

Bespreking voortgang en vervolg risicoclassificatie Toeslagen

Datum: Woensdag 3 juni 2013

Aanwezig:

(Deloitte)

(Deloitte)

(Deloitte)

De volgende "run" voor de risicoclassificatie voor Kinderopvangtoeslag en Huurtoeslag (hierna: KOT en HT) is wederom gekoppeld aan het moment voorafgaand aan formeel beschikken. Dit is voor aankomende maand vastgesteld op 2 augustus 2013.

Naar aanleiding van de aankomende "run" voor risicoclassificatie van de toeslagen voor Kinderopvang en Huur is worden in onderhavige bijeenkomst de vervolgstappen, werkzaamheden, werkverdeling en aanvullende punten besproken.

1. Terugkoppeling gesprek BCA van op 2 juni 2013.
2. Aanwezigheid en werkzaamheden Deloitte.
3. Wishlist aanpassingen modellen voor volgende run.
4. Aanpassingen in de serverindeling.
5. Werkzaamheden en verdeling van ownership over modellen.
 - a) Verdeling ownership modellen
 - b) Updaten brondata en halfproducten
 - c) Trainingsdossiers
 - d) QA controles
 - e) Aanpassingen aan model/filters.
6. Overige punten.
 - a) TVS conversie
 - b) Tijdlijn tot 2 augustus

1. Terugkoppeling gesprek BCA van op 2 juni 2013.

Op 2 juni 2013 hebben gesproken met medewerkers van BCA. Hieruit is voortgekomen dat gepoogd wordt om voor de volgende "run" de halfproducten BVR en FRS meldingen geheel door BCA te laten updaten en aanleveren. BCA levert de halfproducten dan op in een SAS formaat. Wel moet hier nog controle op plaatsvinden voor het halfproduct gekoppeld wordt.

Mogelijk kunnen op termijn meerdere halfproducten (conceptbeschikkingen en ouder-kind) worden opgeleverd door BCA.

2. Aanwezigheid en werkzaamheden Deloitte

Als onderdeel van de brede opdracht tot implementatie van de risicoclassificatiemodellen voor KOT en HT zal door de medewerkers van Deloitte een meer ondersteunende rol worden ingenomen voor de volgende run. Ook zal hiermee de aanwezigheid van Deloitte medewerkers niet meer op een full-time basis in worden gevuld.

Afgesproken wordt dat op dinsdag, woensdag en donderdag aanwezig zal zijn op het kantoor van de Belastingdienst en dat beiden op de woensdag en donderdag aanwezig zullen zijn.

Het zwaartepunt van de werkzaamheden komt hiermee te liggen bij zij zullen door worden ondersteund in de werkzaamheden.

Tevens zijn er een tweetal "trainingsessies" ingepland. De eerste zal plaatsvinden op 24 juni 2013. Deze sessie zal gaan over databewerking in SAS en datavoorbereiding voor het model.

De tweede sessie zal plaatsvinden op 14 augustus 2013 en zal gaan over predictive modelling en de werking van het model.

Tussentijds zal door Deloitte met kennis en kunde ondersteuning worden gegeven bij de werkzaamheden.

3. Wishlist aanpassingen modellen voor de volgende run

Er worden een aantal aanpassingen binnen het model besproken welke voor 2 augustus verwerkt zouden moeten zijn. Het gaat om de volgende aanpassingen:

- Opnemen van de lijst met IP-adressen van aanvragers in de modellen. Tevens is er de wens om IP-adressen als indicator te gebruiken. [] beschikken over lijsten met BSN's en IP-adressen welke "bewezen fout" en "bewezen goed" zijn. Deze lijst wordt opgevraagd.
- Opnemen van de datum van het deeljaar in het huurtoeslag model. Hiermee kan gekeken worden naar de datum van het deeljaar versus de datum van bewoning om zo "false positives" veroorzaakt door een beschikking die niet aansluit op de bewoningsdatum te voorkomen.
- Opnemen van de leeftijd van de aanvrager versus de leeftijd van het oudste kind. Dit kan mogelijk als indicator gebruikt worden in gevallen waarbij de leeftijd van de aanvrager onrealistisch laag is ten opzichte van de leeftijd van het oudste kind waar toeslag voor wordt aangevraagd.

Daarnaast is op 15 juli 2013 een sessie ingepland met medewerkers van BI [] om alle indicatoren voor zowel de KOT als de HT opnieuw te evalueren, eventuele aanvullende indicatoren te bedenken en ook te spreken over de lay-out van de resultatenlijst van de profiling.

4. Aanpassingen in de serverindeling

Er wordt gesproken over de indeling van de map structuur van de server en het gebruik van de server ruimte.

Op dit moment staan alle gegevens en data met betrekking tot de risicoclassificatie op een locatie op de VEPROW52 schijf.

Deze schijfruimte is oorspronkelijk bedoeld als data uitwisseling locatie tussen BCA en andere afdelingen. De afspraak is om elke maand deze ruimte op te schonen [] geeft aan dat deze schijf door meerdere personen te benaderen is en dat onduidelijk is wie nu precies toegang heeft tot deze schijfruimte.

Omdat op deze wijze de veiligheid van de data en beveiliging van de gegevens en informatie niet gegarandeerd kan worden, wordt besloten om op korte termijn alle data vanuit de "Profiling 2013" map te verplaatsen naar de map Handhaving (VEPROW 63). Dit zal gebeuren wanneer iedereen toegang heeft tot deze locatie.

Vervolgens wordt gesproken over de indeling van de serverruimte. De map structuur is ingedeeld naar bron data, halfproducten en run-maand. Op dit moment is er sprake van serververvuiling. Het is belangrijk de bestanden in de mappen ook daadwerkelijk gescheiden te houden voor overzicht. Daarom wordt besloten om voor de transitie naar de VEPROW 63 locatie de mappen geschoond worden:

- Alle "results" worden, voor zover aanwezig, verwijderd uit de halfproduct mappen en uit de bron data mappen. Results mogen alleen in de specifieke run-maand map geplaatst worden en dan alleen voor de resultaten van het gehele model;
- Alle "source" worden, voor zover aanwezig, verwijderd uit de halfproduct mappen. Source data is alleen te vinden in de bron data mappen.

Daarnaast wordt de mogelijkheid bekeken om gebruik te maken op korte termijn van een aanwezige server binnen BI. Hierop kan mogelijk een nieuwe SAS versie van enterprise guide en enterprise miner worden geïnstalleerd om te gebruiken voor de risicoclassificaties. Deze mogelijkheid is ook besproken met [] zal eea bespreken met medewerkers binnen toeslagen om te zoeken naar een oplossing. Voorlopig is er vanuit SAS de toezegging om voor 2 augustus 2013 gebruik te kunnen maken van een SAS laptop met enterprise guide en enterprise miner.

5. Werkzaamheden en verdeling van ownership over de modellen.

De onderdelen van de beiden modellen worden systematisch behandeld.

a. Verdeling ownership over de modellen.

Zowel het model voor HT als voor KOT zal in de aanloop naar 2 augustus 2013 worde geüpdatet en mogelijk worden aangepast (indicatoren). De werkzaamheden die hiermee gepaard gaan worden verdeeld tussen [] met ondersteuning door [] om hen zo veel mogelijk training en ervaring met het model te laten op doen.

Besloten wordt om de modellen te verdelen tussen [] en een ondersteuner vanuit Deloitte aan te wijzen.

[] krijgt hierbij de ownership over het volledige HT model, bijbehorende halfproducten, indicatoren en bron data. [] hierbij ondersteunen en begeleiden daar waar nodig.
 [] krijgt hierbij de ownership over het volledige KOT model, bijbehorende halfproducten, indicatoren en bron data. [] hierbij ondersteunen en begeleiden daar waar nodig.

b. Updaten bron data en halproducten

Bron data

De bron data voor beiden modellen moet worden geüpdatet en geplaatst in de daartoe betreffende map op de VEPROW locatie. Er wordt besloten elke maand te bekijken welke data geüpdatet moet worden.

Onderdeel van het updaten wordt deze maand ook het documenteren van de bron data. Hierbij moet het volgende worden opgenomen:

- Voor welk halfproduct wordt de bron data opgevraagd?
- Bij wie wordt de bron data opgevraagd (loket of persoon) ?
- Hoe wordt de bron data opgevraagd (query)
- Wat is de gemiddelde doorlooptijd van de oplevering (snel te leveren of niet?)
- Hoe wordt de bron data opgeleverd?
- Waar wordt op de VEPROW schijf de data opgeslagen?
- Wat moet er vervolgens met de bron data gebeuren?

Sommige bron data wordt gebruikt voor meerdere halfproducten c.q. modellen. Ook in dit geval wordt de data door slechts 1 persoon opgevraagd.

De ownership van de bron data wordt als volgt verdeeld:

Brondata	Eigenaar
6PP	
BAG	
Beschikking KOT	
Beschikking Huur	
Blacklist	
BVR 07	
BVR 09	
BVR Bewoning	
BVR BSN Leeftijd	
BVR Nationaliteit	
BVR NAW	
HT Labels	
KOT conceptbeschikkingen	
Huur conceptbeschikkingen	
Meldingen FRS	
Ouder_Kind	
SVB	
VHA	
WOZ	
VIT	
COA	

De WOZ bron data die afgelopen maand is opgeleverd moet integraal worden gecontroleerd. Hiervan moet de query worden opgevraagd bij [] en de data worden onderzocht op herkomst en veldwaarden. [] zal dit samen met [] oppakken.

Halfproducten

De halfproducten moeten na het updaten van de bron data elke maand opnieuw worden gedraaid en getoetst (geldt alleen voor de halfproducten waarvan de bron data wordt geüpdatet). Verder moeten de halfproducten worden gecontroleerd voor de koppeling met de indicatoren die worden gebruikt in het eindmodel.

Besloten wordt om bij de verdeling werkzaamheden ten aanzien van de halfproducten deze te relateren aan het ownership van het eindmodel waar het halfproduct betrekking op heeft.

Onderstaand volgt de verdeling, inclusief de aanvullende werkzaamheden voor 2 augustus (nog zonder indicatoren aanpassingen)

Halfproduct	Eigenaar	Werkzaamheden
6PP BAG BVR COA DTM HT DTM KOT HT Model KOT Model IP-lijst MELDING8 MELDINGFRS OUDER_KIND PIT SVB UITGESLOTEN_ADRES UITVAL VHA WOZ		BCA vragen om index op BSN en 01-adres van stream een separate file maken voor model Integrale controle op query voor KOT Integrale controle op velden en waarden

Op de VEPROW locatie komt een bestand met daarin de checklist voor afronding en controle van zowel de halfproducten als de bron data. Op de halfproducten dient een 2^e controle door een ander persoon plaats te vinden voor de run.

c. Trainingdossiers

Voor 2 augustus zullen beiden modellen (KOT en HT) opnieuw getraind moeten worden. Om dit te kunnen doen moeten er nieuwe trainingdossiers beschikbaar komen.

Idealiter komen voor training de resultaten van de risicoclassificaties van april 2013 (HT) en juni 2013 (KOT) beschikbaar.

Er is een interface opgesteld voor medewerkers/behandelaars om de bevindingen ten aanzien van de gescoorde beschikkingen in op te nemen. Zodoende kunnen de dossiers worden terug gevoed aan het model ter training.

Op 25 juni 2013 is hierover ook gesproken met De verwachting is dat op of rond 20 juli 2013 de eerste aantallen BSN's die in juni zijn uitgenodigd voor een baliebezoek voor de KOT beschikbaar komen. Hiermee kan getraind worden. Omstreeks dit moment is ook het aantal baliebezoeken en no-shows voor de HT bekend. Ook hiermee zal dan getraind kunnen worden.

De aantallen vanuit de voorgaande runs zijn echter niet voldoende voor de training van de modellen. Daarom zal gekeken gaan worden naar alternatieven voor trainingdossiers. [] zal dit oppakken. Een van de ideeën is om wederom random beschikkingen te scoren en deze te gebruiken als training (100% goed en 100% fout). Daarnaast zal een inventarisatie worden gedaan bij:

- Fraudeteam, met het voorbehoud van objectiviteit
- Klachtenlijn, voor onterecht gescoorde BSN's
- Een lijst afkomstig van een medewerker bevattende fraude op toeslaggebieden.

Er zal tevens worden uitgedragen dat onvoldoende dossiers voor training zijn weerslag zal hebben op de scoring.

Aan [] zal gevraagd worden of er een mogelijkheid is om structureel steekproeven te laten uitvoeren op tussentijdse scores om deze zo weer terug te voeden aan het model. Ook zal de steekproeven gedaan op vrijdag 22 juni 2013 gebruikt worden voor training.

d. QA controles

In aanloop naar 2 augustus zal een "check list" opgesteld worden wat er allemaal gedaan moet worden aan checks en controles op het model en de resultaten voordat een definitieve resultatenlijst opgeleverd kan worden aan het MT ter besluitvorming.

[] zullen een eerste aanzet maken voor een checklist.

Daarnaast is het ook van belang om tussentijds alle koppelingen, velden, veldwaarden, kolommen en resultaten van de halfproducten te controleren. Hetzelfde geldt voor een snelle controle op logica in distributie van de modelscore, dit om te controleren of het model correct scoort (de juiste indicatoren gebruikt, geen onwaarschijnlijk hoge scores, etc.)

e. Aanpassingen aan het model/filters

Op dit moment worden de resultaten van de risicoclassificatie voor KOT en HT nog in de resultatenlijst in Excel gefilterd (bijvoorbeeld op overledene en op 0-beschikkingen). Het voornemen is om in de run van 2 augustus meer te filteren direct in het SAS bestand. Op deze manier blijft zichtbaar waar precies op is gefilterd en wordt er gecontroleerd gefilterd. Op 15 juli 2013 zal, in de bespreking met betrekking tot de gewenste indicatoren voor KOT en HT, ook gesproken worden over de lay-out van de resultatenlijst.

Hierbij zal uit worden gegaan van het volgende:

- Lay-out van de resultatenlijst moet aansluiten bij behoefte in de workflow;
- Alleen de noodzakelijke gegevens worden gepresenteerd in de resultatenlijst;
- Welke gegevens noodzakelijk zijn voor behandeling van de gescoorde beschikkingen wordt in overleg met experts op of rond 15 juli 2013 vastgesteld;
- De filtering van de gegevens gaat zo veel mogelijk in SAS gebeuren en niet meer in Excel;
- De filtering gebeurt zo veel mogelijk uniform voor KOT en HT (in elk geval voor overledenen en 0-beschikkingen).

De modellen voor KOT en HT dienen tevens volledig (zonder filters) gekopieerd te worden naar de server.

6. Overige punten

a. TVS conversie

In het weekend van 29 en 30 juni heeft een conversie plaatsgevonden in TVS. De conversie betrof het toevoegen van LRK nummers voor de KOT. Het is mogelijk dat de conversie gevolgen heeft voor de data die gebruikt wordt voor de risicoclassificatie voor KOT (bijvoorbeeld ten aanzien van concept beschikkingen). OP dit moment is echter onduidelijk wat het preciese effect is geweest van de conversie op de data in TVS en wat de mogelijke gevolgen voor het KOT model en risicoclassificatie is.

Dit wordt zo snel mogelijk uitgezocht door []

b. Tijdlijn

De volgende pagina geeft de tijdlijn weer tot en met 2 augustus 2013.

TIJDLIJN

