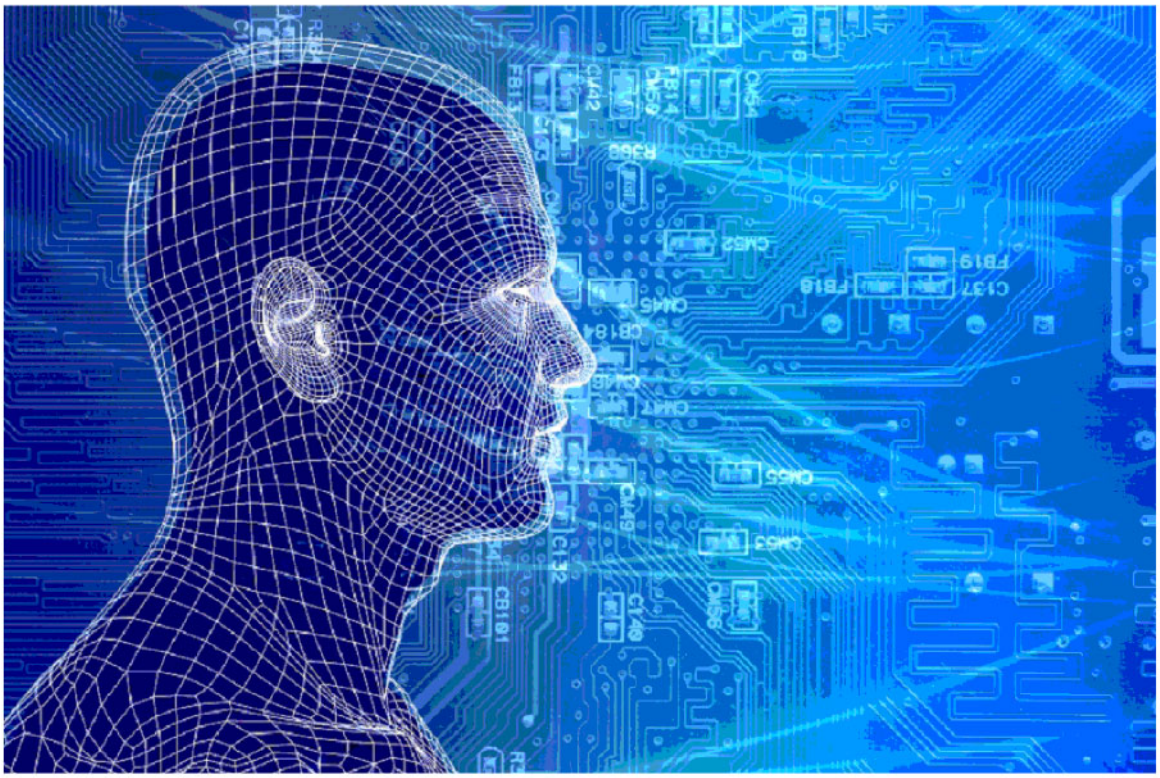


HT_MODEL

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	03-07-2013	10.2.e	Initieel document
1.0	23-12-2015		Aanvullen informatie

1075478 00062

1. INHOUDSOPGAVE

1. Inhoudsopgave 3
2. Beschrijving proces 4
2.1. Input 4
2.1.1. Brondata 4
2.1.2. Halfproducten 5
2.2. Flows 8
2.2.1. 1. Init 8
2.2.2. 2. Combine 8
2.2.3. 3. Score model 10
2.2.4. 4. Finalize 11
2.2.5. 7. Rapportage 16
2.3. Checklist Managementrapportage 20

2. BESCHRIJVING PROCES[1]

Het SAS-script HT_MODEL.egp wordt op de dag van de risicoselectie gedraaid door het risicoselectieteam. Dit resulteert in meerdere outputbestanden, waaronder 'export_temp3', wat als het eindbestand van de risicoselectie wordt beschouwd. Ook wordt een xlsx-bestand geëxporteerd met kengetallen voor de Managementrapportage.

Zowel de master/productie (!) versie van het SAS-script als de resulterende bestanden staan opgeslagen op de C-schijf van de stand alone laptop:

Script:

C:\Risico Selectie\Model \HT_MODEL\SAS Enterprise Guide

Outputbestanden:

C:\Risico Selectie\Model \HT_MODEL\Results

Exportbestand:

C:\Risico Selectie\Model\HT_MODEL\Export

De master versie wordt periodiek handmatig gebackupt naar:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Model\Backup HT_Model

2.1. INPUT

De input in het SAS-script betreft zowel brondata als halfproducten:

2.1.1. BRONDATA

De bronbestanden gebruikt als input voor het model zijn:

1. **Conceptbeschikkingen_ht.sas7bdat**
   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\HT CONCEPTBESCHIKKINGEN

2. **Beschikkingen.sas7bdat**
   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\BESCHIKKING HT

[1] Na overgang tot AWS zullen de locatie en namen brondata/halfproducten/SAS-scripts moeten worden aangepast.

1075478 00062

3. **Nieuwe_aanvragers.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\ONBEKENDE AANVR

4. **Wanbetalers.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\WANBETALERS

5. **Holland2stay.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata 10.2.g

6. **Privacy_doelgroep.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\VIP

7. **Schema_wie_op_welke_lijst_v6.1.csv**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\indeling resultaat

8. **Terugkoppeling steekproef risicoselectie.xlsx**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\Terugkoppeling steekproef

2.1.2. HALFPRODUCTEN

De halfproducten gebruikt als input voor het model zijn:

1. **DTM_HT.sas7bdat**

   Voor het bestand en de documentatie zie de mappen:
   C:\Risico Selectie\Model \DTM_HT\SAS Enterpriseguide (Stand alone laptop)
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\DTM_HT

2. **COA.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\COA

3. **BVR.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\BVR

4. **BAG.sas7bdat**

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\BAG

5. **Meldingenfrs.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\MELDINGFRS

6. **Signaallijstadressen.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\UITGESLOTEN_ADRES

7. **Melding8_huur.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\MELDING8

8. **Vha_huur_halffabrikaat.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\VHA

9. **Woz_halffabrikaat.sas7bdat**

   Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\WOZ

10. **Uitval_uitworp_ht.sas7bdat**

    Voor het bestand en de documentatie zie de map:
    Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\UITVAL

11. **Uitval_uitworp_ht_53.sas7bdat**

    Voor het bestand en de documentatie zie de map:
    Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\UITVAL

12. **IP_lijst.sas7bdat**

    Voor het bestand en de documentatie zie de map:
    Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\IP_LIJST

13. **Aangeleverd.sas7bdat**

    Voor het bestand en de documentatie zie de map:
    Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Bezwaren

14. **Bezwaren_filter_1.sas7bdat**

    Voor het bestand en de documentatie zie de map:
    Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Bezwaren

15. Boete.sas7bdat

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\BOETE

16. Nihilstellers.sas7bdat

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Nihilstellers\Results

17. Vorige_acties_huur.sas7bdat

Voor het bestand en de documentatie zie de mappen:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Terugkoppeling
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\control_klantbehandeling

18. Vluchtelingenwerk.sas7bdat

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\vluchtelingenwerk

19. IGS.sas7bdat

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\IGS

20. RSR_teruglevering_ht.sas7bdat

Voor het bestand en de documentatie zie de map:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\control_klantbehandeling

2.2. FLOWS

Het SAS-script bestaat uit negen procesflows:

1. Init
2. Combine
3. Score Model
4. Finalize
5. Steekproef
6. Uitsluitlijst
7. Rapportage

Checks

Eerdere acties

Het runnen van deze flows resulteert in de outputbestanden en het exportbestand. Onderstaand zijn flows 1. Init, 2. Combine, 3. Score Model, 4. Finalize, en 7. Rapportage beschreven.

De process flows 5. Steekproef, 6. Uitsluitlijst, Checks en Eerdere acties geen onderdeel (meer) zijn van een reguliere risicoselectierun.

2.2.1. 1. INIT

In deze flow worden de scripts geïnitialiseerd.

1. Init

De libraries en variabelen worden gedefinieerd.

2. Macros

De macro's voor het uitvoeren van bijv. herhalende stukken sas-code worden gedefinieerd:

- ExportData, ExportData_zonderheader: standaardmacro voor export van SAS bestanden. Wordt diverse keren gebruikt in process flow 4. Finalize.
- GenerateAdditionalVariables: macro waarin voor geselecteerde variabelen (opgenomen in HT_dataset_metadata.txt) gecheckt wordt of deze missing zijn, en hiervoor een 'missing indicator' (M_*variabele naam*) wordt aangemaakt. Wordt gebruikt in script 'Finalize deduplicate' in process flow 2. Combine. Dit is specifiek voor HT, bij KOT komt deze macro niet voor.
- Normaliseer: macro om score card punten voor alle SCR_ variabelen te normaliseren. Dit maakt de scores op variabelen beter vergelijkbaar en interpreteerbaar. Wordt gebruikt in script 10_Normaliseren in process flow 4.Finalize.

2.2.2. 2. COMBINE

In deze flow worden de beschikkingen met de variabelen uit de halfproducten gecombineerd. Dit gebeurt door het koppelen van de halfproducten en brondata. Verder worden een aantal laatste indicatoren vastgesteld en vindt er een correctie van missing values plaats.

Koppelingen

1: Adresgegevens uit het halfproduct BVR worden gekoppeld aan DTM_HT.sas7bdat op BSN. Dit resulteert in de tabel **HT_MOD_ADRESSEN**.

2: De gegevens uit het halfproduct BAG worden gekoppeld aan **HT_MOD_ADRESSEN** op adres (variabele PHT). Dit wordt opgeslagen in de tabel **HT_MOD_Adressen_BAG** welke een nieuw veld bevat: BAG_MISSING.

3: De uitgesloten adressen (signaallijstadressen,sas7bdat) worden gekoppeld uit de data set verkregen uit de vorige stap. De uiteindelijke dataset heet **HT_MOD_Adressen_BAG_SIGNAALLIJST.sas7bat** en bevat het veld extra kolommen: av_ht_uitgesloten_adres.

4: De gegevens uit het halfproduct VHA worden erbij gekoppeld, zodat de opgegeven huren vergeleken kunnen worden met de huren uit het VHA. De koppeling gebeurt op adresKey. De volgende kolommen zijn extra toegevoegd:

- Perc_huur_VHA_huur
- MISSING_VHA_HUUR
- UIT_VHA_BANDBREEDTE

5: Meldingenfrs wordt gekoppeld aan het bestand verkregen uit de vorige stap op BSN.

6: De lopende vorderingen (halfproduct COA) worden gekoppeld op BSN. Dit resulteert in de tabel **HT_MOD_COA.**

7: Het bestand **twk** (terugwerkende kracht) wordt gecreëerd op basis van de gegevens uit het bronbestand 'Conceptbeschikkingen_HT'. Voor elk BSN (toeslagaanvrager) wordt de minimum waarde van het veld 'DatumIngang' en het veld 'Transactietijdstip' meegenomen.

8: HT_MOD_COA wordt gekoppeld aan de halfproducten Melding8, Wanbetalers, Boete, Nihilstellers en het bestand **twk** op het veld BSN. Hierbij zijn een aantal extra velden toegevoegd:

- Wanbetalen_zorg
- Boete_beschikkingen
- Nihistelling_KOT
- Aant_nihilstelling_KOT

Het resultaat wordt opgeslagen in de tabel **HT_MOD_MELDING8.**

9: De halfproducten Uitval_Uitworp_HT en Uitval_uitworp_HT_53 worden gekoppeld aan **HT_MOD_MELDING8** op BSN. Dit resulteert in het bestand **HT_MOD_UITVAL_UITWORP**.

10: De informatie uit het halfproduct WOZ_halffabrikaat.sas7bdat wordt gekoppeld aan **HT_MOD_UITVAL_UITWORP** om te bepalen of een aanvrager onroerend goed heeft. Ontdubbeling en sortering op BSN en geeft de data set **HT_MOD_FINAL** met nog een extra kolom:

- BSN_MET_WOZ_OP_AVADRES

11: Aan **HT_MOD_FINAL** worden nog een aantal velden toegevoegd met betrekking tot de huurprijs:

- Oppervlakteprijs
- Oppervlakteprijs_omg
- Huurprijs_omg

Dit geeft de tabel: **HT_MOD_FINAL_KOPPELING**

<u>Finalize & deduplicate</u>

1: Meerdere aanpassingen worden gedaan aan de tabel verkregen uit de vorige stap. Enkele variabelen worden toegevoegd, en waarden van bestaande variabelen worden aangepast indien voldaan wordt aan bepaalde criteria.

2: Voor alle variabelen wordt een missing dummy gemaakt met de prefix "M_". Alle missing waarden worden geïmputeerd. Dit kan zijn door middel van een gemiddelde of een mediaan in

1075478 00062

te voegen, maar het kan ook een andere waarschijnlijke/logische waarde zijn voor het geval hij missing is.
De imputatiewaarden staan opgeslagen op de stand alone laptop:
C:\Risico Selectie\Model\HT_MODEL\sascode\generateAdditionalVariables.sas

Dit volledige resultaat wordt opgeslagen in **HT_COMPLEET_LABELED**.

2.2.3. 3. SCORE MODEL

In deze flow wordt de volledige dataset met aanvragen en contra-informatie gescored op non_compliantie. Hierbij wordt de uitkomst van het trainen van het model in SAS Enterprise Miner gebruikt in de vorm van een SAS Base script.

Apply HT Model

1: Het model word 'gescored', i.e. alle observaties in **HT_COMPLEET_LABELED** krijgen een 'score' naar aanleiding van het getrainde scorecard model (dat gemaakt is met behulp van HT_TRAININGSET). Dit model is opgeslagen op de stand alone laptop in de vorm van een SAS Base script:

C:\Risico Selectie\Model\HT_MODEL\sascode\model_ht_ddmm_hhmm.sas

De score betreft voor alle variabelen in het model een aantal Scorecard-punten. . Bovendien geeft het model de variabele P_Y1 als output. Dit is kans (tussen 0 en 1) dat de aanvraag volgens het model 'fout' is, i.e. non compliant. Deze kans wordt in het vervolg gebruikt om de aanvragen op risico te selecteren.

2: Het resultaat uit voorgaande stap, de tabel **score**, wordt aflopend gesorteerd op de waarden van P_Y1, i.e. de eerste observatie heeft de grootste kans op non-compliantie en de laatste observatie de kleinste. Hieraan wordt het veld MODEL_HT_RANK toegevoegd. Dit geeft aan op welke plaats de persoon staat wanneer gesorteerd van 'fout' naar 'goed'.

3: Aan de tabel **score** worden twee velden gekoppeld. Dit zijn de omschrijvingen van nationaliteit voor de eerste en eventueel tweede nationaliteit van de toeslagaanvrager: av_NATIONALITEIT1_bes en av_NATIONALITEIT2_bes. Deze gegevens komen uit het halfproduct BVR en worden gekoppeld op nationaliteit code / nummer.

Het resultaat heet ook weer **score**.

2.2.4. 4. FINALIZE

1075478 00062

In deze flow wordt de scorebepaling door het model verder bewerkt tot de verschillende lijstjes die de uiteindelijke output van de risicoselectie vormen. De processen hieronder worden van links naar rechts beschreven.

2_Uitsluitlijsten

1: De twee (of drie wanneer er twee toeslagjaren worden gedraaid) meest recente uitsluitlijsten uit de brondata UITSLUITLIJST worden ingelezen. Uit deze ingelezen data worden de observaties geselecteerd die een regeling hebben gelijk aan 'HUUR'.

2: De door de risicoselectie gegenereerde uitsluitlijsten van de vorige runs worden ingelezen. Ook hier worden de observaties geselecteerd die een regeling hebben gelijk aan 'HUUR'.

3: De tabellen gegenereerd in de twee bovenstaande stappen worden samengevoegd in één tabel **UITSLUIT**. Hierbij worden drie velden geselecteerd: BSN, Toeslagjaar, en regeling. Daarnaast wordt de tabel gesorteerd op BSN en Toeslagjaar. BSN's die op dit lijstje staan mogen voor het betreffende toeslagjaar en regeling niet nogmaals uitgesloten worden, om eventuele acties/beschikkingen nav een herstelactie of overleggen van stukken niet tegen te houden.

Overledenen (Query builder task)

De tabel **Overleden** wordt gecreëerd. Deze tabel bevat de BSNs van alle overledenen uit de tabel **score** (overleden=1).

Filter and Sort6 (Query builder task)

De tabel **GroeneVlag** wordt gecreëerd. Deze tabel bevat alle BSNs uit de tabel **score** die tijdens de risicoselectie niet uitgesloten mogen worden van beschikken (GroeneVlag=1).

1_VIPS

De brondata privacy_doelgroep wordt ingelezen en gesorteerd op BSN. Vervolgens wordt de informatie opgeslagen in de tabel **VIP**.

Niet uitsluiten (Query builder task)

De tabel **NIET_UITSLUITEN** wordt aangemaakt. Deze tabel bevat alle BSNs van toeslagaanvragers die tijdens de risicoselectie niet uitgesloten mogen worden van beschikken. Deze lijst wordt samengesteld door middel van een Full Join van meerdere tabellen op BSN:

- UITSLUIT
- Overleden
- Vluchtelingenwerk (halfproduct)
- Aangeleverd (halfproduct)
- Bezwaren_filter_1 (halfproduct)
- GroeneVlag
- IGS (halfproduct)

BSNs worden alleen toegevoegd aan de tabel als het toeslagjaar onbekend is of (indien bekend) de toeslagaanvraag het juiste toeslagjaar betreft.

Geen bijvangst (Query builder task)

De tabel **GEEN_BIJVANGST** wordt gecreëerd. Deze tabel bevat alle BSNs van toeslagaanvragers die tijdens de risicoselectie niet onder de bijvangst (harde stoppers) mogen worden geplaatst. Deze lijst wordt samengesteld door middel van een Full Join van meerdere tabellen op BSN:

- Overleden
- Aangeleverd (halfproduct)
- Bezwaren_filter_1 (halfproduct)

Geen onbekende aanvrager (Query builder task)

De tabel **GEEN_ONB_AANV** wordt aangemaakt. Deze tabel bevat alle BSNs van toeslagaanvragers die tijdens de risicoselectie niet tot de onbekende aanvragers mogen worden gerekend. Deze lijst wordt samengesteld door middel van een Full Join van meerdere tabellen op BSN:

- Vluchtelingenwerk (halfproduct)
- Aangeleverd (halfproduct)
- Bezwaren_filter_1 (halfproduct)
- Overleden
- 10.2.g (halfproduct)

3 Join

1075478 00062

Aan de tabel **score** worden één voor één een aantal lijsten met een Left Join gekoppeld op BSN. Hierbij worden een aantal nieuwe variabelen gecreëerd. Dit betreffen variabelen die geen input zijn voor het risicoselectiemodel (de risicoscore per post is hiervoor immers al berekend in 3. Score Model), maar wel invloed kunnen hebben op de uiteindelijke actie die aan de post wordt toegewezen. De volgende tabellen en halfproducten worden (op deze volgorde) gekoppeld:

- Vorige_acties_huur (halfproduct)
- Ip_lijst (halfproduct)
- VIP
- NIET_UITSLUITEN
- GEEN_ONB_AANV
- GEEN_BIJVANGST

Het resultaat, de tabel **HT_score,** bevat nu alle toeslagaanvragen en contra-informatie met daarbij per aanvraag of en op welke manier de aanvraag/aanvrager beschermd is.

4 Indelen

1: Aan **HT_score** worden twee velden toegevoegd: Voor alle toeslagaanvragen wordt bepaald wat het type aanvraag is ('mutatie-omhoog danwel nieuw'; 'mutatie-omlaag-gelijk danwel lopend') en of de aanvraag onder de bijvangst valt. Dit resulteert in de tabel **HT_Export**.

2: Uit alle nieuwe aanvragers (brondata nieuwe_aanvragers) worden de onbekende aanvragers voor huurtoeslag geselecteerd en gesorteerd op BSN. De informatie wordt opgeslagen in de tabel **onbekende_aanvrager_Huur**.

3: De tabellen aangemaakt in de vorige stappen worden aan elkaar gekoppeld om voor alle toeslagaanvragen, opgenomen in **HT_Export**, te bepalen of de aanvrager onbekend is of niet (variabele onbekende_aanvrager). Een aanvrager is onbekend indien deze is opgenomen in **onbekende_aanvrager_Huur** en het Jaarbedrag behorende bij de aanvraag groter is dan 0. Deze stap geeft de tabel HT_score_all als output.

4: De indelingslijsten (Brondata Schema_wie_op_welke_lijst_v6.1) worden ingelezen. Dit schema bepaalt op welke actielijst de aanvrager komt te staan.

5: De indelingslijsten worden door middel van een Left Join gekoppeld aan **HT_score_all** op meerdere velden. Dit resulteert in tabel **EXPORT_Indeling** waarin een nieuw veld is opgenomen: 'Mogelijke_Actie'. Dit veld geeft per aanvraag de voorlopige actie, die later in het proces mogelijk nog aangepast wordt .

1075478 00062

5 Acties

In dit proces worden de voorlopige acties (bepaald in stap 5) voor enkele posten aangepast. De definitieve acties worden opgenomen in het veld 'Actie'.

1: De beschikkingen in **EXPORT_Indeling** worden aflopend gesorteerd op Rank, i.e. de meest risicovolle aanvraag staat bovenaan.

2: De posten met de voorlopige actie (veld 'Mogelijke Actie') 'Toevoegen aan uitsluit-hoog-risicolijst' worden geselecteerd. Hiervan wordt de top *X* (variabel getal), ofwel de X meest risicovolle posten definitief op de uitsluitlijst geplaatst wegens hoog risico. Deze posten krijgen de waarde 'Toevoegen aan uitsluit-hoog-risicolijst' in het veld Actie. Voor de andere geselecteerde posten wordt de waarde in het veld 'Mogelijke_Actie' aangepast naar 'Op onderzoeks-lijst vanwege hoog risico'. Deze aanvragen komen in ieder geval niet meer op de uitsluitlijst wegens hoog risico, maar mogelijk nog wel op de onderzoekslijst.

3: De posten met voorlopige actie (veld 'Mogelijke Actie') 'Op onderzoeks-lijst vanwege hoog risico' worden geselecteerd. Hiervan wordt de top *X* (variabel getal), ofwel de *X* meest risicovolle posten, definitief op de onderzoekslijst geplaatst wegens hoog risico. Deze posten krijgen de waarde 'Op onderzoeks-lijst vanwege hoog risico' in het veld Actie. Voor de andere geselecteerde posten wordt de waarde in het Actie 'Geen enkele'. Voor deze posten zal geen actie worden uitgevoerd.

4: Voor de overige posten, niet geselecteerd in stap 2 en 3 wordt 'Actie' gelijk aan 'Mogelijke_Actie'. Ofwel, de voorlopige acties worden definitief gemaakt.

Aanvragen met contra-informatie en uit te voeren acties wordt opgeslagen in het 'eindbestand **EXPORT_temp3**. Hierna worden nog enkele scripts uitgevoerd die selecties maken op deze centrale tabel, voor verdere verwerking of rapportagedoeleinden.

7 Bijvangst

1: Uit **EXPORT_temp3** worden alle toeslagaanvragen geselecteerd die onder de bijvangst (harde stoppers) vallen. Dit zijn de aanvragen die aan één van de volgende eisen voldoet:

- index(Actie,"bijvangst")
- (is_bijvangst=1 AND onbekende_aanvrager = 1);

Hieraan wordt het veld 'Reden' toegevoegd, welke de reden voor de bijvangst (uitgesloten adres of Vertrokken Onbekend Waarheen) aangeeft. Deze informatie is opgeslagen in **Bijvangst_voorbereiding.**

2: Het halfproduct RSR_teruglevering_ht wordt aan **Bijvangst_voorbereiding** gekoppeld door middel van een Left Join op BSN. Doel van deze koppeling is om voor elke toeslagaanvrager te bepalen of er in de afgelopen zes maanden al een aanvraag is behandeld en door wie. Deze informatie wordt opgenomen in het veld 'Behandeling'. Behandeling kan de volgende waarden bevatten: 'Door ons'; 'Missing'; 'Door ander'. Ook kan het veld voor een BSN leeg zijn. In dat laatste geval is er voor het desbetreffende persoon geen aanvraag behandeld in de afgelopen zes maanden.

3: De outputtabel uit de vorige stap wordt geëxporteerd onder de naam 'RISICOSELECTE_HT_Bijvangst_*&dateRun*.csv' op de volgende locatie:

C:\Risico Selectie\Model\HT_MODEL\Export

9 Onderzoekslijst

1: Op basis van de gegevens uit **EXPORT_temp3** wordt de tabel **OnderzoeksLijst** gecreëerd. Alle aanvragen met een actie betreffende de onderzoekslijst worden hieraan toegevoegd. In het verleden is deze lijst aangelegd om posten die niet direct uitgesloten werden van beschikking, maar wel volgens het model een hoog risico op niet compliantie droegen, toch te kunnen bijhouden en eventueel aan te bieden voor behandeling. Dit is echter nooit in de praktijk gebeurt.

Daarnaast wordt de extra kolom 'Reden' aangemaakt. Deze bevat de waarde 'Uitgesloten adres' of 'VOW' indien de aanvraag tot de bijvangst behoort en de waarde is 'Hoog risico' in de andere gevallen.

2: De outputtabel uit de vorige stap wordt geëxporteerd onder de naam 'RISICOSELECTE_HT_OnderzoeksLijst_*&dateRun*.csv' op de volgende locatie:

C:\Risico Selectie\Model\HT_MODEL\Export

10 Normaliseren

De scorecard punten opgenomen in **EXPORT_temp3** worden genormaliseerd met behulp van de macro 'Normaliseer' (zie hierboven). Dit resulteert in het outputbestand **Export_norm_HT**.

11 Overzichtstabel

1: Een overzichtstabel van de resultaten uit **EXPORT_temp3** wordt aangemaakt. Deze tabel geeft het aantal posten over schillende dimensies, die gebruikt werden voor

rapportagedoeleinden. De waarden in enkele velden worden aangepast en er worden nieuwe velden aangemaakt:

- In_Model_gegaan
- Geclassificeerd
- Type_Mutatie
- BSN_niet_uitsluiten
- Overleden_erbij_betrokken
- Nulbeschikking
- Deeljaar_Actueel2
- Eerdere_Actie_op_geweest
- Op_Bezwarenlijst
- VIP

2: Uit de tabel gecreëerd in de vorige stap worden een aantal velden geselecteerd. Vervolgens wordt per dimensie het totaal aantal BSNs berekend en opgenomen in het veld 'aantal_BSNs'. Dit resulteert in de outputtabel **Overzichtstabel_HT**.

3: Met de informatie in de tabel **Overzichtstabel_HT** wordt een frequentie tabel gegenereerd.

4: De overzichtstabel wordt geëxporteerd onder de naam *'Overzichtstabel Huur &dateRun.csv'* op de volgende locatie:

C:\Risico Selectie\Model\HT_MODEL\Export

12_JW

De tabel **lijst_JW_HT** wordt gecreëerd. Dit is een lijst met BSNs (en contra-informatie) die uitgesloten zijn van beschikken. **EXPORT_temp3** wordt gekoppeld aan de lijst met uitgesloten beschikkingen **Uitgesloten_HT** door middel van een Left Join op BSN. Hier worden een aantal variabelen geselecteerd en worden nieuwe variabelen gecreëerd:

- Regeling (waarde is altijd 'HT')
- Uitgesloten (waarde is altijd 'J')
- Batch (waarde 'mu' of 'ea')
- Herkomst (waarde 'rsr' of 'FRM')

Deze lijst vormt de invoer voor een proces (onderhouden door Toeslagen collega [redacted] dat rapporteert over de uiteindelijke afhandeling van posten die vanwege de risicoselectierun zijn uitgesloten van beschikken.

2.2.5. 7. RAPPORTAGE

In dit script worden de kengetallen gegenereerd die worden gebruikt als input voor het periodieke Management Rapportage betreffende de resultaten van de risicoselectie.

Libname

Vaststellen van libraries, locaties inputbestanden en huidige datum. Daarnaast moet voor elke run opnieuw de periode en datum van de desbetreffende risicoselectierun handmatig worden ingevoerd.

1. Behandeling

Hier wordt het aantal conceptaanvragen en lopende aanvragen berekend dat ter behandeling is aangeboden. Daarnaast wordt bepaald hoeveel conceptaanvragen op de uitsluitlijst staan. Beide berekeningen maken gebruik van het bestand export_temp3.

Aangezien op dit moment alleen nog de conceptbeschikkingen worden aangeboden, is het aantal aangeboden posten gelijk aan het aantal posten op de uitsluitlijst.

Een post komt op de uitsluitlijst indien de variabele Actie de waarde 'Toevoegen aan uitsluit-hoog-risicolijst' bevat.

2a. Steekproef

In deze flow wordt berekend hoe groot de geselecteerde populatie toeslagaanvragen is. Uit deze populatie wordt een steekproef getrokken om te controleren of de selectie van aanvragen terecht was. Het bepalen van de populatiegrootte betreft dezelfde berekening als vermeld in bovenstaande flow: 1. Behandeling. De populatie geselecteerde posten zijn de aanvragen welke de actie 'toevoegen aan uitsluit-hoog-risico-lijst' hebben.

De grootte van de steekproef wordt tijdens de risicoselectierun echter hardcoded bepaald en kan niet automatisch d.m.v. een SAS-script in het MT Rapportage worden toegevoegd. De toevoeging van dit kengetal aan het rapportage is een handmatige handeling.

2b. Importeer terugkoppeling steekproef rs.xlsx

Hier wordt het bestand Terugkoppeling steekproef rs.xlsx geïmporteerd. Het percentage posten uit de steekproef terecht geselecteerd als hoog risico, heeft betrekking op de posten getrokken uit de populatie zoals beschreven in bovenstaande flow: 2a. Steekproef

2c. Voeg resultaten steekproef samen

In deze flow worden de resultaten uit de flows 2a. en 2b. samengevoegd in één bestand.

3. Harde Stoppers

In deze flow wordt het aantal harde stoppers van de posten in het bestand export_temp3 berekend. Voor huurtoeslag zijn er twee harde stoppers voor beschikken:

1. Het adres waarvoor huurtoeslag is aangevraagd, is uitgesloten van beschikken.
2. De toeslagaanvrager is vertrokken, maar het is onbekend waarheen (VOW).

Een post valt onder de bovenstaande harde stoppers indien:

1. De variabele Actie de waarde 'Toevoegen aan uitsluit-bijvangstlijst' bevat EN de waarde van variabele av_ht_uitgesloten adres gelijk is aan 1 (harde stopper 1)
2. De variabele Actie de waarde 'Toevoegen aan uitsluit-bijvangstlijst' bevat EN de waarde van variabele B_KEYADRES1 leeg is of gelijk aan '&&' (harde stopper 2)

4. Onbekende aanvrager

In deze flow wordt het aantal onbekende aanvragers voor de huidige risicoselectierun vastgesteld. Ook hier is export_temp3 het gebruikte bronbestand. Een toeslagaanvrager is onbekend indien het in export_temp3 de actie 'Toevoegen aan uitsluitlijst onbekende aanvrager' bevat en de variabele onbekende_aanvrager gelijk is aan 1.

5. Hitrate

Hier wordt voor elk van de vijf afgelopen risicoselectieruns berekend hoeveel procent van de aangeboden posten is behandeld, wat de behaalde hitrates zijn en de bijbehorende opbrengsten voor het toeslagjaar. Deze berekeningen worden gedaan obv de gegevens in de bestanden rsr_teruglevering.sas7bdat en historisch_overzicht_HT.sas7bdat :

1: Sorteer het bestand historisch_overzicht_HT.sas7bdat op datum van de risicoselectie en neem de laatste vijf voorgaande perioden.

2: Bepaal voor elke periode:

a. alle huurtoeslagaanvragen met actie wegens hoog risico. Dit zijn alle aanvragen uit rsr_teruglevering.sas7bdat waarvoor geldt: toeslagtype=2 EN Hoofdbatch = 'Risico'. Bepaal vervolgens alle voltooide acties: WO_status='Completed'. Tel alle voltooide acties bij elkaar op en deel door het totaal aantal huurtoeslagaanvragen aangeboden obv hoog risico. Dit geeft het percentage behandelde posten.
b. alle behandelde toeslagaanvragen met als resultaat een lagere of nul-beschikking. Dit zijn de aanvragen waarbij: Resultaat = nul OF Resultaat = lager. Tel het aantal posten waarvoor dit geldt en deel dit door het totaal aantal huurtoeslagaanvragen aangeboden obv hoog risico. Dit getal geeft de hitrate.
c. De jaaropbrengst voor de behandelde aanvragen met lager of nul-beschikking. Dit is gegeven door het attribuut HHR_bedrag. Tel alle waarden bij elkaar op. Dit geeft de totale opbrengst jaartoeslag voor de desbetreffende periode.

3: Creëer een bestand dat voor alle vijf perioden het percentage behandelde posten, de behaalde hitrate en opbrengst bevat.

6. Risicoscores

In deze flow wordt voor de huidige en voorgaande risicoselectieperiode de verdeling toeslagaanvragen per categorie risicoscores bepaald. Deze gegevens worden gehaald uit de bestand export_temp3 en historisch_overzicht_HT.sas7bdat.

1: Bepaal voor elke aanvraag de categorie risicoscores waartoe het behoort. Voeg een nieuw attribuut toe aan de dataset met als waarde de categorie.

2: Groepeer de aanvragen per categorie risicoscores en tel de verdeling posten per categorie in aantallen en percentages.

3: Voeg zowel de verdeling in aantallen als in percentages toe aan het bestand historisch_overzicht_HT.sas7bdat.

4: Sorteer en ontdubbel het bestand historisch_overzicht_HT.sas7bdat zodat voor elke runperiode alleen de gegevens van de laatste verversdatum behouden blijven.

5: Sorteer het bestand historisch_overzicht_HT.sas7bdat op datum van de risicoselectie en selecteer de risicoscoreverdelingen van de laatste twee perioden.

7. Aanvragen onbehandeld

Hier wordt de verdeling nieuwe toeslagaanvragen/mutaties omhoog en lopende aanvragen/mutaties omlaag-gelijk per categorie risicoscores bepaald. Daarnaast wordt voor elke categorie berekend hoeveel van deze aanvragen dit toeslagjaar nog niet behandeld zijn / worden.

Deze berekeningen worden gedaan obv de gegevens in de bestanden export_temp3.sas7bdat en rsr_teruglevering.sas7bdat :

1: Bepaal voor alle aanvragen uit export_temp3.sas7bdat tot welk type ze behoren. Een aanvraag behoort tot nieuwe aanvragen/mutaties omhoog indien de variabele TYPE_aanvraag de waarde 'mutatie-omhoog danwel nieuw' heeft. Indien de waarde gelijk is aan 'mutatie-omlaag-gelijk danwel lopend', is er sprake van een lopende aanvraag of een mutatie omlaag/gelijk.

2: Bereken voor elk van de twee verzamelingen toeslagaanvragen:

a. tot welke categorie risicoscores de aanvragen behoren.
b. De verdeling posten in aantallen per categorie risicoscores
c. Per categorie risicoscores : Het aantal posten dat deze periode niet behandeld is/wordt. Dit zijn de toeslagaanvragen die niet op de uitsluitlijst staan.
d. Per categorie risicoscores: Welke aanvragen gedurende het huidig toeslagjaar nog niet behandeld zijn. Deze berekening wordt gedaan op BSN niveau door voor alle onbehandelde aanvragen deze periode het BSN van de toeslagaanvrager te koppelen met de BSNs in het bestand rsr_teruglevering. Indien het BSN voor het desbetreffende toeslagjaar niet in rsr_teruglevering voorkomt, is het een onbehandelde aanvraag voor het huidige toeslagjaar.
e. Per categorie risicoscores: Het aantal onbehandelde posten in het huidige toeslagjaar.

8. Voeg samen

Hier worden alle resultaten uit de beschreven flows samengevoegd tot één output-bestand. Dit bestand bevat:

1. Huidige datum
2. Runperiode
3. Aantal aanvragen ter behandeling aangeboden
4. Aantal aanvragen toegevoegd aan de uitsluitlijst
5. Aantal posten in populatie geselecteerd voor steekproef
6. Aantal posten getrokken in steekproef
7. Percentage van steekproef terecht geselecteerd als hoog risico
8. Aantal harde stoppers per categorie
9. Aantal onbekende aanvragers
10. Percentage behandelde aanvragen, hitrate en opbrengst vijf voorgaande perioden
11. Per categorie risicoscores: Het aantal en percentage posten voor de huidige en vorige periode
12. Per categorie risicoscores: Het aantal nieuwe aanvragen/mutaties omhoog en lopende aanvragen/mutaties omlaag-gelijk voor de huidige periode en hiervan het aantal aanvragen onbehandeld in het huidige toeslagjaar.

9. Export

Het output-bestand gecreëerd in de vorige flow wordt hier geëxporteerd in een excel-bestand naar de C-schijf.

2.3. CHECKLIST MANAGEMENTRAPPORTAGE

Deze checklist is een lijst welke doorgelopen dient te worden vóór het runnen van de procesflow '7. Rapportage'.

Voor het runnen van de managementrapportage wordt aangenomen dat:

1. Er een juist en volledig eindbestand (Export_temp3) van de risicoselectie gemaakt is. De rapportage flow moet dus altijd aan het einde van de risicoselectierun gedraaid worden.
2. De parameters in het script '1. Init' zijn juist gezet. Let op: hier is vanwege de managementrapportage een extra variabele (datum_risicoselectie) bij gekomen.
3. De juiste versie van het bestand 'rsr_teruglevering' staat in de folder *C:\Risico Selectie\Halfproducten\control_klantbehandeling*

Om de managementrapportage te runnen moet het volgende gedaan worden:

1075478 00062

1. Verkrijg het bestand 'terugkoppeling steekproef risicoselectie'. Dit excel-bestand bevat resultaten uit de steekproef getrokken na de proefrun, één tot enkele dagen vóór de risicoselectierun. Deze steekproef wordt getrokken uit de top X hoog risico-aanvragen in een voorlopige set aan conceptbeschikkingen. Het excel-bestand met kengetallen wordt elke risicoselectierun opnieuw aangeleverd vanuit Handhavingsregie. Indien mogelijk wordt deze vóór de start van de run op het netwerk geplaatst:

   Q:\VEPROW63\TSL_DM_Handhavingsregie\RISICOSELECTIE\Terugkoppeling steekproef risicoselectie

   Vervolgens dient het risicoselectieteam dit bestand te kopiëren naar de stand alone laptop:

   C:\Risico Selectie\Brondata\Terugkoppeling steekproef\Source

2. Zet in het script '1. Init' van het SAS Enterprise Guide project 'HT_Model' de datum van de risicoselectie (variabele datum_risicoselectie) naar de datum van de betreffende risicoselectierun.

3. Run de process flow '7 Rapportage.' in het SAS Enterprise Guide project 'HT_Model'.

4. Deze process flow schrijft alle output naar 1 .csv bestand:

   C:\Risico Selectie\Model\HT_Model\Export\MT_Rapportage_data_HT.xlsx.

   De gegevens uit deze file moet handmatig gekopieerd worden naar de tab 'Data HT' in de file *'C:\Risico Selectie\Rapportage\MT_Rapportage.xlsx'* . Let er hierbij goed op dat de gegevens naar exact dezelfde cellen worden gekopieerd (selecteer cellen, kopieer, en plak op tab 'Data HT' in cel A1.). Bij deze actie mogen er geen gegevens uit voorgaande risicoselectieruns in de tab achtergebleven zijn.

5. Dit zorgt ervoor dat automatisch de juiste getallen getoond worden in de tab 'Resultaten HT' in het rapport. Doe hier echter voor de zekerheid nog een sanity check op!

6. Er is 1 informatie-element dat handmatig toegevoegd moet worden: cellen D120 en D121 die gaan over de laatste datum waarop het risicoselectiemodel opnieuw getraind is en het aantal trainingscases dat sinds de voorgaande training van het model is toegevoegd.

Wanneer ook de informatie voor KOT en Zorg juist is opgenomen in het rapport (zie hiervoor de documentatie van 'KOT_Model' en het halfproduct 'Uitsluiten_Onbekende_Aanvragers_Zorg') en de volledige controlematrix voor de risicoselectie is afgerond, kan het rapport toegestuurd worden aan de regisseur van de risicoselectie, die het rapport uiteindelijk deelt met het management.

1075478 00062