Gegevensbeschermings-effectbeoordeling (GEB)

Doelgericht ingrijpen op een aangevraagde toeslag (M1354): Het Risicoclassificatiemodel

Ministerie van Financiën, directie Toeslagen

Utrecht, 11 maart 2021

Vaststelling verwerkersverantwoordelijke: Selecteer/typ datum
Naam: Typ naam/functie

Advies functionaris voor gegevensbescherming: Selecteer/typ datum
Naam: Typ naam/functie

Advies CIO: Selecteer/typ datum
Naam: Typ naam/functie

Gegevensbeschermings-effectbeoordeling (GEB)

Ministerie van Financiën, directie Toeslagen
Directie Toeslagen

Contact:

Ministerie van Financiën, directie Toeslagen

Contact Persoon:

Versie: 2.0

1075458 00042

Inhoudsopgave

Inleiding 5
A. Beschrijving kenmerken gegevensverwerkingen 6
1. Voorstel 6
2. Persoonsgegevens 8
3. Gegevensverwerkingen 11
4. Verwerkingsdoeleinden 15
5. Betrokken partijen 18
6. Belangen bij de gegevensverwerking 19
7. Verwerkingslocaties 19
8. Techniek en methode van gegevensverwerking 19
9. Juridisch en beleidsmatig kader 45
10. Bewaartermijnen 47
B. Beoordeling rechtmatigheid gegevensverwerkingen 48
11. Rechtsgrond 48
12. Bijzondere persoonsgegevens 50
13. Doelbinding 51
14. Noodzaak en evenredigheid 52
15. Rechten van de betrokkene 55
C. Beschrijving en beoordeling risico's voor de betrokkenen 57
16. Risico's 57
D. Beschrijving voorgenomen maatregelen 60
17. Maatregelen 60
Bijlage 1: Bronnen
Bijlage 2: Halfproducten
Bijlage 3: HT indicatoren
Bijlage 4: KOT indicatoren
Bijlage 5: Control Matrix
Bijlage 6: Instemming verzoek monitoring OR
Bijlage 7: AU processen

1075458 00042

Inleiding

Per 25 mei 2018 is de General data protection regulation (GDPR) in werking getreden. Beter bekend als de Algemene verordening gegevensbescherming (AVG). In deze wet is vastgelegd dat het verplicht is om een Gegevensbeschermingseffectbeoordeling (hierna: GEB) uit te voeren.

Het doel van een GEB is om privacyrisico's in kaart te brengen van nieuwe projecten en initiatieven of bestaande verwerkingen van persoonsgegevens. Het draagt bij aan het vermijden of verminderen van privacyrisico's. Door middel van een GEB wordt op systematische wijze inzichtelijk gemaakt hoe groot de kans is dat de privacy van betrokkenen geschaad wordt en welke maatregelen er nodig zijn. Na het uitvoeren van de GEB heeft de organisatie antwoorden op deze vragen en kan er op deze basis een proces beschreven worden over de te nemen maatregelen als er zich calamiteiten of incidenten voordoen.

Deze GEB is uitgevoerd bij directie Toeslagen. Voor deze GEB wordt het process "doelgericht ingrijpen op een aangevraagde toeslag" (binnen Toeslagen benoemd het risicoclassificatiemodel) beschreven zoals dat bij Toeslagen gebruikt wordt.
Het Risicoclassificatiemodel (RCM) wordt ingezet om een inschatting te maken van het risicogehalte van toeslagaanvragen, oftewel de kans dat een aanvraag na controle aangepast of afgekeurd moet worden. Deze risicoclassificatie helpt Toeslagen bij het maken van behandelkeuzes zoals bijvoorbeeld meteen uitbetalen of eerst bewijsstukken opvragen. Deze verwerking valt onder de hoofdverwerking Toezicht.

Op basis van art. 35 AVG is deze GEB vereist omdat de volgende criteria van toepassing zijn:
a) een systematische en uitgebreide beoordeling van persoonlijke aspecten van natuurlijke personen, die is gebaseerd op geautomatiseerde verwerking, waaronder profilering, en waarop besluiten worden gebaseerd waaraan voor de natuurlijke persoon rechtsgevolgen zijn verbonden of die de natuurlijke persoon op vergelijkbare wijze wezenlijk treffen;

De AP[1] geeft aan dat het uitvoeren van een GEB verplicht is voor bepaalde verwerkingen. Op het proces "doelgericht ingrijpen op een aangevraagde toeslag" zijn de volgende criteria van toepassing:
- Grootschalige verwerkingen en/of stelselmatige monitoring van (bijzondere) persoonsgegevens in het kader van fraudebestrijding.
- Grootschalige verwerkingen en/of stelselmatige monitoring van persoonsgegevens die leiden tot of gebruik maken van inschattingen van de kredietwaardigheid van natuurlijke personen.
- Profilering

De onderstaande criteria, die afkomstig zijn van de Europese Privacy-toezichthouder, de European Data Protection Board (EDPB), zijn van toepassing:
- Evaluatie of scoretoekenning.
- Matching of samenvoeging van datasets
- Innovatief gebruik of innovatieve toepassing van nieuwe technologische of organisatorische oplossingen
- Op grote schaal verwerkte gegevens

De volgende artikelen van het Handvest van de Grondrechten van de EU worden ook op één of andere manier geraakt door de verwerkingen die vallen onder het proces "doelgericht ingrijpen op een aangevraagde toeslag":

[1] Autoriteit Persoonsgegevens

1075458 00042

Artikel 8: Bescherming van persoonsgegevens
Artikel 20: Gelijkheid voor de wet
Artikel 41: Recht op behoorlijk bestuur
Artikel 48: Vermoeden van onschuld en rechten van de verdediging

Deze GEB wordt minimaal eens in de 3 jaar, of bij grote wijzigingen herijkt. Het gaat dan om wijzigingen aan het proces of aan systemen. Het proces van herijken werkt als volgt. In een multidisciplinaire workshop wordt getoetst of de GEB nog de werkelijke verwerking c.q. situatie beschrijft. Daarnaast wordt gekeken naar de verwerkte gegevens en of de gegevens nog proportioneel en subsidiair zijn. Verder wordt de risicoanalyse voor betrokkenen opnieuw gemaakt en toetsen we de maatregelen in de actualiteit.

Na dit proces wordt er een addendum toegevoegd aan de GEB en voorgelegd aan het Management Team DGT ter akkoord en opvolging van de maatregelen. De GEB krijgt dan een nieuwe datum en versienummer. Afhankelijk van de inhoud wordt er advies gevraagd aan de Functionaris Gegevensbescherming en Privacy Officer. Vervolgens wordt er binnen de implementatie getoetst in hoeverre de maatregelen zijn doorgevoerd. Deze evaluatie volgt na 1 jaar nadat de GEB is vastgesteld. De evaluatie wordt gedaan door medewerkers van BDO/Bedrijfsvoering, het team Integrale beveiliging en Privacy (IBP Toeslagen).

A. Beschrijving kenmerken gegevensverwerkingen

Beschrijf op gestructureerde wijze de voorgenomen gegevensverwerkingen, de verwerkingsdoeleinden en de belangen bij de gegevensverwerkingen.

Onder A wordt de eerste stap beschreven van de PIA: een overzicht van de relevante feiten van de voorgenomen gegevensverwerkingen. Als de feiten onduidelijk zijn, werkt dit door in de beoordeling.

1. Voorstel

Beschrijf het voorstel waar de gegevensbeschermingseffectbeoordeling op ziet en context waarbinnen deze plaatsvindt op hoofdlijnen.

De missie van de directie Toeslagen is het juist, tijdig en rechtmatig uitbetalen van toeslagen. Om dit zo efficiënt mogelijk te doen is handhaving onmisbaar. Wij willen daarbij maatwerk leveren door de juiste combinatie van dienstverlening, voorkomen van fouten, toezicht en fraudebestrijding. Toekenningszekerheid in een zo vroeg mogelijk stadium is een belangrijk onderdeel van deze strategie.
Maandelijks verwerkt directie Toeslagen[2] honderdduizenden mutaties in toeslagaanvragen. Periodiek (ongeveer eens per maand) worden alle mutaties in een keer beschikt: de wijziging in de aanvraag wordt effectief gemaakt door een formele beschikking aan te maken. Na een formele beschikking wordt de toeslag uitgekeerd.

[2] Belastingdienst/Toeslagen

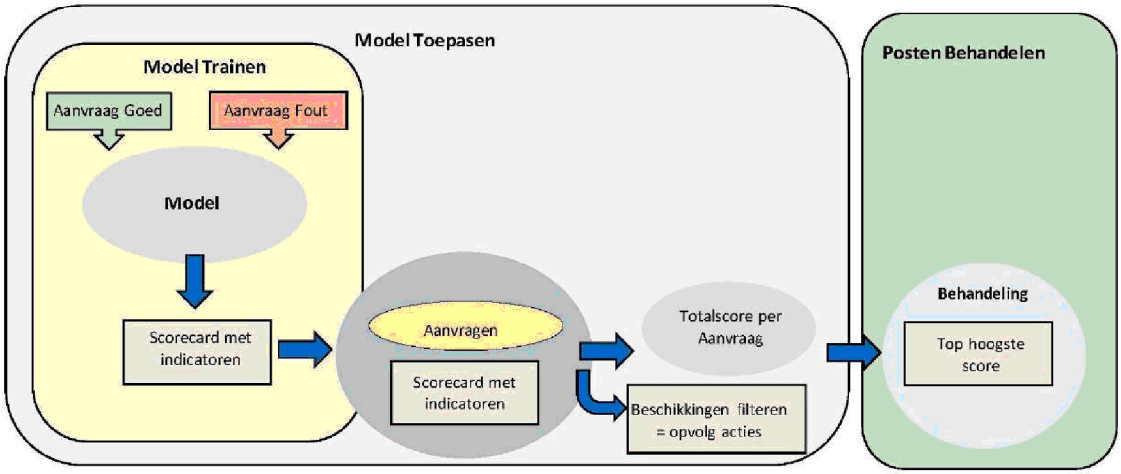
1075458 00042

Om grip te houden op mogelijke foutieve aanvragen en mutaties, wordt vlak voor het periodieke beschikkingsproces voor alle aanvragen (waarvoor een conceptbeschikking is afgegeven) een risicoscore berekend door middel van een lerend risicoclassificatie model. Deze score drukt de kans uit dat een aanvraag volgens het model onjuist is. De aanvragen met de hoogste risicoscore worden aangeboden aan behandelaars van de Toezicht team voor handmatige controle op juistheid. Het proces waarmee deze risicoscore berekend en opgevolgd wordt heet risicoclassificatie.

Het Risicoclassificatie model wordt toegepast voor Huurtoeslag en Kinderopvangtoeslag, afhankelijk van de indicatoren die worden gebruikt in het model.
Het Risicoclassificatiemodel wordt getrained volgens de supervised learning[3] methode. Dat betekent dat het model leert van bekende voorbeelden van rechtmatig/onrechtmatig gebruik en de geleerde kennis toepast op onbekende cases.

Het Risicoclassificatie proces bestaat uit 3 onderdelen:

A. Het trainen van het model;
B. Het toepassen van het model en het filteren van beschikkingen (opvolgacties);
C. Het behandelen van de geselecteerde posten.



Onderdelen A en B worden verder in deze GEB geanalyseerd. Onderdeel C maakt al deel uit van GEB M8776 Toezicht en zal hier dus niet in detail behandeld worden.

Het proces raakt ook andere processen die betrokken zijn in andere GEB's, zoals:
M8776 Toezicht
M861 Behandeld Fraudesignalen
M2077 Handmatige behandeling van onvolledige en/of tegenstrijdige toeslagenaanvragen (Uitval/uitworp)
M860 Handmatig behandelen aanvragen Toeslagen (Uitval/Uitworp)
M675 Definitief Toekennen Toeslag

Hoe het risicoclassificatie proces begint
HHR geeft opdracht aan DAT om een bepaald aantal toeslagaanvragen te selecteren met de hoogste kans op fouten. Het aantal posten wordt bepaald aan de hand van de beschikbare

[3] https://en.wikipedia.org/wiki/Supervised_learning

1075458 00042

behandelcapaciteit en er wordt bepaald waar aanvragen aan moeten voldoen om in aanmerking te komen voor selectie door het model (selectie van de indicatoren). De indicatoren zijn variabelen zoals aantal kinderen, leeftijd, huurprijs, etc.

Hoe het model wordt getraind
Het risicoclassificatiemodel is een zelflerend model dat wordt getraind door het toevoegen van goede (rechtmatig) en foute (onrechtmatig) aanvragen. De selectie van goede en foute cases wordt gebaseerd op eerdere behandelingen of ervaringen. De trainingcases bevatten alleen BSN nummers, maar om het model te kunnen trainen zijn meer gegevens nodig. Die extra gegevens zijn de indicatoren die uit verschillende bronnen komen (halfproducten en beschikkingen). De indicatoren worden geselecteerd tijdens brainstormsessies tussen HHR en de behandelaars van het Toezicht team. Die selectie is gebaseerd op huidige en eerdere ervaringen uit het toezichtpraktijk. Het model wordt gevoed door het bestand met de datapreparatie in SAS te verwerken; op die manier wordt het model getraind. Tijdens dit proces wordt er bepaald welke indicatoren voorspellen of een aanvraag goed of fout gekeurd wordt en hoe zwaar deze indicatoren wegen. Het resultaat is een scorecard met te beoordelen indicatoren en hun gewicht. Met de beschikbare scorecard worden later alle aanvragen getoetst door het model. Randvoorwaarden voor een betrouwbare scorecard zijn een frequente toevoer van nieuwe testcases zodat het model blijft leren van data die representatief is.

Hoe het model wordt toegepast
Op de scorecard krijgt elke indicator een bepaalde score toegewezen. Om het model te draaien zijn gegevens van alle aanvragen nodig. De input voor het model zijn de gegevens die gerelateerd zijn aan de indicatoren van de scorecard. Het model wordt daarna in SAS gedraaid.
Het Risicoclassificatiemodel toetst alle aanvragen in een maand aan de hand van de scorecard. Elke indicator op de scorecard heeft een waarde[4] en deze waarde resulteert voor iedere aanvraag in een uiteindelijke totaalscore die wordt gebruikt als risicoscore.
De aanvragen met de hoogste risicoscore (ongeveer 30 per maand, afhankelijk van beschikbare capaciteit) worden aangeboden aan behandelaars van het Toezicht team voor handmatige controle op juistheid. De rest van de aanvragen krijgen een formele beschikking.

Behalve de risicoscore zijn er nog een aantal kenmerken van de aanvraag die bepalen wat er met de conceptbeschikking gebeurt. Deze kenmerken worden per toeslagaanvraag in het model vastgesteld. Vervolgens wordt aan de hand van deze kenmerken per aanvraag bepaald wat de relevante opvolgactie is (bijvoorbeeld uitworp of VT/DT proces).

Het streven van directie Toeslagen is om 95% van de toeslagenaanvragen en wijzigingen volledig geautomatiseerd af te handelen. Als dat niet lukt is handmatige behandeling nodig.
Daarom wordt vóór de start van het massale automatische continuering (MAC) proces het risicoclassificatiemodel toegepast. Bij aanvragen waaraan een hoog risico wordt toegekend is handmatige behandeling nodig. Deze aanvragen worden voorlopig van het MAC-proces uitgesloten.

2. Persoonsgegevens

Som alle categorieën van persoonsgegevens op die worden verwerkt. Geef per categorie van persoonsgegevens tevens aan op wie die betrekking hebben. Deel deze

[4] Bijvoorbeeld: Het model beslist dat de indicator 'aantal kinderen' in combinatie met de indicator 'afstand KOO' een hogere score opleveren; dit is gebaseerd op de informatie van onrechtmatige aanvragen die onderdeel is van de trainingscases. Het kan bijvoorbeeld blijken dat aanvragen met minder dan 3 kinderen en meer dan 25 km afstand hoger scoren dan andere aanvragen.

persoonsgegevens in onder de typen: gewoon, bijzonder, strafrechtelijk en wettelijk identificerend.

De gegevens die worden verwerkt binnen het risicoclassificatiemodel zijn afkomstig uit verschillende externe bronnen. Met die gegevens worden halfproducten gemaakt die worden gekoppeld aan beschikkingen. Die datapreparatie wordt gebruikt om het model te trainen en toe te passen. Het aantal persoonsgegevens dat wordt gebruikt binnen de datapreparatie is afhankelijk van de indicatoren die geselecteerd worden.

Gewone persoonsgegevens

ID	Veldnaam	Bron1[5]	Bron 2[6]	Betrekking	Relevant voor onderdeel
1	NAW gegevens: • Postcode en Huisnummer • Woonland • Postadres, Adres woning	BRP[7], CIV[8]		Av[9], Tp[10]	A, B, C
2	Persoonlijke gegevens: • Geboortedatum • Leeftijd • Aantal kinderen • Overlijdensdatum	BRP, FLG[11](CIV)		Av, (Tp, Mb[12] en kind voor geboortedatum)	A, B, C
3	Inkomen gegevens: • Bijstand (WWB-uitkering)* • Inkomen • Dagtaak • Draagkracht	FLG (CIV)	Av, Tp, Mb	Av, Tp	A, B,C
4	Huurtoeslag gegevens: • Huurprijs woning • soort woning • Dienst en Recreatieruimten kosten • Energiekosten, • Rekenhuur • Schoonmaakkosten • Standplaatshuur • Woonobject identificatie • Maximale huurgrens	VHA[13]	Av	Av	A, B, C
5	Recht op kinderbijslag op basis van de AKW[14] **	SVB[15]		AV, Tp	A, B,C
6	Gegevens KOO[16]: • Naam • Adres en plaats opvanglocatie en gastouderbureau • Aantal uren opvang • Ingangsdatum kinderopvangtoeslag	SZW (LRK)[17], CIV, DUO	Av	Av, KOO	A, B, C

[5] Externe bronnen behalve TVS, trainingcases en intensief toezicht die interne bronnen zijn.
[6] Gegevens worden door de aanvrager zelf verstrekt.
[7] Basisregistratie Persoonsgegevens
[8] Centraal Informatie Voorzieningen
[9] Aanvrager
[10] Toeslag Partner
[11] Fiscale Loon Gegevens
[12] Medebewoner
[13] Verhuurders Administratie
[14] Algemene Kinderbijslagwet
[15] Sociale Verzekeringsbank
[16] Kinderopvang Organisatie
[17] Ministerie van Sociale Zaken en Werkgelegenheid

1075458 00042

	• periode waarop het kind kinderopvang heeft • uurtarief • soort opvang (dag-/of buitenschools, kindercentrum of gastouder) • LRK-id van opvanglocatie				
7	Aanvraag gegevens: • Type regeling • Ingangsdatum aanvraag • Wijzigingen van de aanvraag op datum en tegemoetkomingsjaar • Product- en procesresultaten van de behandeling van de aanvraag zoals de hoogte van de toeslag (beschikking).	CIV	Av	Av	A, B, C
8	Huishouden gegevens	Directie Toeslagen (TVS)		Av	A, B,C
9	Persoonsrelaties: • Ouder kind relatie	CIV	Av	Av, Kind	A, B, C
10	Nederlanderschap*	BRP		Av	A, B
11	Status voorderingen en beschikkingen: • Toeslagjaar • Beschikkings-nummer • Datum-dagtekening • Jaarbedrag	CIV			
12	Wanbetaler zorg status*	CVZ[18] (CIV)		Av	A, B
13	WOZ onroerend goed bezitting	CIV		Av	A, B, C
14	Medewerkers ID (interne bron)	ICT Infra logging		Medewerker	A, B, C

* Deze gegevens (WWB-uitkering, Nederlanderschap, Wanbetaler zorg status) worden niet meer als indicator gebruikt maar ze maken nog steeds deel uit de datapreparatie. Ze zullen in Q2 2021 structureel uit de datapreparatie verwijderd worden (zie Risico's en maatregelen).

**Het gegeven "Recht op kinderbijslag op basis van de AKW" is gerelateerd aan de KGB[19] regeling en niet aan KOT of HT. Het gebruik kan onrechtmatig zijn, daarom is het als een risico geregistreerd (zie 16.Risico's, risico 2.)

Wettelijke identificerend persoonsgegevens

Id	Veldnaam	Bron 1	Bron 2	Betrekking op	Relevant voor onderdeel
15	BSN[20]	BVR, DUO[21], COA[22], CIV, TVS, trainingcases, toezichtlijst[23]	Av	Av, Tp, Mb, Kind	A, B, C

[18] College voor Zorgverzekeringen
[19] Kinderbondbudget
[20] Burgerservicenummer
[21] Dienst Uitvoering Onderwijs
[22] Centrale Ontvangers Administratie
[23] Lijst met BSN's die bevat de bewezen fraudegevallen behandeld door het IntensiefToezichtteam

1075458 00042

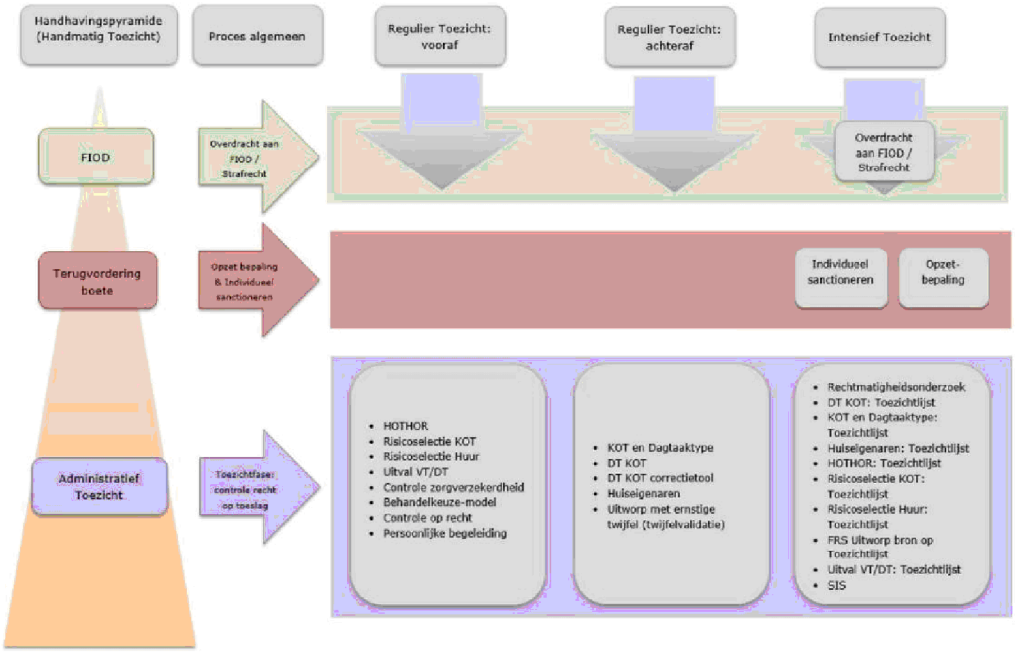
3. Gegevensverwerkingen

Geef alle voorgenomen gegevensverwerkingen weer.

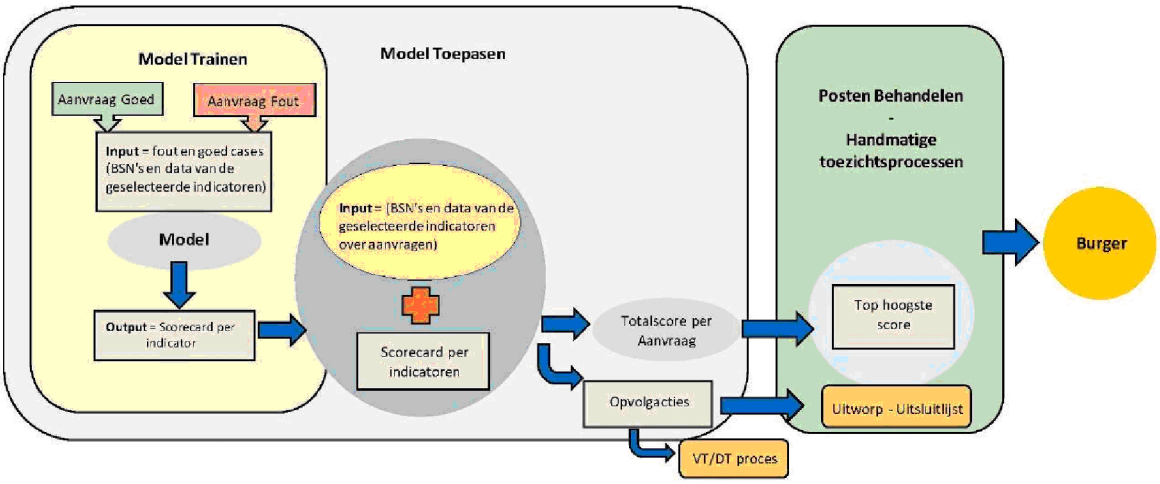
Het process "doelgericht ingrijpen op een aangevraagde toeslag" valt onder de Toeslagen bedrijfsprocessen 'Toezicht – handhavingsregie'. De Belastingdienst beoogt met 'handhaving' het gedrag van burgers en bedrijven zodanig te beïnvloeden dat zij structureel (fiscale) regels naleven (compliance). Dit moet zorgen voor de borging van de continuïteit van belastingopbrengsten en de rechtmatige betaling van toeslagen. Directie Toeslagen handelt bij het handhaven binnen de juridische kaders van de wet en conform de beginselen van behoorlijk bestuur.

Onder handhaving wordt door directie Toeslagen alle activiteiten verstaan die bijdragen aan structurele regelnaleving (compliance) door toeslaggerechtigden. Dit houdt voor directie Toeslagen specifiek in dat de burger
(a) juiste en volledige informatie verstrekt bij de aanvraag van een toeslag en bij het doorgeven van wijzigingen die relevant zijn voor de aanspraak op een toeslag,
(b) tijdig en adequaat reageert op verzoeken van de directie Toeslagen om informatie en
(c) eventuele terugvorderingen tijdig betaalt.
In aanvulling op deze algemene compliance doelstelling streeft directie Toeslagen met haar handhavingsbeleid het waarborgen na van de rechtmatigheid van de toekenning van toeslagen, ofwel de mate waarin wordt gehandeld in overeenstemming met de geldende wet- en regelgeving.

Onderstaand is ingezoomd op de handmatige toezichtsprocessen:

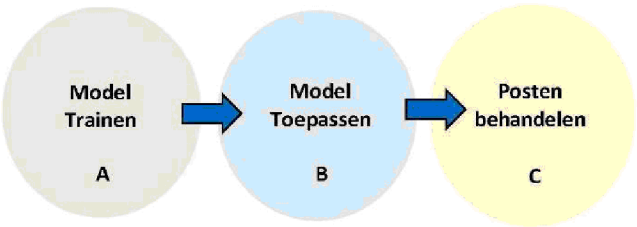


De output van het risicoselectiemodel (risicoclassificatiemodel) wordt gebruikt binnen het handmatige toezichtsproces van administratief Toezicht. Hieronder wordt ingezoomd op de verwerking zelf:



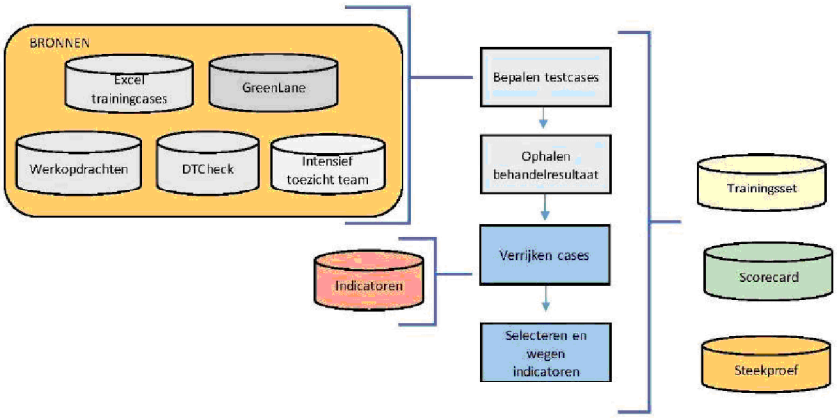
Het risicoclassificatiemodel proces bevat 3 onderdelen:

A. Het trainen van het model
B. Het toepassen van het model en het bepalen van opvolgacties
C. Het behandelen van de geselecteerde posten



De GEB is ingedeeld volgens deze driedeling, waarbij dus alleen onderdeel A en B beschreven worden in deze GEB.

A. Trainen van het model



Het RCM is 'lerend': aan de hand van aanvragen uit het verleden waarvan bekend is of deze fout (onrechtmatig) of goed (rechtmatig) waren. Dit is bekend als zijnde de 'trainingscases'. Hierbij wordt uitgerekend welke combinaties van indicatoren risicovol zijn.
De trainingscases vormen het ruggengraat van de risicoclassificatiemodel: zonder een gebalanceerde populatie van kwalitatief goede trainingscases is het onmogelijk een goed model te trainen, en hierop in beleid en uitvoering op te steunen.
Doel van de trainingscases proces is om de toevoer van goede trainingscases te continueren. Hierdoor wordt de trainingset geleidelijk ververst, en groeit het model automatisch mee met de patronen van niet compliant cases die recent geconstateerd zijn.
Het model wordt gevoed met trainingscases waaraan historische contra-informatie toegevoegd wordt. Deze reconstructie van trainingscases wordt 'tijdreizen' genoemd.
Alleen wanneer nieuwe trainingscases geïdentificeerd worden zal het tijdreizen aangepast en opnieuw gerund moeten worden. Wanneer geen nieuwe trainingscases beschikbaar zijn zal de trainingset, en daarmee ook het model, gelijk blijven.

Er worden vijf interne bronnen voor de trainingscases gebruikt:

1. Trainingscases in Excel
2. GreenLane
3. Werkopdrachten (TVS)
4. Intensief toezicht teams
5. DTCheck

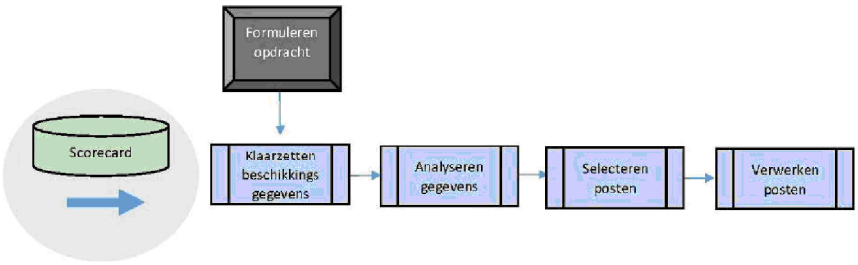
Er wordt per bron bepaald of het behandelresultaat goed of fout is geweest. De bronnen bevatten BSN nummers van aanvragen uit het verleden waarvan bekend is dat deze fout (onrechtmatig) of goed (rechtmatig) zijn geweest. Voor elke bron zijn businessrules opgesteld die bepalen wanneer een case 'fout' of 'goed' genoemd wordt. (zie hoofdstuk 8.Techniek en methode van gegevensverwerking - 2. Bepalen behandelresultaat (goed / fout))

Aan de trainingscases voegt DAT gegevens over de aanvraag en aanvrager toe die pas in de vervolgstap worden geanalyseerd op voorspellende kracht. Dit zijn de indicatoren die gebruikt worden om het scoremodel te bepalen.
De indicatoren zijn variabelen (bijvoorbeeld: aantal kinderen of aantal verhuizingen) die een indicatie kunnen zijn of onrechtmatigheid. De indicatoren zijn geselecteerd door HHR en PR op basis van eerdere onrechtmatige cases. De informatie die gerelateerd is aan de indicatoren komt uit verschillende bronnen (halfproducten en beschikkingen, zie hoofdstuk 8.Techniek en methode van gegevensverwerking - 3. Klaarzetten beschikking gegevens) en wordt aan het BSN gekoppeld. Dit bestand met BSN's en gekoppelde informatie van de geselecteerde indicatoren wordt gebruikt om het model te kunnen trainen. De verschillende versies van dit bestand worden in SAS opgeslagen. Alleen geautoriseerde DAT medewerkers hebben toegang tot het bestand.

Maar er zijn een aantal oude SAS bestanden die ook op de Q-schijf worden bewaard. Dit stamt nog uit de tijd dat we niet de beschikking hadden over de SAS Grid en alles nog op een aparte pc op moesten slaan. Die bestanden moeten nog bewaard worden om het model te kunnen trainen, maar het is niet meer nodig om die bestanden op de Q-schijf te bewaren - het kan ook alleen op de SAS Grid. Dit zal in 2021 worden aangepast (zie Risico 9 in Hoofdstuk 16 en 17 Risico's en Maatregelen)

Door middel van een logistisch regressiemodel wordt er tijdens het trainen van het model een scorecard gemaakt. Deze scorecard bevat de indicatoren met hogere scores die daarna worden gebruikt om de risicoscore van de aanvragen te bepalen.

B. Toepassen van het model



De opdrachtgever HHR[24] wil data inzetten om risicogebaseerd keuzes te maken over de aan te bieden posten voor toezichtbehandeling in de fase van voorlopige toekenning: de posten waar het risico het grootst is dat de aanvraag/wijziging niet juist en/of rechtmatig is. De opdrachtnemer D.A.T. maakt hiervoor gebruik van het risicoclassificatiemodel waarin door middel van meerdere indicatoren het risico wordt bepaald of een aanvraag (deels) onrechtmatig is.

[24] Handhavingsregie

1075458 00042

Vergelijkbaar met het trainen van het model selecteert HHR een aantal indicatoren die worden gekoppeld aan de BSN's van alle lopende en nieuwe aanvragen. De verschillende versies van dit bestand worden in SAS opgeslagen. Zoals net bij A benoemd zijn er een aantal oude SAS bestanden op de Q-schijf opgeslagen en dit zal in Q2 2021 aangepast worden. Alleen geautoriseerde DAT medewerkers hebben toegang tot het bestand.
De toepassing van het risicoclassificatiemodel resulteert in een rangschikking van de aanvragen op basis van de risicoscore.
Samen met de selectie op basis van harde criteria (zie hoofdstuk 8.Techniek en methode van gegevensverwerking - 5.1 Filteren beschikkingen – opvolgacties) wordt de top van deze rangschikking voor handmatige behandeling aangeboden aan toezichtmedewerkers (team Toezicht). In deze behandeling wordt de aanvraag gecontroleerd op het rechtmatig en juist toekennen van de toeslag voordat er een voorlopige toekenning wordt afgegeven.

De teams van Toeslagen die betrokken zijn bij het Risicoclassificatiemodel zijn: HHR in de rol van opdrachtgever, D.A.T. in de rol van opdrachtnemer en Toezicht in de rol van behandelaar van de te controleren posten (uitvoering).

C. Het behandelen van de geselecteerde posten.

Het Toezicht team zal als behandelaar de geselecteerde posten controleren. De verwerking van de bovengenoemde gegevens heeft als verwerkingsdoel het vaststellen van de rechtmatigheid van de toekenning. Door het controleren van de lopende toeslag met de ontvangen bewijsstukken van de belanghebbende of contra-informatie van onder andere kinderopvangorganisaties en verhuurder kan worden vastgesteld of de toeslag een aanpassing behoeft. Hiermee wordt zorg gedragen voor een juiste vaststelling van de toeslagen.
De behandelaar kan na het controleren van informatie de aanvraag handmatig corrigeren als de rechtmatigheid geconstateerd is. Vanaf dat moment kan de aanvraag verder verwerkt worden. Onrechtmatige aanvragen zullen een opvolgend proces volgen. Dit proces wordt in meer detail geanalyseerd in andere GEB's zoals M861 Behandelen Fraudesignalen en M8776 Toezicht.

Dit proces C valt binnen de scope van GEB M8776 Toezicht. Hiermee valt het buiten scope van deze GEB en wordt het hier niet verder in detail beschreven.

4. Verwerkingsdoeleinden

Beschrijf de doeleinden van de voorgenomen gegevensverwerkingen.

Id	Veldnaam	Verwerkingsdoel
1	Postcode en Huisnummer, Woonland, Postadres, Adres woning	Verwerking van persoonsgegevens in het kader van de uitvoering van hun taak en om uitwisseling te kunnen maken met (authentieke) bronnen. Het adres is nodig om de geografische locatie te bepalen en de afstand tot het opvangadres te berekenen (indicator afst_adr_opvang).
2	Geboortedatum, Aantal kinderen, Overlijdensdatum	Geboortedatum wordt gebruikt om met voldoende zekerheid te kunnen stellen of het

		echt om de gegevens van persoon X gaat (vergewisplicht). Voor kinderopvangtoeslag zijn de leeftijden 4 en 12 jaar relevant als indicatie om te bepalen of er recht is op kinderopvangtoeslag op basis van dagopvang of op basis van naschoolse opvang[25]. Dit is afhankelijk van of een kind naar school gaat. Voor de huurtoeslag dient de aanvrager meerderjarig te zijn.[26] Voor de huurtoeslag is de leeftijd van belang bij het vaststellen van de hoogte van het recht op huurtoeslag. Ouderenhuishouden, medebewoner en leeftijd van het kind i.v.m. vrijstellingsdeel van het eventuele inkomen.
3	Toeslagpartner heeft werkt, draagkracht, inkomen Toeslagpartner heeft bijstand (WWB-uitkering)*	Om te kunnen bepalen bij welke bron de contra-informatie over gewerkte uren kan worden geverifieerd. Ter bepaling van de draagkracht[27] wordt het toetsingsinkomen[28], van de toeslagaanvrager, toeslagpartner, kind (>18jaar) en evt. medebewoner in aanmerking genomen. WWB-uitkering (hoogte) werd als risico indicator in het model opgenomen. WWB-uitkering was een fraudesignaal in het verleden.
4	Huurprijs woning, soort woning, Huurprijs woning, soort woning Dienst en Recreatieruimten kosten, Energiekosten, Rekenhuur, Schoonmaakkosten, Standplaatshuur, Woonobject identificatie, Maximale huurgrens	Om te kunnen bepalen of er recht is op huurtoeslag, en de hoogte van de huurtoeslag te kunnen bepalen.
5	Statusgegeven: Tp of Av heeft recht op kinderbijslag op basis van de AKW voor een kind	Deze gegevens zijn nodig omdat het kindgebonden budget afhankelijk is van het recht op kinderbijslag.[29] Het werd als risico indicator in het model opgenomen. Dit gebruik kan onrechtmatig zijn omdat deze gegevens niet aan KOT en HT gerelateerd zijn. Als onderdeel van de maatregelen zal ook de rechtmatigheid van het gebruik onderzocht worden.
6	Gegevens KOO: Naam, Adres en plaats opvanglocatie en gastouderbureau, Aantal uren opvang, Ingangsdatum kinderopvangtoeslag, Periode waarop het kind kinderopvang heeft genoten (begin- en einddatum in deze	Deze gegevens zijn noodzakelijk om vast te stellen of er sprake is van recht op kinderopvangtoeslag en dit te communiceren met de burger en de KOO in het aanvraag- en mutatieproces. De hoogte van de kinderopvangtoeslag is afhankelijk van het aantal uren kinderopvang per kind in het berekeningsjaar[30].

[25] In artikel 1.1. lid 1 Wko en artikel 1 sub b Besluit kinderopvangtoeslag.
[26] Artikel 1, lid 1, onder c van de wet op de zorgtoeslag en artikel 8 wet op de huurtoeslag.
[27] In artikel 7 lid 1 Awir.
[28] In artikel 8 lid 1 Awir.
[29] Artikel 2, lid 1 wet op het kindgebonden budget.
[30] In artikel 1.7 lid 1 sub b ten eerste Wko.

	kalendermaand), uurtarief, Soort opvang (dag-/of buitenschools, kindercentrum of gastouder), LRK-id van opvanglocatie	De hoogte van de kinderopvangtoeslag is afhankelijk van de voor die kinderopvang te betalen prijs[31]
7	Aanvraag gegevens: Type regeling, Ingangsdatum aanvraag, Wijzigingen van de aanvraag op datum en tegemoetkomingsjaar, Product- en procesresultaten van de behandeling van de aanvraag, zoals de hoogte van de toeslag (beschikking).	Om de periode van recht op toeslag vast te kunnen stellen
8	Huishouden gegevens: aantal personen	Voor het vaststellen van het recht op en de hoogte van een tegemoetkoming zijn het aantal personen op een adres van belang.
9	Persoonsrelaties: Ouder kind relatie	Verwerking van persoonsrelaties ten behoeve van het vaststellen van toeslagpartnerschap en de samenstelling van het huishouden teneinde de toeslagen correct te kunnen vaststellen.
10	Nederlanderschap ja/nee*	Het bepalen of sprake is van wel of geen Nederlanderschap is van belang om te kunnen vaststellen of een belanghebbende recht heeft op een toeslag. De manier waarop dit werkt is dat er een script naar de bron in BvR kijkt en bepaalt of de waarde van Nederlanderschap ja of nee is. Nationaliteit als gegeven wordt niet op een andere manier verwerkt in het proces. Deze gegevens worden ook niet opgeslagen, behalve de output van het script (Nederlanderschap ja/nee of 0/1)
11	Status voorderingen en beschikkingen	Om de financiële situatie van de aanvrager te kunnen bepalen.
12	Wanbetaler zorg status*	Gegevens van gevoelige aard. Het werd als risico indicator in het model opgenomen: aanvrager heeft geen zorgpremie afgedragen (hoger risico).
13	WOZ onroerend goed bezitting	Om te kunnen bepalen of er recht is op huurtoeslag, en de hoogte van de huurtoeslag te kunnen bepalen.
14	Medewerkers ID	Gegevens van gevoelige aard. In het kader van login en monitoring van bepaalde acties (zie hoofdstuk 8)

De volgende gegevens worden sinds een aantal jaar geleden niet meer als indicator gebruikt in het model, maar de gegevens worden nog wel verwerkt als bron (ze zijn onderdeel van de datapreparatie):
3. Toeslagpartner heeft bijstand (WWB-uitkering)
10. Nederlanderschap ja/nee
12. Wanbetaler zorg status

Deze zullen in 2021 structureel uit het model verwijderd worden (zie Risico's en Maatregelen). Bij de start van de risicoselectie in 2013 is men op zoek gegaan naar mogelijke indicatoren die invloed hebben op het risico op fouten. Deze indicatoren waren daar een aantal van maar omdat ze geen effect op het model hebben (er geen invloed op de risicoscore is gezien tijdens testen) is er besloten om ze niet meer te gebruiken. Ze worden op "reject" gezet in het model. Dat betekent dat het model deze indicatoren of variabele niet meer gebruikt.

[31] In artikel 1.7 lid 1 sub b ten eerste Wko.

Id	Veldnaam	Verwerkingsdoel
15	BSN	Verwerking van persoonsgegevens[32] in het kader van de uitvoering van hun taak en om uitwisseling te kunnen maken met (authentieke) bronnen. Om het model te trainen worden trainingscases gebruikt. Die bevatten een lijst van BSN's met het resultaat fout of goed. Daarnaast wordt het model getraind met een lijst met posten die vanwege de opzet niet compliant zijn. Die lijst komt van het Intensief toezicht team en bevat alleen BSN nummers. Het BSN wordt gebruikt als unieke identifier om de cases te kunnen identificeren en te kunnen koppelen aan de informatie uit de indicatoren (meer informatie in hoofdstuk 12. Bijzondere persoonsgegevens).

5. Betrokken partijen

Benoem welke organisaties betrokken zijn bij welke gegevensverwerkingen. Deel deze organisaties per gegevensverwerking in onder de rollen: verwerkingsverantwoordelijke, verwerker, verstrekker en ontvanger. Benoem tevens welke functionarissen binnen deze organisaties toegang krijgen tot welke persoonsgegevens.

Directie Toeslagen heeft convenanten gesloten met een aantal organisaties. Dit zijn meestal overheidsorganisaties maar dat hoeft niet altijd zo te zijn. Hierbij gaat het om afspraken over:

- de gezamenlijke handhaving van overheidsregelingen
- het onderling uitwisselen van informatie

De overheid wil dat haar organisaties meer samenwerken waardoor de overheid als geheel efficiënter gaat werken. In dat verband wordt onderling ook informatie uitgewisseld. Uiteraard worden daarbij de wettelijke bepalingen strikt in acht genomen zoals bijvoorbeeld de privacyregelingen. De directie Toeslagen hecht aan transparantie en publiceert daarom de samenwerkingsconvenanten met andere organisaties.

Rol	Naam
Verwerkingsverantwoordelijke	Minister van Financiën, voor deze de directeur-generaal directie Toeslagen, voor deze de algemene directeur Toeslagen
Externe verstrekkers	10.2.g, Gemeente, UWV[34], FLG, VHA, SVB, SZW, CIV, IBS-COA
Interne verstrekkers	CAP[35], Intensief toezicht team

[32] In artikel 10 Wet algemene bepalingen Burgerservicenummer (Wabb).
[33] Verhuurder van tijdelijk woonruimte. Vanuit directie Toeslagen zijn er afspraken gemaakt met verhuurders met meer dan 10 woningen m.b.t. het standaard leveren van diverse gegevens, waaronder in ieder geval de kale huurprijs en indien van toepassing de kosten voor serviceposten, of voor een woonwagenstandplaats. Deze gegevens worden op jaarlijkse basis gedeeld met de Verhuurdersadministratie van directie Toeslagen en worden vervolgens verwerkt in TVS.
[34] Uitvoeringsinstituut Werknemersverzekeringen
[35] Centrale Administratieve Processen

Interne verwerker	HHR, DAT, PR[36], CAP-UPG[37], Toezicht
Interne ontvanger	Werkvoorbereiding Queryloket Handhavingsregie (postbus)

6. Belangen bij de gegevensverwerking

Beschrijf alle belangen die de verwerkingsverantwoordelijke en anderen hebben bij de voorgenomen gegevensverwerkingen.

Het belang van burgers is dat zij tijdig het juiste toeslagbedrag krijgen. Mocht het feit zich voordoen dat zij geen recht hebben, dan is het in het belang van de aanvrager dat deze ook geen voorschot krijgt, zodat er niet onterecht toeslag uitgekeerd wordt en zodat geen terugvordering plaats hoeft te vinden. Daarnaast is de burger erbij gebaat dat hij zo snel mogelijk het juiste toeslagbedrag krijgt en daarbij niet onterecht in lange toezichtprocessen verzeilt raakt.

Het belang van Toeslagen is het juist, tijdig en rechtmatig uitbetalen van toeslagen. Het RCM is daarbij behulpzaam om zo efficiënt en effectief mogelijk de handmatige behandelcapaciteit in te zetten en burgers daardoor ook zoveel mogelijk meteen het juiste toeslagbedrag uit te keren ipv dat achteraf het toeslagbedrag toch anders blijkt te zijn.

Daarnaast speelt het maatschappelijk belang dat er verwacht mag worden dat de overheid zorgvuldig omgaat met en gebruik maakt van gegevens van burgers. De gegevens van burgers op individueel niveau moeten ook inzichtelijk (gemaakt kunnen worden) zijn.

7. Verwerkingslocaties

Benoem in welke landen de voorgenomen gegevensverwerkingen plaatsvinden.

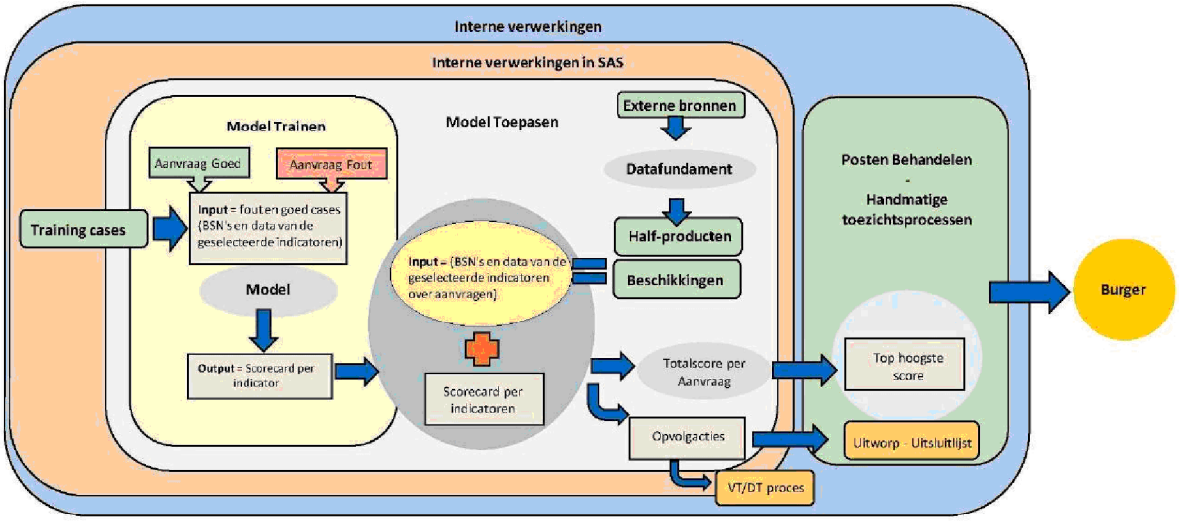
Uitsluitend in Nederland.

8. Techniek en methode van gegevensverwerking

Beschrijf op welke wijze en met gebruikmaking van welke (technische) middelen en methoden de persoonsgegevens worden verwerkt. Benoem of sprake is van (semi-)geautomatiseerde besluitvorming, profilering of big data-verwerkingen en, zo ja, beschrijf waaruit een en ander bestaat.

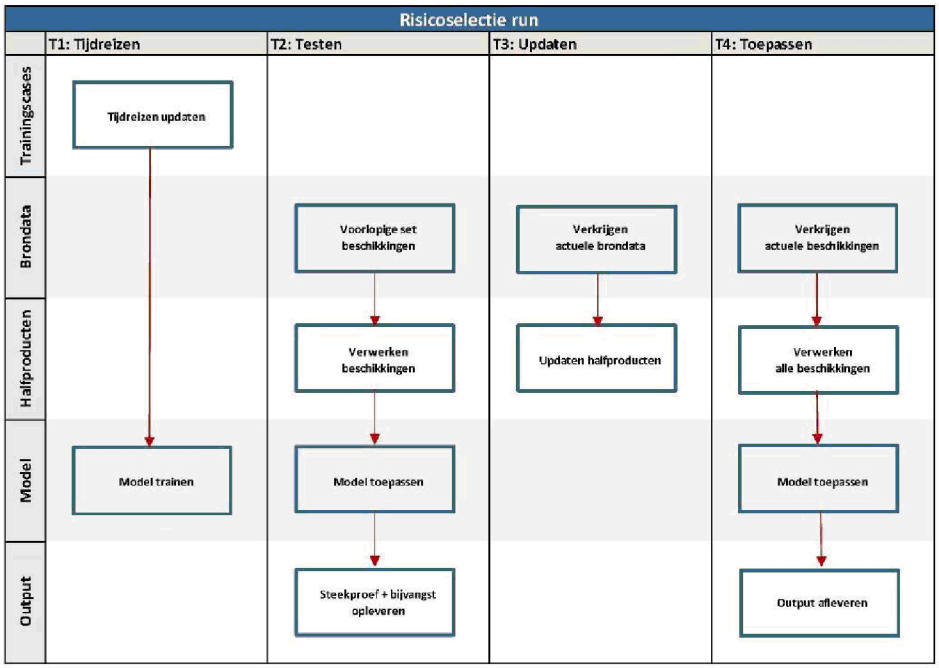
Het risicoclassificatiemodel is lerend en kan op basis van historische aanvragen waarvan de rechtmatigheid/onrechtmatigheid bekend is (trainingscases) de kans voorspellen dat een aanvraag onrechtmatig is. Het model wordt in SAS Enterprise Miner getraind. Tijdens een run van het model wordt in de meeste gevallen eerst een nieuw model berekend. Dit nieuwe model wordt vervolgens toegepast op de meest actuele set beschikkingen. In de praktijk wordt dit niet zo heel vaak gedaan; de huidige scorecard stamt nog uit 2019.

[36] Productieregie
[37] CAP-UPG (Unit Productiekwaliteit en Gebruiksbeheer)



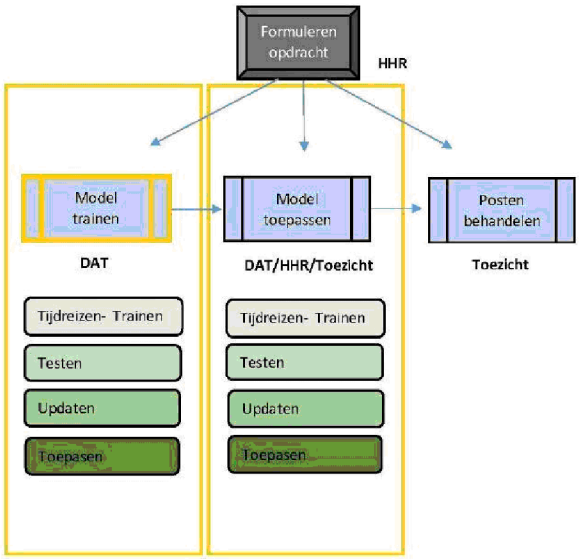
Deze toepassing beslaat vier fasen:

1. **Trainen:** de set met trainingscases wordt, indien beschikbaar, aangevuld met nieuwe cases. Het model wordt op basis van de nieuwe trainingsets getraind. Tijdreizen is de samenstelling van de trainingset op basis van historische contra-informatie.
2. **Testen:** het opnieuw getrainde model wordt getest op een voorlopige set van conceptbeschikkingen. Voor de meeste halfproducten wordt hierbij gebruik gemaakt van brondata van de vorige run. Dit leidt tot een steekproef die handmatig beoordeeld wordt door een team medewerkers die fungeren als 'steekproevers'.
3. **Updaten:** de meeste halfproducten worden ververst met actuele brondata. Als de opvolging van steekproef hier aanleiding toe geeft, wordt het model inhoudelijk aangepast. Het model wordt dan nogmaals toegepast, en uit de resultaten wordt wederom een steekproef getrokken.
4. **Toepassen:** alle actuele conceptbeschikkingen worden verkregen en wordt het model voor de laatste keer toegepast. Dit leidt tot definitieve output van het model, op basis waarvan aanvragen stopgezet, onderzocht en gerapporteerd kunnen worden.

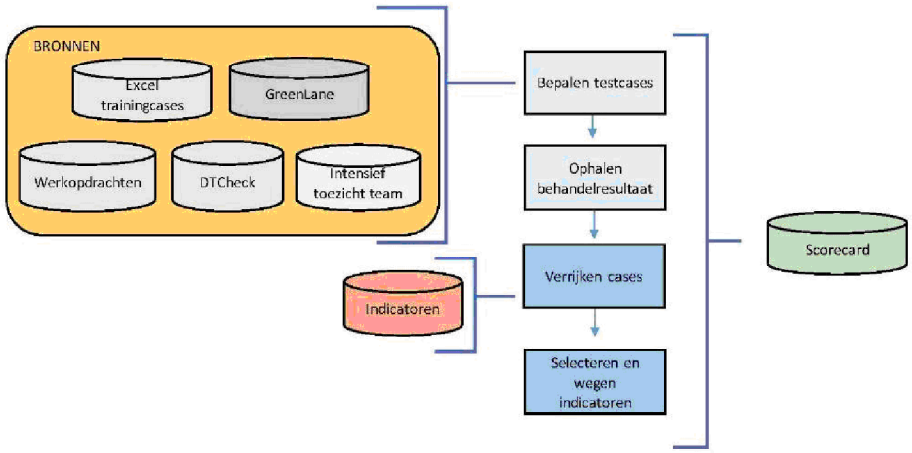


Het RCM proces bevat 3 onderdelen:

A. Het trainen van het model
B. Het toepassen van het model en filteren van beschikkingen (opvolgacties)
C. Het behandelen van de geselecteerde posten



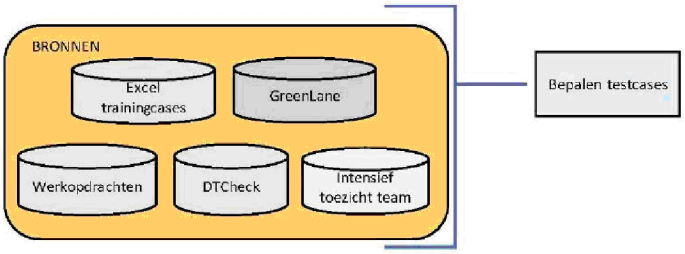
A. Trainen van het model



Proces	Omschrijving	Organisatie
1. Bepalen testcases	Geautoriseerde[38] medewerkers van DAT en HHR bepalen welk type handmatig behandelde aanvragen worden gebruikt als testcase	DAT
2. Bepalen behandelresultaat (goed / fout)	Geautoriseerde medewerkers van DAT geven het label goed of fout aan de testcases	DAT
3. Verrijken cases	Geautoriseerde medewerkers van DAT koppelen indicatoren aan de testcases	DAT, HHR, Toezicht
4. Selecteren en wegen indicatoren	Geautoriseerde medewerkers van DAT bepalen of hoeveel weging de indicatoren hebben in het voorspellen van goed of fout	DAT

1. Bepalen testcases

[38] De autorisaties voor het systeem SAS worden door de Teamleader DAT geregeld.



Voor het bepalen van de testcases zijn vijf bronnen beschikbaar die geautoriseerde medewerkers van DAT gebruiken om een trainingsset te vormen waarmee een scorecard gemaakt wordt:

- **Trainingscases (Excel bestand):** in het verleden (vanaf mei 2013) zijn handmatig trainingscases verzameld, zowel voor KOT als HT. De oorsprong hiervan is velerlei, en niet altijd vastgelegd. De kwaliteit hiervan is daarom in het algemeen (vooral voor de periode mei – december 2013) wisselend. Ondanks de wisselende kwaliteit vormt deze verzameling posten een substantieel deel van de trainingset. In december 2015 is besloten om cases geïdentificeerd in 2013 te bevroren, en geen nieuwe cases voor toeslagjaar 2013 meer toe te voegen. Daarom worden alle cases met toeslagjaar 2013 of eerder hier al uitgesloten van verdere verwerking. Met ingang van april 2019 zijn ook de jaren 2014 en 2015 bevroren. Sinds 2019 zijn geen cases meer toegevoegd.
  De cases zijn in SAS Enterprise tool verwerkt. De output is een Excel bestand die een lijst van BSNs met het resultaat fout of goed bevat. De output is in SAS bewaard. Een backup is ook in een map op de Q-schijf bewaard. Het bestand is toegankelijk voor alle D.A.T. medewerkers.
- **GreenLane:** als onderdeel van de GreenLane[39] pilots zijn in meerdere iteraties een groot aantal posten behandeld door Toezicht. Deze behandelaars hebben gegeven een groot aantal stukken contra-informatie beoordeeld of deze posten 'de GreenLane op mochten', oftewel: vanwege een laag risico zonder verdere behandeling definitief beschikt mochten worden.
  De cases die door de rechters als GreenLane waardig beoordeeld zijn, worden in deze flow als compliante case meegenomen in de trainingset.
  Het GreenLane project is gestopt en er worden geen nieuwe trainingscases meer toegevoegd vanuit dit project.
  De gegevens (in een onversleuteld Excel bestand) waren vanuit productieregie op de Q-schijf geplaats. Die waren daarna ingelezen in SAS. Het output bestand bevat een lijst van BSNs met het resultaat goed. De output is in SAS bewaard. Een backup is ook in een map op de Q-schijf bewaard. Het bestand is toegankelijk voor alle D.A.T. medewerkers.
- **Werkopdrachten:** in deze stap worden de posten toegevoegd waarvan we uit de werkopdrachten weten dat ze in toezicht hebben gezeten. Vervolgens worden daar de beschikkingen aan gekoppeld van vóór en na het toezicht. Afhankelijk van het soort toeslag en verschil tussen het bedrag voor en na de controle wordt bepaald of een post als goed of fout wordt geclassificeerd. Deze verwerking gebeurt in SAS. Gegevens uit het data fundament (TVS en beschikkingen) worden gebruikt. Het output bestand bevat een lijst van BSNs met het resultaat goed of fout. De output is in SAS bewaard. Het bestand is toegankelijk voor alle D.A.T. medewerkers (zie Risico's en Maatregelen).
- **Intensief toezicht teams:** de Intensief toezicht teams houden een centrale lijst bij van posten die vanwege opzet niet compliant zijn. Een selectie hiervan kan als fout case toegevoegd worden aan het model. Ook deze stroom is lopend, hoewel aantallen relatief beperkt zijn.

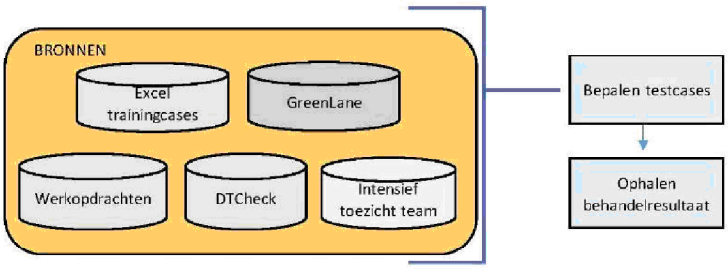
[39] GreenLane project uit 2016: Verschuiving van integraal toezicht naar onderbouwd selectief toezicht.

1075458 00042

De gegevens (in een onversleuteld Excel bestand) worden door de Intensief toezicht team op de Q-schijf geplaats. Die worden daarna ingelezen in SAS. Het output bestand bevat een lijst van BSNs die in SAS bewaard wordt. Een backup is ook in een map op de Q-schijf bewaard. Het bestand is toegankelijk voor alle D.A.T. medewerkers.

- **DTCheck:** als extra check willen we posten hebben die niet gecontroleerd zijn en die als goed getypeerd kunnen worden. Dit is nu alleen gedaan voor toeslagjaar 2017. Daarbij is gekeken naar het verschil tussen VT[40] en DT[41] en meldingen die aanvragers hebben gedaan. Uit het bestand wat zo verkregen is, is een gestratificeerde steekproef getrokken zodat de steekproef op een aantal achtergrondkenmerken overeenkomt met de populatie. Deze verwerking gebeurt in SAS. Gegevens uit het data fundament (beschikkingen) worden gebruikt. Het output bestand bevat een lijst van BSNs met het resultaat goed of fout. De output is in SAS bewaard. Het bestand is toegankelijk voor alle D.A.T. medewerkers.

2. Bepalen behandelresultaat (goed / fout)



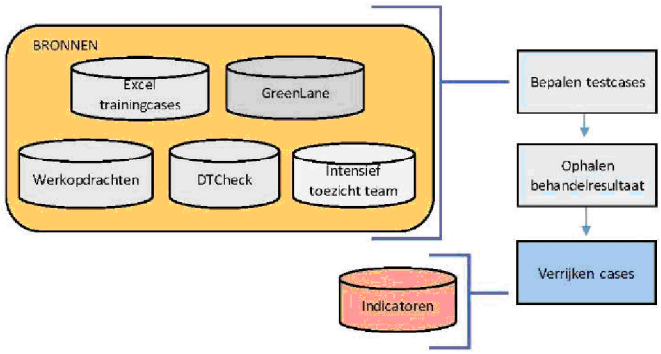
Voor elke bron zijn businessrules opgesteld wanneer een case 'fout' of 'goed' wordt genoemd. Per herkomst is de aanpak hiervan verschillend:

- **Excel:** bij opzet van het model en bij gebrek aan betere trainingscases is ervoor gekozen om een selectie van BSN's die over een langere periode geen mutaties hebben doorgevoerd allemaal als 'Goed' te labelen, ondanks dat hiervoor niet is vastgesteld of de case daadwerkelijk compliant was. Door steeds meer cases te toevoegen wordt de kans kleiner of een oneerlijk samengestelde trainingset.
- **GreenLane:** alle posten die in deze iteraties als compliant beoordeeld zijn, kunnen als 'goed' case toegevoegd worden aan de trainingset.
- **Werkopdrachten:** aan de werkopdrachten worden beschikkingen gekoppeld van vóór en na het toezicht. Afhankelijk van het soort toeslag en verschil tussen het bedrag voor en na de controle wordt bepaald of een post als goed of fout wordt geclassificeerd.
- **Intensief toezicht team:** alle posten die hieronder vallen, worden als 'fout' toegevoegd aan de trainingset.
- **DTCheck:** alle posten die hieronder vallen, worden als 'goed' toegevoegd. Omdat we over weinig goede cases beschikken hebben we voor 2017 goede cases gehaald uit DT-beschikkingen die nauwelijks afwijken van de VT.

3. Verrijken cases

[40] Voorlopige Toekenning
[41] Definitieve Toekenning

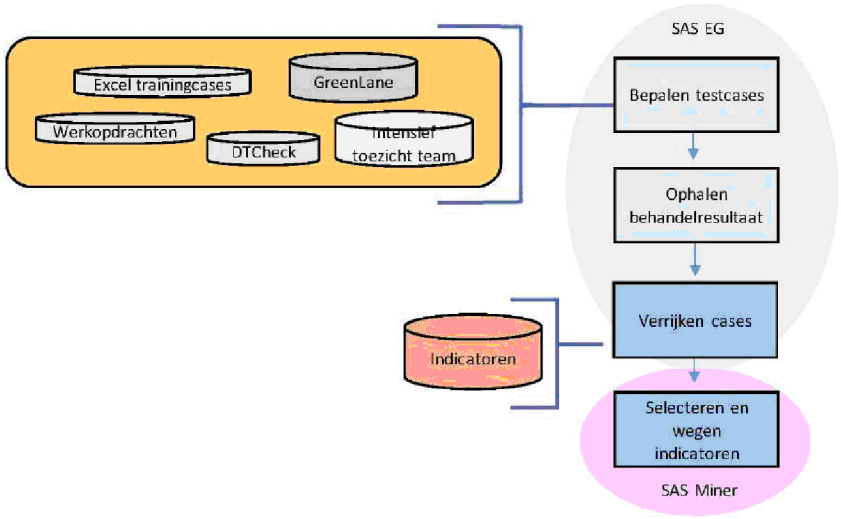
1075458 00042



Aan de cases voegt een geautoriseerde medewerker van DAT gegevens over de aanvraag en aanvrager toe die pas in de vervolgstap worden geanalyseerd op voorspellende kracht. Dit zijn de indicatoren[42] die gebruikt worden om het scoremodel te bepalen. De indicatoren worden toegevoegd vanuit verscheidene bronnen (zie bijlage 3 en 4).

Deze gegevens worden gekozen op relevantie voor de toeslagsoort en op basis van heuristieken van medewerkers van DAT, HHR en Toezicht. De resulterende trainingsset en de indicatorenlijst worden opgeslagen in SAS en op de Q-schijf.

4. Selecteren en wegen indicatoren



De verrijkte gegevens worden met statistische-analyse-software SAS Enterprise Miner geanalyseerd op voorspellende kracht waarna hun weging in de voorspelling wordt bepaald. Door interactive

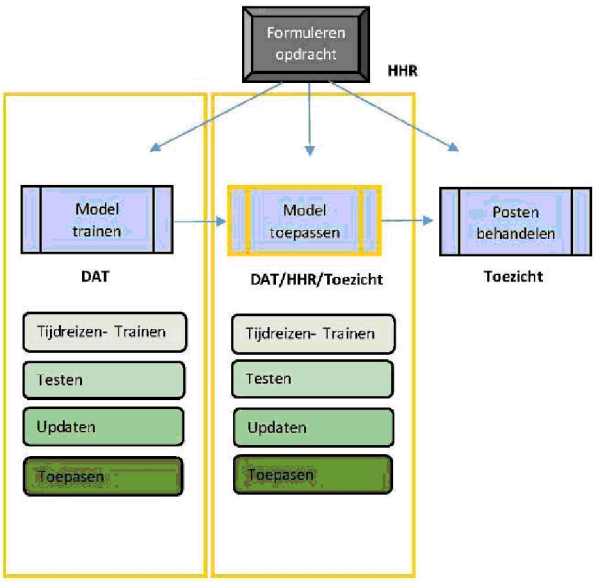
[42] De indicatoren zijn tijdens de bouwfase van het model gecreëerd tijdens een brainstormsessies tussen HHR en de behandelaars van het Toezicht team. Die selectie was gebaseerd of huidige en eerdere ervaringen uit het toezichtpraktijk.

grouping[43] wordt elke indicator onderverdeeld in bins, waarna bepaald wordt of de indicator verklarend genoeg is om te voorspellen of een beschikking goed of fout verklaard wordt (volgens de Gini coëfficiënt[44]).

Door middel van een logistische regressiemodel wordt een scorecard[45] gemaakt, waaruit blijkt hoeveel weging elke bin heeft in het voorspellen van goed of fout. De som van de gescoorde indicatoren geeft de eindscore, waarvan een hogere score betekent dat de aanvraag een relatief hogere kans op fout heeft dan een lagere score.
De resulterende scorecard wordt opgeslagen als code in SAS en als een Excel bestand op de Q-schijf.

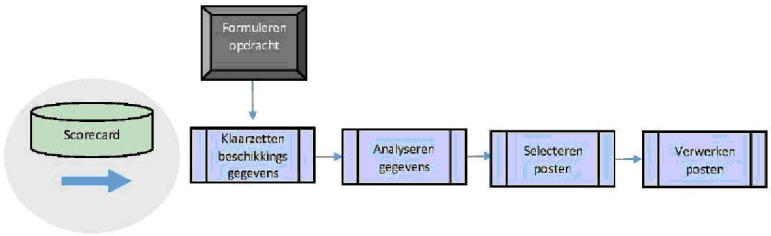
De scorecard wordt daarna gebruikt om de risicoscore van de aanvragen te bepalen (zie onderdeel B. hieronder voor toepassing van het model).

B. Toepassen van het model



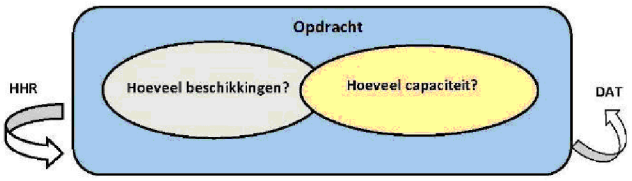
Hoofdproces

[43] Zie voor een beschrijving:
https://documentation.sas.com/?docsetId=emref&docsetTarget=p1qzwz7onopjgcn11uc04i18urg7.htm&docsetVersion=15.1&locale=en
[44] Gini coefficient: ook wel Gini-index, is een statistische maatstaf van de ongelijkheid in een verdeling.
[45] Zie voor een beschrijving:
https://documentation.sas.com/?docsetId=emref&docsetTarget=n181vl3wdwn89mn1pfpqm3w6oaz5.htm&docsetVersion=14.3&locale=en



Proces	Omschrijving	Organisatie
1. Formuleren opdracht	Een intelligence analist van HHR geeft opdracht aan DAT een bepaald aantal toeslagaanvragen te selecteren met de hoogste kans op fout. Een controller van productieregie bepaalt wat de behandelnorm per type post is en een intelligence analist bepaalt hoeveel (afhankelijk van de behandelcapaciteit er beschikbaar is die maand) en waaraan aanvragen moeten voldoen om in aanmerking te komen voor selectie door het model. Een data analist voert de daadwerkelijke selectie uit.	HHR, PR
2. Voorbereiden scorecard	Geautoriseerde medewerkers van DAT bepalen met behandelde posten welke eigenschappen van een aanvraag voorspellend zijn voor het fout zijn van een aanvraagbeschikking, wat resulteert in een scorecard (zie trainen van het model)	DAT
3. Klaarzetten beschikkingsgegevens	DAT ontvangt van verschillende bronnen gegevens in een beveiligde omgeving en koppelt deze gegevens aan de beschikkingen	DAT
4. Analyseren gegevens	Geautoriseerde medewerkers van DAT analyseren de beschikkingen aan de hand van gekoppelde gegevens en de scorecard, wat resulteert in gerangschikte beschikkingen	DAT
5. Selecteren posten	Geautoriseerde medewerkers van DAT selecteren te behandelen posten aan de hand van de gerangschikte beschikkingen en behandelcapaciteit. Een Toezicht medewerker voert een steekproef uit op de geselecteerde posten om te controleren of de verwachte risicoscore juist is. Geautoriseerde intelligence analisten van HHR controleren het resultaat en herinitiëren het proces als nodig. Als alle stappen volgens verwachting zijn verlopen, biedt de intelligence analist de gewenste aantallen en welke posten niet meegenomen hoeven te worden aan, aan een data analist.	DAT, Toezicht, HHR
6. Verwerken posten	Geautoriseerde medewerkers van DAT leveren de beschikkingen aan het Toeslagen queryloket. Het queryloket maakt werkopdracht aan waaruit bundels volgen die worden aangeleverd aan CAP-UPG om ingelezen te worden in TVS voordat het formeel beschikken gaat draaien. Als ze in TVS staan, worden ze door Toeslagen werkvoorbereiding aangeboden aan medewerkers van IST om de geselecteerde posten te behandelen	DAT, Toezicht

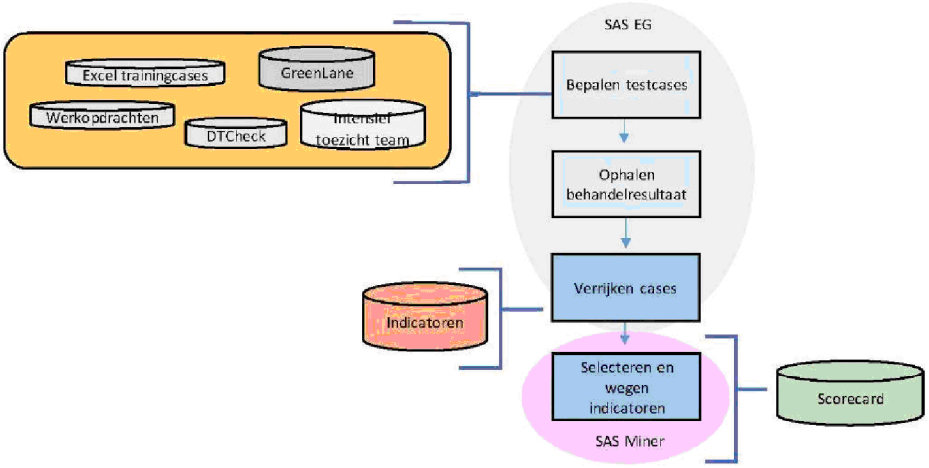
1. Formuleren opdracht



Proces	Omschrijving	Organisatie
1.1 Behoefte aan het classificeren van beschikkingen wordt uitgesproken	HHR bepaalt dat beschikkingen voor huurtoeslag, kinderopvangtoeslag en KGB dienen te worden classificeren.	HHR
1.2 Beschikbare capaciteit wordt bepaald	HHR en PR[46] bepalen hoeveel posten hoog risico KOT, hoog risico HT en hot-hor[47] aangeboden worden voor handmatige behandeling. Het document waarin dit is vastgelegd wordt gedeeld met DAT per e-mail. Dit document bevat geen persoonsgegevens.	HHR, PR

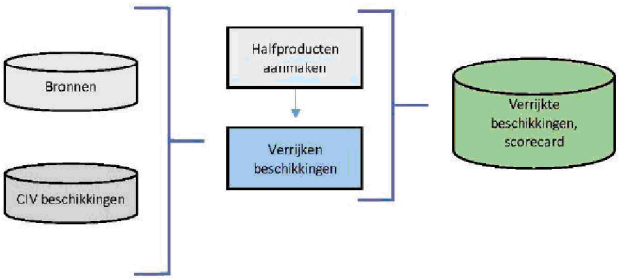
2. Voorbereiden Scorecard

[46] Productieregie
[47] Hot Hor: hoge teruggaaf/hoog risico



Het proces om de scorecard voor te bereiden is reeds beschreven in sectie A. Het trainen van het model. Het doel van het voorbereiden van de scorecard is het bepalen welke indicatoren[48] voorspellen of een aanvraag goed of fout gekeurd wordt en hoe zwaar deze indicatoren wegen. Het resultaat is een scorecard met te beoordelen indicatoren en hun weging. Randvoorwaarden voor een betrouwbare scorecard zijn een frequente toevoer van nieuwe testcases en voldoende gegevens over behandelde aanvragen en aanvragers. Hiermee blijft de trainingset aan testcases up-to-date, zodat de scorecard nieuwe types fouten kan blijven vaststellen. Aan de hand van de resulterende scorecard vindt in een vervolgstap de beoordeling plaats van de relatieve kans op het foutief bevinden van nieuwe aanvragen.
Het model wordt sinds een jaar geleden niet meer getraind maar het proces wordt herhaald voor het voorbereiden van de scorecard. Steekproeven worden weinig uitgevoerd vanwege gebrek aan capaciteit.

3. Klaarzetten beschikkingsgegevens



Het doel van het klaarzetten van beschikkingsgegevens is het verzamelen van gegevens over de aanvraag en aanvrager voor analyse en filtering in vervolgstappen. DAT ontvangt daarvoor ruwe

1075458 00042

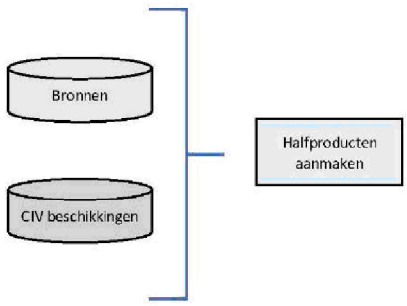
gegevens van dataleveranciers (zie bijlagen 1) en zet deze om tot koppelbare data. De samenvoeging van deze data, al koppelbare gegevens uit het datafundament en beschikkingen, vormt een grote dataset. De gegevens in deze dataset maken het mogelijk een uitspraak te doen over de relatieve kans op het foutief beoordelen van de aanvraag. Ook kan met deze gegevens in een vervolgstap een filtering plaatsvinden op basis van criteria die HHR stelt.

Proces	Omschrijving	Organisatie
3.1 Halfproducten aanmaken	DAT ontvangt van verschillende bronnen (zie bijlage 1) gegevens en bereidt deze voor om te koppelen aan beschikkingen	DAT
3.2 Verrijken beschikkingen	Geautoriseerde medewerkers van DAT koppelen gegevens uit halfproducten en het datafundament aan beschikkingen	DAT

Brondata is afkomstig van diverse afdelingen binnen Toeslagen. De manier waarop deze data aangevraagd en verkregen wordt verschilt per dataset. In alle gevallen is de inhoud van de aanvraag bij de leverende partij bekend (meestal is sprake van gestandaardiseerde queries). De verversing van brondata impliceert in de regel dus enkel een herhaling van bekende zetten.
Per run wordt bepaald voor welke bronnen nieuwe data opgevraagd wordt. Voor iedere dataset die ververst moet worden, wordt aan de betreffende contactpersoon per email een nieuwe data-aanvraag gestuurd. Hiervoor wordt een opvraaglijstje bijgehouden.

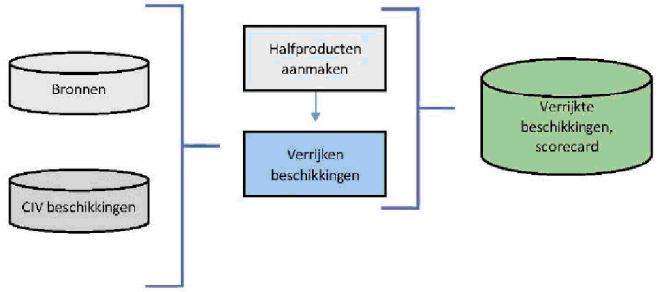
Gegevens worden vaak per e-mail ontvangen als Excel of csv bestanden (zie bijlage 1 en 2 "ontvangstmethode" bij bronnen en halfproducten).
In veel gevallen vindt levering van de data plaats via een gedeelde netwerkschijf. Gegevens uit CIV worden na het afronden van een officiële aanvraag en een WMK toets door CAP klaargezet in een uitwisselmap op de netwerkschijf. Vervolgens worden deze door een data-analist gekopieerd naar de SAS Grid van Toeslagen, waar alleen data-analisten die belast zijn met de risicoclassificatie bij kunnen. Na ontvangst wordt de brondata direct gecontroleerd op plausibiliteit en bruikbaarheid. Deze controles worden bijgehouden in de risicoclassificatie controlematrix (zie bijlage 5). Bij bevindingen zal overleg plaatsvinden met de leverende partij, deze blijft in principe verantwoordelijk voor de kwaliteit van de data.

3.1 Halfproducten aanmaken



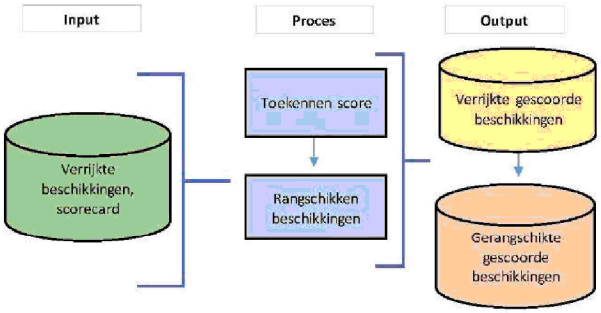
Na ontvangst van nieuwe brondata worden de hiervan afhankelijke halfproducten ververst. Elk halfproduct zet brondata om in een tabel die, vaak op het BSN, aan de (concept)beschikkingen gekoppeld kan worden. Het gaat in de halfproducten voornamelijk om 'contra-informatie': kenmerken van de aanvrager of aanvraag die in de vorm van indicatoren gebruikt worden als input voor het risicoclassificatiemodel.
Alle halfproducten bestaan uit een enkel SAS Enterprise Guide project. In de meeste gevallen zal het betreffende project hiervoor niet of nauwelijks aangepast hoeven te worden, en zal enkel het betreffende project opnieuw gedraaid moeten worden.
De verwerking wordt gecontroleerd op fouten en implausibele resultaten. De resultaten van deze controles worden bijgehouden in de risicoclassificatie controlematrix.
Nadat het proces gedraaid heeft worden de halfproducten in het model (HT_model en KOT_model) aan de beschikkingen gekoppeld.

3.2 Verrijken beschikkingen



Koppeling van details uit de aanvraag vanuit de (concept)beschikkingen en contra-informatie vanuit de halfproduct. De gegevens uit de halfproducten worden gekoppeld aan de gegevens van de beschikking dataset die op het datafundament staat. Met deze data worden bepalingen gedaan zoals afstand KOO-aanvrager, hoogste maandbedrag, etc.

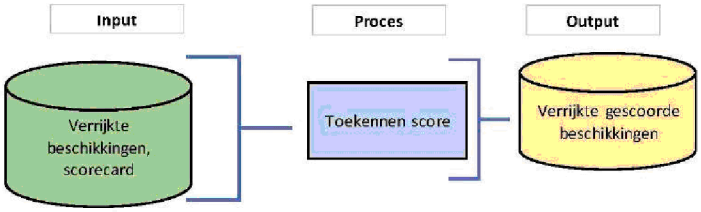
4. Analyseren Gegevens



Proces	Omschrijving	Organisatie

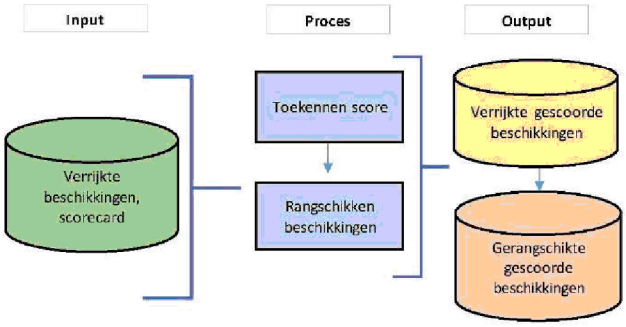
4.1 Toekennen score	Geautoriseerde medewerkers van DAT kennen per beschikking een score toe op basis van de scorecard	DAT
4.2 Rangschikken beschikkingen	Geautoriseerde medewerkers van DAT rangschikken gescoorde beschikkingen op hoogste kans van foutief	DAT

4.1 Toekennen score



Toekenning van risicoscores aan alle (concept)beschikkingen aan de hand van het getrainde risicoclassificatiemodel. De output van deze stap is de ruwe output van het model. Voor alle in het model gebruikte indicatoren is het aantal ScoreCard-punten bepaald.
De scorecard wordt gebruikt om per BSN elke verklarende variabele een score te geven. De scores van verklarende indicatoren worden bij elkaar opgeteld. De hoogste score is de relatief hoogste kans op een foute aanvraag en vice versa.

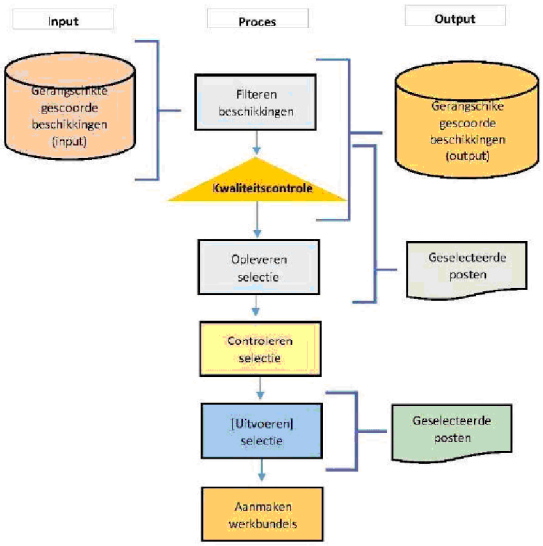
4.2 Rangschikken beschikkingen



Door rangschikken is het mogelijk een selectie te maken van de groepsaanvragen waar de verwachte hitrate het hoogste is.

Het resultaat wordt als een Excel bestand op de netwerkschijf opgeslagen.
Om te finaliseren worden de gescoorde aanvragen nabewerkt zodat aan elke aanvraag een opvolgactie toebedeeld kan worden, en de diverse vormen van output samengesteld worden.

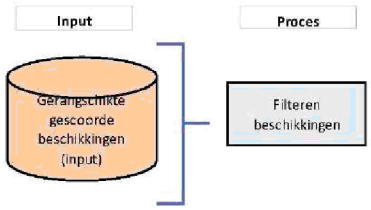
5. Selecteren Posten



Proces	Omschrijving	Organisatie
5.1 Filteren beschikkingen	Geautoriseerde medewerkers van DAT filteren beschikkingen op basis van criteria waaraan een beschikking moet voldoen om in aanmerking van selectie te komen	DAT
5.2 Kwaliteitscontrole	Het proces zelf en de output worden gecontroleerd aan de hand van de controlematrix	DAT
5.3 Opleveren selectie	Geautoriseerde medewerkers van DAT leveren aan een intelligence analist van HHR een lijst met geselecteerde beschikkingen	DAT, HHR, PR
5.4 Controleren selectie	Een Toezicht medewerker voert een steekproef uit op de geselecteerde posten om te controleren of de verwachte risicoscore juist is, geautoriseerde intelligence analisten van HHR controleren het resultaat en koppelt terug als een aanpassing gewenst is.	Toezicht, HHR, PR
5.5 Uitvoeren selectie	Geautoriseerde medewerkers van DAT leveren een definitieve selectie van beschikkingen aan een intelligence analist van HHR en stuurt de geselecteerde posten aan TQL[49] om te verwerken	DAT, HHR, TQL
5.6 Aanmaken werkbundels	TQL zorgt ervoor dat werkbundels worden aangemaakt	TQL

[49] Toeslagen Query Loket

5.1 Filteren beschikkingen – (opvolgacties)



Behalve de risicoscore zijn er nog een aantal kenmerken van de aanvraag of aanvrager die bepalen wat er met de conceptbeschikking gebeurt. Deze stap is ook nodig om te bepalen of een aanvraag onderdeel van de risicoselectie is of niet en om de juiste behandeling aan de burger te geven:

- Is de toeslagaanvraag 'op nul gezet'
- Is voor het BSN van de aanvrager recent een bezwaar afgehandeld?
- Is de aanvrager een 'onbekende aanvrager'?
- Wordt de betreffende BSN voor de betreffende regeling al uitgesloten van beschikken (staat BSN op de 'uitsluitlijst') of was dit recent het geval?
- Is sprake van geen recht op toeslag vanwege 'bijvangst'?
- Is de laatste beschikking een conceptbeschikking?
- Is de aanvraag nieuw of leidt een mutatie tot een hogere toeslag dan voorheen?
- Vormt de aanvraag vanwege een hoog bedrag een hoog risico ('Hot Hor' – alleen voor KOT)?
- Is er een overledene bij de aanvraag betrokken?

Deze kenmerken worden per toeslagaanvraag in het model vastgesteld. Vervolgens wordt aan de hand van deze kenmerken per aanvraag bepaald wat de relevante opvolgactie is. De volgende opvolgacties zijn mogelijk:

- Geen Actie
- Op uitsluitlijst vanwege onbekende aanvrager
- Op uitsluitlijst vanwege Hot Hor (alleen KOT)
- Op uitsluitlijst vanwege bijvangst
- Op uitsluitlijst vanwege hoog risico

Op nul gezet

'Op_nul_gezet' is een indicator die aangeeft of de aanvraag een actuele conceptbeschikking met een jaarbedrag van 0 euro is (1=Ja, 0=Nee). De bepaling van deze indicator vindt plaats in DTM_HT en DTM_KOT.

Als de aanvraag op nul gezet is, dan is de afgeleide conclusie dat het gaat om een aanvraag die door een toeslagmedewerker behandeld is, waarbij de conclusie is doorgevoerd dat geen sprake is van een rechtmatige aanvraag.

Om te voorkomen dat een aanvraag die op nul gezet is, vervolgens door de risicoclassificatie uitgesloten wordt van beschikken, volgt op deze aanvragen nooit een uitsluiting en is de actie altijd 'Geen Actie'.

Recent afgehandeld bezwaar
Het halfproduct 'bezwaren' geeft een lijst van BSN's waarvoor recent (sinds de vorige run van formeel beschikken van het betreffende Toeslagjaar) door Toeslagen een bezwaar is afgehandeld. De kans is groot dat een conceptbeschikking voor een BSN met een afgehandeld bezwaar een consequentie is van deze afhandeling (dit kan zowel toekenning als afwijzing van het bezwaar zijn). Om te voorkomen dat in die situatie de conceptbeschikking uitgesloten wordt, volgt in het geval van een recent afgehandeld bezwaar nooit de actie 'uitsluitlijst'.

Onbekende aanvrager
Voor de zorgtoeslag worden handmatige controles uitgevoerd op het zorgverzekerd zijn van burgers. Om risicogericht toezicht uit te voeren wil directie Toeslagen voor aanvragers zorgtoeslag weten of zij wel recht hebben op deze toeslag.
Onbekende aanvragers zij burgers die woonachtig zijn op een Nederlands adres maar zijn uitgesloten voor formeel beschikken omdat zij de status onbekende aanvragers hebben.
Een onbekende aanvrager is een aanvrager:

- met een BSN dat jonger is dan 1 jaar;
- waarvoor geen enkele toeslagregeling een formele beschikking met een jaarbedrag groter dan 0 euro is uitgereikt;
- na de vorige risicoclassificatierun een nieuwe conceptbeschikking is ontstaan.

Deze aanvragers worden in het algemeen beschouwd als een hoog risico, omdat ze bij Toeslagen relatief onbekend zijn. Aanvragen van onbekende aanvragers worden daarom in beginsel uitgesloten van beschikken. De onbekende aanvragers worden hierop uitgenodigd op een Belastingkantoor om bewijsstukken (bijv. huurcontract of overeenkomst met kinderopvanginstelling) te overleggen.

Bijvangst
Bijvangst zijn toeslagaanvragen waarvoor op basis van harde criteria (op basis van de Awir) vastgesteld kan worden dat in geen geval sprake zal zijn van recht op toeslag.
Voor HT zijn deze criteria:

- Voor de BSN is in de BVR geen adres bekend (Vertrokken Onbekend Waarheen - VOW) of
- de aanvraag heeft betrekking op een adres dat is uitgesloten voor huurtoeslag (vastgesteld obv halfproduct 'uitgesloten adres')

Voor KOT zijn deze criteria:

- De aanvrager woont niet in Nederland, België of Duitsland of
- Een kind waarvoor KOT wordt aangevraagd woont niet in Nederland, België of Duitsland of
- De leeftijd van het oudste kind waarvoor KOT wordt aangevraagd is hoger dan 15.

Omdat bij bijvangst geen sprake kan zijn van recht op toeslag wordt de betreffende aanvraag in principe uitgesloten.

Uitzonderingen zijn:

- Er is een overledene bij de aanvraag betrokken. Om pijnlijke situaties te voorkomen wordt in dat geval niet uitgesloten.
- Er is een recent afgehandeld bezwaar voor de aanvrager.
- De beschikking is geen concept – in dat geval valt er niets uit te sluiten.
- Het betreft geen nieuwe aanvraag en geen mutatie met een hoger bedrag dan voorheen.

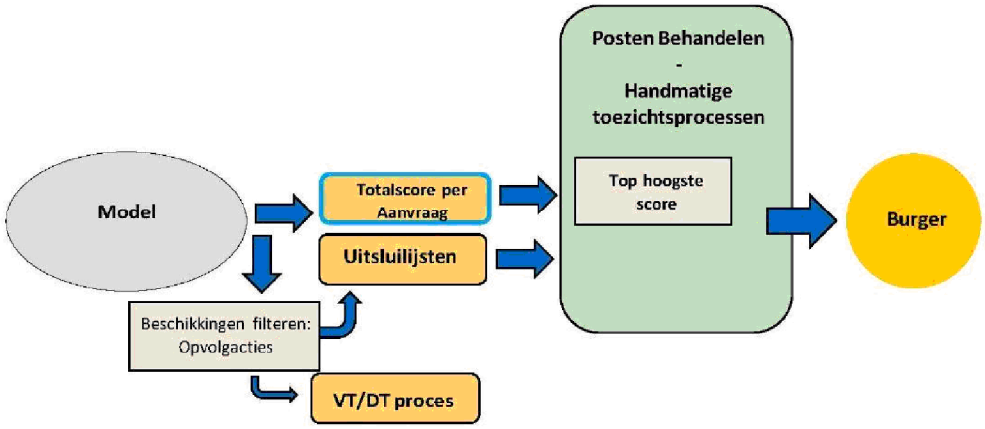
Uitsluitlijst
De uitsluitlijst is een lijst van BSN's die voor een bepaalde toeslagregeling en voor een bepaald tijdvak uitgesloten worden van beschikken. De risicoclassificatieruns leveren input voor deze uitsluitlijst, maar de uitsluitingen kunnen ook uit andere onderdelen van de organisatie komen.

1075458 00042

In het proces worden historische uitsluitingen gebruikt om te voorkomen dat iemand die op het moment van selectie uitgesloten is, of dat recent was, voor een tweede keer of nogmaals uitgesloten wordt. Per run wordt echter inhoudelijk getoetst of die vertaling passend is.
Een belangrijke aanname is dat iemand die recent uitgesloten is en op het moment van selectie niet meer, door Toeslagen behandeld en daarom vrijgegeven is. Na een dergelijke vrijgave is het niet de bedoeling dat de conceptbeschikking alsnog uitgesloten en geblokkeerd wordt. Alle BSN's die recent zijn uitgesloten van beschikken, worden daarom niet opnieuw uitgesloten.
In theorie is het mogelijk dat na een afvoering van de BSN van de uitsluitlijst, de aanvrager een mutatie opgeeft die leidt tot een nieuwe conceptbeschikking. Die mutatie wordt door de risicoclassificatie vanwege bovenstaande implementatie niet op risico geselecteerd. Omdat bij zowel de opvoering als afvoering van een BSN niet geregistreerd wordt wat de reden voor opvoering dan wel afvoering was, en bij de generatie van een conceptbeschikking niet bekend is of de beschikking een gevolg is van een handeling van de aanvrager of een medewerker van Toeslagen, kan hier in het proces geen rekening mee gehouden worden.

HOT-HOR:
De posten hoog risico KOT, hoog risico HT en HOT-HOR worden aangeboden voor handmatige behandeling. Dit zijn actuele conceptbeschikkingen met een hoog bedrag (dus een hoog risico). Deze stap heeft ook als doel het beschermen van de burger zodat geen hoge terugvorderingen worden ontvangen.

De outputbestanden (uitsluitlijsten) worden gebruikt binnen het handmatig administratief Toezicht proces. De mogelijke risico's die kunnen ontstaan door deze lijsten is al geanalyseerd in de GEB M861 Behandelen Fraudesignalen.

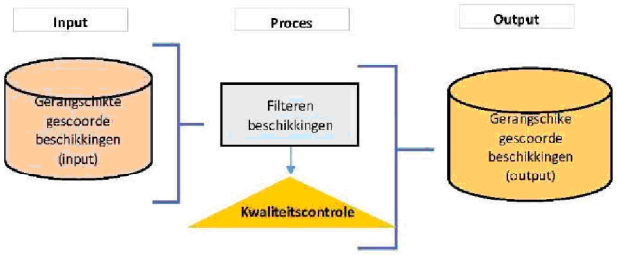


Als resultaat van het filteren van beschikkingen zijn de volgende acties mogelijk:

- Een aanvraag wordt gemarkeerd als uitworp (uitsluitlijst) wanneer op grond van de risicogerichte beslisregels/kenmerken twijfel bestaat over de rechtmatigheid. Dat betekent dat nadere handmatige controle van de berekende toeslag is nodig. Dit proces wordt verder gedetailleerd in GEB Handmatige behandeling van onvolledige en/of tegenstrijdige toeslagenaanvragen (Uitval/uitworp) M2077.

- Een aanvraag wordt binnen het VT /DT[50] proces behandeld (dit is een geautomatiseerd proces; zie verder GEB Definitief Toekenning Toeslag M675 en Automatisch continueren (MAC) toeslag M1647)

5.2 Kwaliteitscontrole

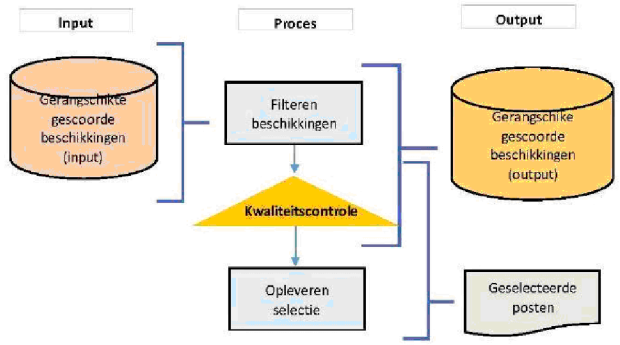


Nadat het model gedraaid is wordt de inputdata, het proces zelf en de output gecontroleerd aan de hand van de controlematrix. Pas nadat alle controles afgerond zijn, worden de diverse vormen van output met de rest van de organisatie gedeeld.

Om de kans op fouten tijdens een risicoclassificatierun zo klein mogelijk te maken is een controlematrix samengesteld (zie bijlage 5).
De matrix geeft een lijst met handmatig uit te voeren checks. Deze checks zijn open geformuleerd, er is in de meeste gevallen geen sprake van een norm of gekwantificeerd criterium waaraan voldaan moet zijn. Dit legt de verantwoordelijkheid voor de invulling van de controles bij de controleurs neer. Deze controleurs zullen dus vanuit hun expertise moeten beoordelen of aan de benodigde voorwaarden voldaan moet worden.

De controles voor tijdreizen en de modellen (HT en KOT) worden in principe onder een vierogenprincipe uitgevoerd om fouten te voorkomen die grote gevolgen voor de kwaliteit van het model kunnen hebben. De andere controles worden uitgevoerd door de collega (DAT) die het betreffende proces tijdens de risicoclassificatierun behandelt.

5.3 Opleveren selectie



[50] Voorlopige Toekenning / definitief Toekenning

De definitieve run resulteert in verschillende vormen van output:

- Een inhoudelijke uitsluitlijst ('met reden'), die gebruikt wordt om aan de uitsluiting opvolging te geven (bijvoorbeeld in communicatie met de betreffende burger). De reden geeft aan op welke grond de post is geselecteerd.
- Een overzichtstabel, waarin de aantallen posten over een aantal categorieën gegeven worden getoond. Deze tabel vormt de basis voor een rapportage over de resultaten van de risicoclassificatierun.
- Voor KOT: een lijst met alle Hot Hor posten die echt behandeld gaan worden.
- Een eindbestand waarin per BSN alle risicoscores zitten.

De bestanden worden in SAS bewaard en in csv formaat op de Q-schijf als back-up opgeslagen.

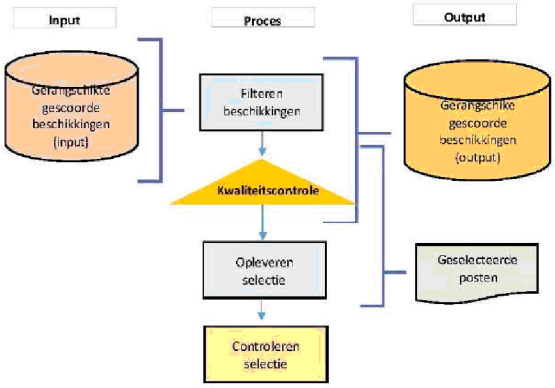
De resultaten van een run worden per e-mail opgeleverd naar de postbus van QueryLoket, Werkvoorbereiding en Kantoorregisseurs.
In de e-mail worden de volgende bestanden opgeleverd in csv formaat:

- De uitsluitlijsten (één per regeling KOT en HT)
- De overzichtstabel
- De lijst met alle Hot Hor posten

Deze bestanden heeft het Queryloket nodig om de posten uit te sluiten en Productie Regie om werkbundels aan te maken.

De manier hoe de bestanden worden opgeleverd zal in Q2 2021 worden aangepast volgens de nieuwe richtlijnen voor veilig uitleveren van AU[51] (zie Risico 8 en 9 in Hoofdstuk 16 en 17 Risico's en Maatregelen)

5.4 Controleren selectie



Iedere maand worden Hoog Risico posten uitgezet ter controle. Deze posten worden geselecteerd op basis van beschikbare toezichtcapaciteit en de risicoscore. Om de selectie te valideren wordt vóór het definitief selecteren van de posten, steekproeflijsten opgesteld door HHR en uitgezet bij de Toezicht teams. Vanuit de selectie worden meerdere lijsten/acties gedraaid, deze worden gecontroleerd door HHR & Toezicht.

[51] Analytics United

1075458 00042

Proefrun en definitieve run
Voor het draaien van de proefrun bespreken de data en intelligence analisten met elkaar welke aantallen er opgeleverd gaan worden. Bij de proefrun wordt er een aantal Excel lijsten uitgezet ter controle bij het Toezicht team. De bestanden worden per e-mail gestuurd. Als de ingevulde Excel lijsten terug gestuurd worden bereken HHR aan de hand van de hoog risico-lijsten het percentage posten dat door de steekproevers als risicovol wordt gezien.

- ✓ Er wordt met kwaliteitsregie afgestemd of er een kwaliteitscheck is gedaan.
- ✓ Het percentage risicovol per regeling wordt uitgerekend. De uitkomsten en redenen worden afgestemd met de data-analisten en er wordt een beslissing over de selectie bij de definitieve run genomen.
- ✓ De uitkomsten van het verslag worden besproken met de data-analisten en gedocumenteerd.

De steekproef heeft verschillende doeleinden. Het is een check om te beoordelen of een toeslagaanvraag terecht door het risicoclassificatiemodel als risicovol is aangemerkt. De behandelaar controleert of de aangeboden BSN's daadwerkelijk risicovol zijn. Dit doet hij o.b.v. enkele factoren (bijv. voor KOT: hoge opvanguren, verhouding inkomen - aantal opvanguren - kosten en voor Huur: verhouding inkomen - huurprijs, ronde huurprijzen, hoge servicekosten, evt. oppervlakte woning en aantal personen op de beschikking). Dit is belangrijk want het model selecteert de posten voor Risicoclassificatie en impliciet dus ook waar wij toezicht op inzetten, de controle door de behandelaar zorgt er voor dat die kan bevestigen of het model de juiste/risicovolle aanvragen eruit haalt. Daarnaast worden de resultaten van de steekproef gebruikt om de verdeling tussen Hoog Risico HUUR, Hoog Risico KOT en HOT-HOR te bepalen. Het kan namelijk voorkomen dat in de ene maand de Hoog Risico HUUR risicovoller is dan Hoog Risico KOT, met tot gevolg dat er naar verhouding meer HUUR uitgezet wordt.

Onderdeel van de steekproef is de validatiecheck, waar wordt gekeken naar het percentage van de posten di terecht als hoog risico geselecteerd zijn. Deze validatiecheck wordt door een van de behandelaren van het Toezicht team per e-mail naar HHR en DAT gestuurd. Zie voorbeeld:

Categorie	Dimensie	Aantal
Steekproef	Percentage van steekproef terecht geselecteerd als hoog risico	98
Steekproef	Posten uit populatie getrokken in steekproef	49

Als onderdeel van het validatieproces wordt naar de 'hitrate' gekeken: het percentage posten wat gecorrigeerd is na behandeling.

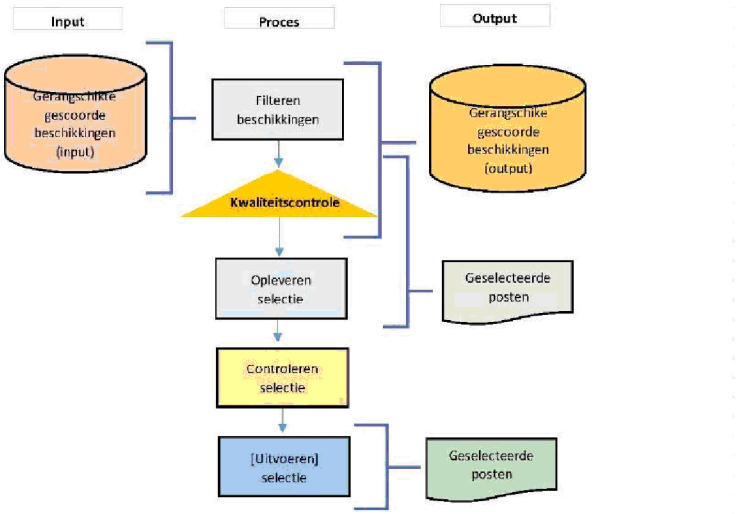
Naast hitrate wordt er een statistische toets op accuratesse gedaan (gini-score: als die te laag is, is het model te weinig verklarend).

De combinatie van de hierboven benoemde technieken biedt extra betrouwbaarheid omdat het niet alleen de verklarende waarde van je model bekijkt, maar ook de werking in de praktijk. Een model kan heel goed verklarend zijn maar niets opleveren.

Momenteel wordt de proefrun op de dag van de definitieve run uitgevoerd. Een verandering ten opzichte van de runs van voor november 2019 is dat er geen maandelijkse steekproef wordt uitgevoerd op de nieuwe aanvragen/mutatie Hoog Risico KOT en HUUR. Deze steekproef wordt enkel uitgevoerd als het model getraind wordt. Redenen die hebben geleid tot aanpassingen in

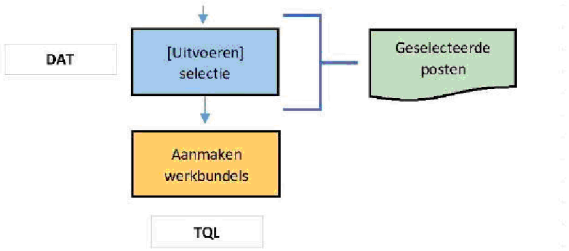
steekproeven zijn o.a. de afnemende toezichtcapaciteit, risicogerichte keuzes, dat steekproeven veel tijd vergen en dat HHR meer kan met de beschikbare data.

5.5 Uitvoeren selectie



Na de validatie van de selectie levert DAT na een definitieve run een definitieve selectie van beschikkingen aan HHR. De geselecteerde posten worden per e-mail naar de postbus van TQL verstuurd om te verwerken.

5.6 Aanmaken werkbundels



TQL zorgt ervoor dat werkbundels worden aangemaakt zodat behandelaars de posten verder kunnen controleren.

6. Verwerken posten