volgens LRK.
		av_aant_kot_besch	Aantal beschikkingen in dit toeslagjaar
		av_aant_verhuiz_opgegeven	Aantal verhuizingen opgegeven in TVS.
		av_aant_verhuizingen_gba	Aantal verhuizingen volgens de BVR.

		av_adr_aant_volwassenen	Aantal ingeschreven volwassenen per adres volgens de BVR.
		av_dagtaaktype	Dagtaaktype van de aanvrager zoals bekend in TVS.
		AV_doelgrp_oms_KOT	Doelgroep omschrijving aanvrager KOT
		av_groei_inkomen	Groei van het jaarinkomen (inkomen TVS tov inkomen AIG).
		av_leeftijd	Leeftijd van de aanvrager volgens BVR
		av_missend_voorgaande_ink	Indicatie of inkomen missend is in AIG
		av_nul_inkomen_opgegeven	Indicatie of opgegeven inkomen in TVS nul is.
		av_openst_vord_rood	**Bedrag aan openstaande toeslagen-vorderingen per BSN.**
		av_recht_op_kgb	Indicatie of aanvrager of partner recht heeft op KGB voor alle kinderen volgens informatie van SVB.
		av_verschil_verhuizingen	Verschil tussen aantal in TVS opgegeven verhuizingen en aantal verhuizingen volgens de BVR.
		av_woont_niet_in_nlbedu	Indicatie of aanvrager in Nederland, Belgie of Duitsland woont (1 = niet 0 = wel).
		Bewindvoerder_JN	Volgens BVR is er sprake van een bewindvoerder Indicatie of BSN eerder een onvoldoende
		boete_beschikkingen	informatie heeft aangeleverd of slechte schatter zijn gebleken (1 = ja; 0 = nee).
		Briefadres_Gemeente_JN	Volgens BVR is er sprake van een briefadres
		BSN_Open_Bedr_KOT_t1	**Openstaand bedrag aan vorderingen KOT**

		Eigen_bijdrage_per_maand	**Verschil tussen aangevraagde uren * aangevraagd tarief - de toegewezen toeslag.**
		einddatum_bekend	Indicatie of einddatum van de toeslagaanvraag bekend is in TVS (1=ja; 0=nee)
		Feitelijk_Woonadres_JN	Volgens BVR is er sprake van een feitelijk woonadres (vastgesteld door Belastingdienst)
		heeftpartner	Is er sprake van een toeslagpartner?
		ind_bso_go	Indicatie of toeslagaanvraag betrekking heeft op buitenschoolse opvang door een gastouder (1=ja; 0=nee)
		ind_bso_kc	Indicatie of toeslagaanvraag betrekking heeft op buitenschoolse opvang door een kindercentrum (1=ja; 0=nee)
		ind_do_go	Indicatie of toeslagaanvraag betrekking heeft op dagopvang door een gastouder (1=ja; 0=nee)
		kind_woont_bij_av	Indicatie of het kind waarvoor KOT wordt aangevraagd volgens de BVR woonachtig is bij de aanvrager (1 = wel bij aanvrager; 0 = niet)
		Kot_Huur_Zorg_Kgb	Is er sprake van KOT/ Huur/ Zorg/ KGB. Bv. 1 1 0 0 ==> KOT en Huur, 1 0 1 0 ==> KOT en Zorg
		LastDT_Minus_LastVT_Tmin2_Kot	Verschil tussen laatste DT KOT en laatste VT KOT, 2 jaar geleden
		LastDT_Minus_LastVT_Tmin3_Kot	**Verschil tussen laatste DT KOT en laatste VT KOT, 3 jaar geleden**
		leeft_jongste_kind	**Leeftijd van het jongste kind waarvoor KOT wordt aangevraagd volgens BVR.**
		Leeftijd_bij_geboorte	Leeftijd van de aanvrager volgens de BVR op het moment van de toeslagaanvraag.

		maximale_status_vord_rood	**Status waarin de toeslag-vordering zich bevindt. Hoogste status vanaf 2014** 1 ind_bet = CBT-regeling 2 ind_bet = uitstel_bezwaar OF ind_nietbet = uitstel 7 beslagopdracht, DB verzonden, DB betekend 9 ten uitvoerlegging 10 ind_nietbet = oninbaar 5 rest
		Omschrijving_45	Classificatie van hoeveel inwoners van betreffende gemeentecode
		Omschrijving_45_31	Idem voor BEWINDVOERDER-ADRES
		Omschrijving_45_7	Idem voor FEITELIJK WOONADRES
		Omschrijving_45_9	Idem voor BRIEFADRES VOLGENS GEMEENTE
		Omschrijving_47	Classificatie van zeer stedelijk tot niet stedelijk van betreffende gemeentecode
		Omschrijving_47_31	Idem voor BEWINDVOERDER-ADRES
		Omschrijving_47_7	Idem voor FEITELIJK WOONADRES
		Omschrijving_47_9	Idem voor BRIEFADRES VOLGENS GEMEENTE
		Partn_doelgrp_oms_KOT	Doelgroep omschrijving toeslagpartner KOT
		perc_ink_eigen_bijdrage	**Eigen bijdrage per jaar t.o.v. draagkracht huishouden**
		Te_Jong_Voor_BSO	Indicatie of kind waarvoor toeslag voor buitenschoolse opvang is aangevraagd te jong is voor recht op toeslag voor buitenschoolse opvang.
		Te_Oud_Voor_BSO	Indicatie of kind waarvoor toeslag voor buitenschoolse opvang is aangevraagd te oud is voor recht op toeslag voor buitenschoolse opvang.

		Te_Oud_Voor_DO	Indicatie of kind waarvoor toeslag voor dagopvang is aangevraagd te oud is voor recht op toeslag voor dagopvang.
		terugwerkende_kracht	Aantal dagen tussen datum waarop toeslag is aangevraagd en de ingangsdatum van het tijdvak waarop de aanvraag betrekking heeft.
		trainingsetnummer	Nummer van de trainingset
		uitval_kot	Aantal uitvalmeldingen in TVS voor de actuele toeslagaanvraag KOT.
		zelfde_kind_do_en_bso	Indicatie of voor hetzelfde kind over hetzelfde tijdvak zowel toeslag voor dagopvang als voor buitenschoolse opvang aangevraagd wordt.
		INNINGSGEGEVENS Openstaande toeslagschulden van de burgers en de fase van de inning. Dit bestand met inningsgegevens zal na ontvangst worden vereenvoudigd zodat de soort schuld niet meer herkenbaar is. Er is dan alleen nog maar te zien of het een rode of een blauwe schuld betreft. Het oude bestand met inningsgegevens zal na de vereenvoudiging worden verwijderd van SASGRID. Op de Q-schijf staan de oorspronkelijke bestanden op een beveiligde omgeving.	
	Bijzonder	-	
	Strafrechterlijk	-	
	Wettelijk identificerend	BSN	
A4. Verwerkingsdoeleinden	Selectie van ouders met een verhoogd risico op problematische toeslagschulden KOT. Bij de ouders met de grootste kans op problematische toeslagschulden KOT wordt vervolgens Persoonlijke begeleiding aangeboden. Deze		

	verwerking staat beschreven in GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.		
A5. Betrokken partijen	**Verwerkings-verantwoordelijke**	Minister van financiën	
	Verwerker	-	
	Verstrekker	- Gegevens in het datafundament, SAS-GRID en open source CBS-data - BCA/BICC team Inning verstrekt wekelijks een bestand met inningsgegevens (zie A2 laatste regel)	
	Ontvanger	IV en Handhavingsregie	
A6. Belang van de gegevensverwerking	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.		
A7. Verwerkingslocaties	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.		
A8. Techniek en methode gegevensverwerking	De opwerking van de data gebeurt op Teradata en SASGRID. Deze gebieden zijn alleen te benaderen door geautoriseerde data-analisten van Toeslagen. 1. Selecteren van ouders Elke maand worden de in A2 beschreven **finale indicatoren** als onderdeel van het voorspellend model berekend per burger in SAS-EG. Vervolgens wordt het reeds ontwikkelde voorspellend model gedraaid in SAS-MINER op basis van deze **finale indicatoren**. Output van dit model wordt geordend op voorspeld risico (van hoog naar laag risico). De X aantal aanvragers (per BSN) met de grootste kans op problematische schulden KOT worden in een Excel bestand per email aan HHR geleverd. 2. Trainen van het model Het voorspellend model is gemaakt in SAS-MINER en wordt periodiek getraind. De gegevens die hiervoor gebruikt worden zijn beschreven onder A2. Het voorspellend model problematische schulden maakt slechts gebruik van de datapreparatie van het Risicoselectiemodel		

	Toeslagen (RSM). Dit gebeurt alleen bij het opstellen van de trainingscases voor het voorspellend model problematische schulden (en niet bij het draaien van het model): Trainingscases van het voorspellend model problematische schulden (met gegevens uit een bepaald moment uit het verleden en het feit of er sprake is van dwanginvordering een half jaar later) worden opgesteld m.b.v.: - Geprepareerde gegevens voor het RSM (uitgevoerd op hetzelfde moment in het verleden). Niet de gehele datapreparatie van het RSM wordt hiervoor gebruikt, maar een extract hieruit. - Andere gegevens afkomstig uit datafundament en CBS-gegevens. Vóórdat de scoringsmodule van het RSM wordt gedraaid, vindt datapreparatie plaats (deel 1 van het RSM waarin nog niet gescoord wordt). Hierbij worden allerlei gegevensbronnen gecombineerd, wat leidt tot een groot aantal velden. Dit zijn de gegevensbronnen van het RSM (beschreven in de GEB van RSM). Deze velden worden elke maand opgeslagen op SASGRID. Een aantal van deze opgeslagen velden (niet allemaal) worden ook gebruikt voor het trainen van het voorspellend model problematische schulden. Bij de trainingssets wordt bij alle records het veld BSN vervangen door versleutelde gegevens. De overige velden van de trainingssets kunnen niet worden herleid naar personen.	
A9. Juridisch en beleidsmatig kader	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.	
A10. Bewaartermijnen	De trainingsets worden 3 jaar op Teradata bewaard om een zo uitgebreid mogelijke set aan trainingscases te hebben waarmee het model wordt getraind. De outputbestanden worden 3 jaar op SASGRID bewaard t.b.v. evaluatiedoeleinden.	

B11. Rechtsgrond	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.	
B12. Bijzondere persoonsgegevens	n.v.t.	
B13. Doelbinding	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.	
B14. Noodzaak en evenredigheid	**Proportionaliteit** (Hebben we de gegevens de we gebruiken nodig om het gestelde doel te bereiken?) 1. <u>Selecteren van ouders</u> De **finale indicatoren** die worden gebruikt om burgers te selecteren, zijn de gegevens die het beste voorspellen of bij ouders in de toekomst problematische toeslagschulden KOT zullen ontstaan. Het voorspellend vermogen van het model wordt aanzienlijk minder als één of meer van deze indicatoren niet wordt gebruikt. In het verleden hebben we ook gezien dat bij het verwijderen van indicatoren uit privacy overwegingen (wanbetaler zorg en blauwe schulden) de voorspellende werking minder werd. 2. <u>Trainen van het model</u> De volledige set aan potentiële indicatoren die wordt gebruikt om het model te trainen, is nodig om te voorkomen dat causale verbanden die nu niet aanwezig zijn maar die in de toekomst gaan ontstaan niet worden onderkend bij het voorspellen van problematische toeslagschulden KOT. **Subsidiariteit** (Kun je hetzelfde doel bereiken met andere gegevens, of andere cohorten(bijvoorbeeld: veel, minder, weinig) van gegevens?) De gegevens zijn beoordeeld op subsidiariteit. Als gevolg hiervan wordt het input bestand met belastingschulden versimpeld, zie ook de laatste alinea van A2.	
B15. Rechten van de betrokkene	Zie GEB M7842 'Rechtmatig toekennen voorschot kinderopvangtoeslag door proactief signaleren'.	

C16. Risico's	1. De output wordt via de interne mail gestuurd aan Handhavingsregie (Excel-bestand zoals beschreven in A8). Een mogelijk risico hierbij is dat het Excel-bestand door Handhavingsregie niet op de juiste manier bewaard of verwijderd wordt. Met als mogelijk gevolg dat niet-bevoegde mensen toegang krijgen tot deze gegevens. (De bestanden worden trouwens ook op SASGRID opgeslagen) 2. Gebruik van postcodegebieden om te bepalen of een ouder in een stedelijk of landelijk gebied woont, zou bij het toepassen als finale indicatoren via een omweg tot een bias voor bepaalde postcodegebieden kunnen leiden. (Momenteel zijn dit geen finale indicatoren, maar potentiële indicatoren.) 3. Burgers wordt voorafgaand aan Persoonlijke begeleiding gevraagd of zij begeleiding willen ontvangen. Burgers die begeleiding ontvangen, hebben op elk moment in het proces de mogelijkheid om aan te geven dat zij de begeleiding willen stoppen. Burgers hebben echter geen mogelijkheid om aan te geven dat hun data niet mag worden gebruikt voor het trainen van het model. 4. Door het gebruik van trainingssets om het model te trainen bestaat er een risico dat er een bias ontstaat voor een bepaalde deelpopulatie binnen de doelgroep. Er kan bijvoorbeeld bias ontstaan als trainingsets voornamelijk posten bevatten die in het verleden geselecteerd waren als hoog risico en daarom beoordeeld werden. Deze trainingssets zijn dan geen weerspiegeling van de gehele populatie.	
D17. Maatregelen	1. Er wordt aangesloten op de richtlijnen van Analytics United. Zie bijlage Richtlijnen Analytics United, d.d. 3-12-2020.	

	2. Bij elk besluit om de finale indicatoren te wijzigen, zal samen met de privacy officer opnieuw moeten worden beoordeeld of deze combinatie van indicatoren nog steeds aan alle privacy eisen voldoet en niet tot een bias leidt. 3. De trainingssets worden gepseudonimiseerd waarbij bij alle records het veld BSN wordt vervangen door versleutelde gegevens. De overige velden van de trainingssets kunnen niet worden herleid naar personen. 4. In dit voorspellend model zijn de trainingsets wél een weerspiegeling van de populatie. De volledige populatie (zie onder A1) wordt op diverse tijdstippen in het verleden bepaald. En vervolgens wordt gekeken of deze burgers een half jaar later een dwanginvordering voor kinderopvangtoeslag hebben gekregen. De trainingssets zijn daardoor dus niet gebaseerd op een deelpopulatie binnen de doelgroep, maar gebaseerd op de gehele populatie. Hierdoor is de kans op bias geminimaliseerd.	

Advies Integrale Beveiliging & Privacy
Akkoord met deze deel GEB. Alle privacyrisico's zijn voldoende in beeld gebracht en geadresseerd. Vanuit Integrale Beveiliging en Privacy geen verder advies aan het MT. De deel GEB wordt als bijlage opgenomen bij de VB KOT GEB. AVG Registernummer M7842: Rechtmatig toekennen voorschot Kinderopvangtoeslag door proactief signaleren. De M7842 GEB wordt momenteel herijkt onder coördinatie van [redacted].

Bijlage Richtlijnen Analytics United, d.d. 3-12-2020

1) Het Excel-bestand wordt met een wachtwoord beveiligd.

2) De Ontvanger moet eerst verklaren dat ze de gegevens niet verder zullen verspreiden of op een ongeoorloofde plaats opslaan. Dit moet altijd gedaan zijn voordat ontvanger toegang krijgt tot het bestand.

3) De oplevermail zal versleuteld en met de afleveropties 'Kopieren voorkomen' en 'Onderwerp markeren als vertrouwelijk' en 'Coderen' worden verstuurd zodat de mail niet inclusief bijlage kan worden doorgezonden.

4) In de mail zal een standaardtekst worden opgenomen:

De gegevens die zijn verstrekt, zijn uitsluitend verstrekt ten behoeve van uw vraag om gegevens, hetzij eenmalig of voor een doorlopend informatieverzoek (i.e., dashboard). Zonder toestemming van de verwerkingsverantwoordelijke is het niet toegestaan om de gegevens voor een ander doel te gebruiken dan waarvoor deze zijn aangevraagd en/of ter beschikking te stellen aan derden. U bent, als ontvanger van de verstrekte gegevens, vanaf het moment van ontvangst verantwoordelijk voor de bescherming van de aan u verstrekte gegevens.