



MINISTERUL MUNCII  
ȘI SOLIDARITĂȚII SOCIALE

Agencia Națională pentru Plăți și Inspecție Socială

DIRECȚIA AUDIT PUBLIC INTERN

## **Determinarea unor profile în auditul administrării beneficiilor de asistență socială cu utilizarea unor tehnici de inteligență artificială**

În perioada 21.03.2022- 30.06.2022, la nivelul DAPI-ANPIS s-au desfășurat două misiuni de audit public intern de regularitate respectiv: **„Auditarea activității de stabilire, acordare și punere în plată a alocației pentru susținerea familiei”** și **„Evaluarea activității de acordare și plată a alocației de stat pentru copii”**. Pe baza datelor și informațiilor colectate pe parcursul misiunii de audit și în special a datelor cuprinse în Balanța intrări-ieșiri (DLI) pentru perioada 2012-2021, în cadrul misiunii au fost realizate analize de date în vederea identificării unor trenduri și/sau excepții, pe baza unor modele de analiză statistică avansate prin utilizarea programului informatic „R” (The R Project for Statistical Computing).

Prin valorificarea unor procese de analiză și tendințe, le-a fost facilitată auditorilor cunoașterea domeniului specific, care este necesară pentru dezvoltarea raționamentului profesional<sup>1</sup>.

Activitățile referitoare la beneficiile de asistență socială au la bază o serie de procese ce apar și se dezvoltă, în principal, ca urmare a unor diverși factori care acționează în aceeași direcție sau în direcții opuse și cu o amploare diferită. Procesele sau fenomenele sunt interconectate între ele prin legături diverse, care nu sunt remarcate și cunoscute de prima dată. Scăderea riscului de nedectare și implicit de

---

<sup>1</sup> Conform Manualului privind auditul de conformitate al Curții de Conturi, utilizarea raționamentului profesional în toate etapele acțiunii de control se referă la aplicarea cunoștințelor relevante dobândite prin pregătire și experiență profesională, pentru luarea de decizii potrivite cu privire la cursul acțiunii, în funcție de circumstanțele în care se desfășoară auditul. Conceptul de scepticism profesional este fundamental pentru toate tipurile de acțiuni de verificare. Auditorul public trebuie să planifice și să realizeze controlul cu o atitudine de scepticism profesional, recunoscând că anumite circumstanțe pot cauza discrepante între obiectivul specific analizat și criteriile de audit identificate. O atitudine de scepticism profesional înseamnă că, auditorul public realizează o evaluare critică, cu o atitudine alertă și iscoditoare a gradului de suficiență și de adecvare a probelor de audit abținute pe parcursul controlului.

audit implică utilizarea unor metode moderne precum data mining și inteligența artificială . Institutul Internațional al auditorilor Interni (IIA ) cuprinde de altfel astfel de direcții de dezvoltare .<sup>2</sup>

Pe de alta parte, calitatea raționamentului profesional presupune cunoașterea de către auditori a unor profile de administrare. Determinarea unor profile poate avea la bază o analiză cantitativă nesupervizată ( pe baza inteligenței artificiale exclusiv ) , care să cuprindă elemente specifice proceselor de administrare a beneficiilor de asistență socială.

În cadrul Direcției Audit Public Intern din cadrul ANPIS aceasta se realizează prin parcurgerea de către auditori a unor studii cantitative<sup>3</sup> ce se desfășoară în etapa conceperii studiului preliminar. Studiile au un caracter formalizat .

Analiza statistică a raporturilor de dependență dintre procese și fenomene pune accent pe determinarea relației dintre două sau mai multe caracteristici. Importanța percepției legăturii dintre fenomene sau procese, creează posibilitatea reală de control și influențare a acestuia. Relațiile de interdependență dintre procese și fenomene, îmbracă forme diverse și este destul de greu de remarcat. Analiza legăturii dintre o variabilă dependentă (rezultativă, efect, explicată) și una sau mai multe variabile independente (factoriale, cauzale, explicative) presupune metode de analiză care utilizează inclusiv tehnici de inteligență artificială .

În acest context, pentru acest studiu , am utilizat în primul rând o astfel de metodă pentru care s-au utilizat ca variabile :

- ponderea beneficiarilor reținuți la plată în numărul total de beneficiari - ca variabilă explicată;
- ponderea beneficiarilor suspendați în numărul total de beneficiari, ponderea beneficiarilor noi intrați în plată în numărul total de beneficiari și numărul total de beneficiari - ca variabile explicative.

Pe cale de consecință , studiul are la bază analiza unor indicatori derivați (definiți printr-un raport ) astfel :

- a. numărul de beneficiari reținuți / suspendați într-o lună ca urmare a identificării unor anumite elemente de eroare, raportat la numărul de beneficiari aflați în plată .
- b. numărul de intrări noi în lună, față de numărul de beneficiari aflați în plată.

Datele aferente au fost analizate pentru fiecare județ (agenție teritorială) și lună de administrare pentru perioada ianuarie 2012 – mai 2022 și au avut în vedere alocația de stat pentru copii (ASC) și alocația de sprijin familial (ASF). Față de acest set de variabile ne-am propus practic să căutăm răspunsul la întrebarea : „**De ce variabile este influențată ponderea reținerilor la plată?**”

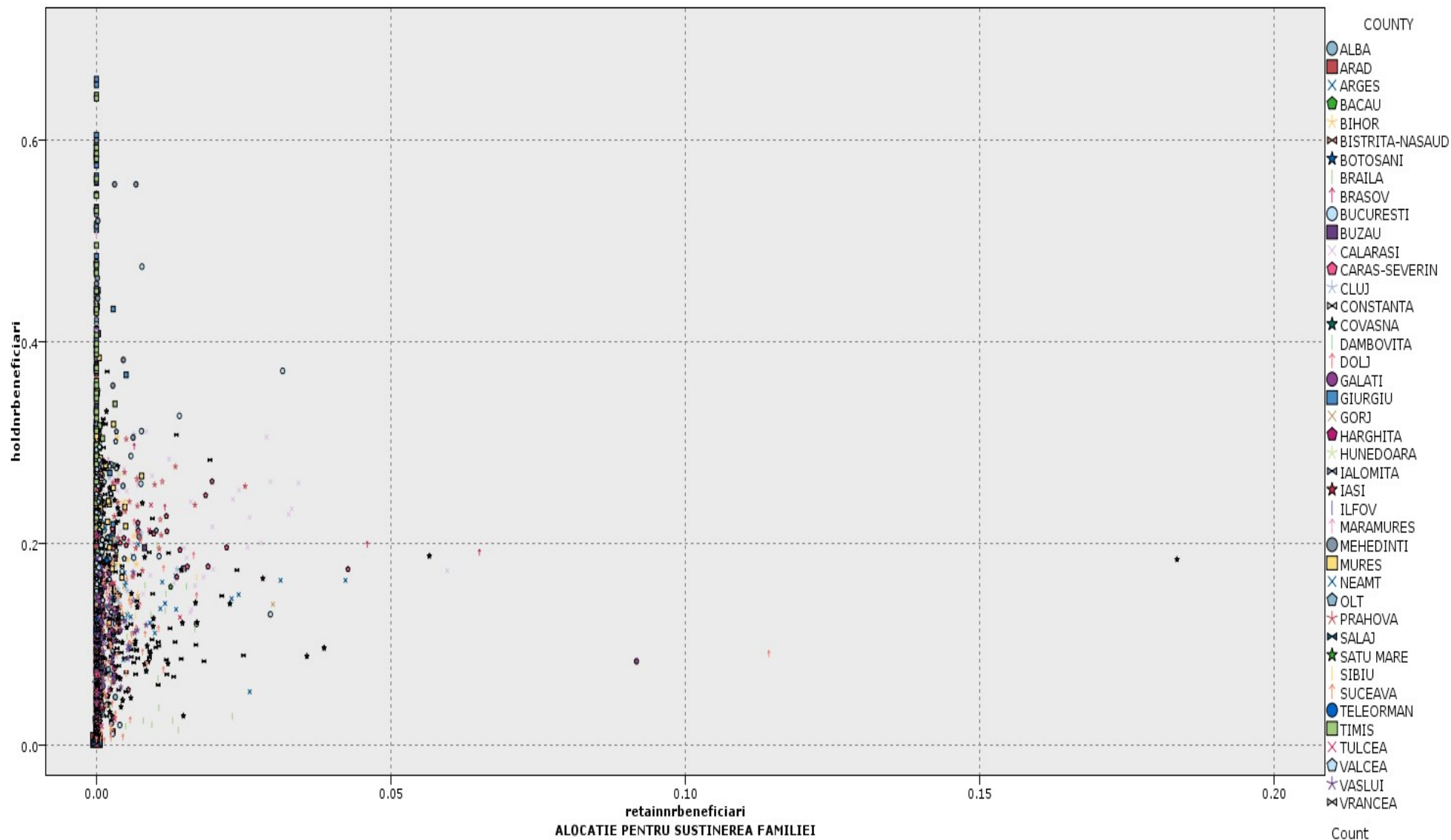
Procesele de rețineri la plată sunt realizate în principal, când starea de fapt impune o reacție rapidă privind neefectuarea unei plăți. Ipoteza considerată a fost că, reținerile la plată pot fi influențate de două variabile specifice procesului de administrare, respectiv intrările noi și suspendările la plată. În următoarele două grafice am analizat cum este influențată variabila dependentă - ponderea beneficiarilor reținuți la plată în numărul total de beneficiari, față de ponderea beneficiarilor suspendați, în numărul total de beneficiari la alocația pentru susținerea familiei și respectiv la alocația de stat pentru copii.

În aceste grafice, (scatterplot) fiecare punct are două coordonate și reprezintă valoarea indicatorilor pentru fiecare județ ( care este marcat cu o culoare și un semn), fiecare lună din intervalul ianuarie 2012 – mai 2022 pentru prestația analizată.

---

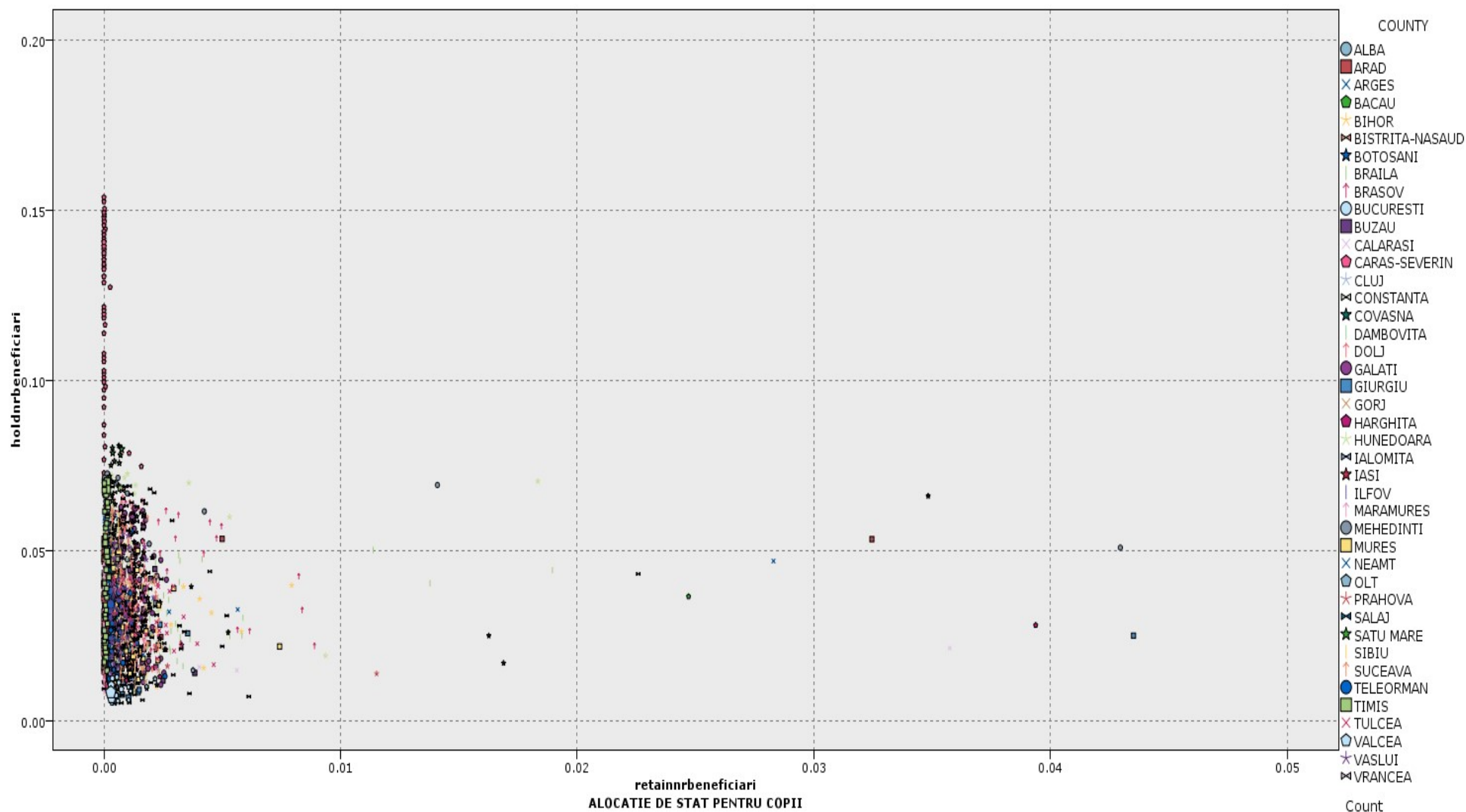
<sup>2</sup> <https://www.theiia.org/en/products/learning-solutions/webinar/practical-application-of-rpa-ai-and-machine-learning-in-internal-audit/>

<sup>3</sup> pe parcursul orelor de studiu individual



Graficul nr.1: Rețineri la plată ASF pentru intervalul ianuarie 2012 – mai 2022

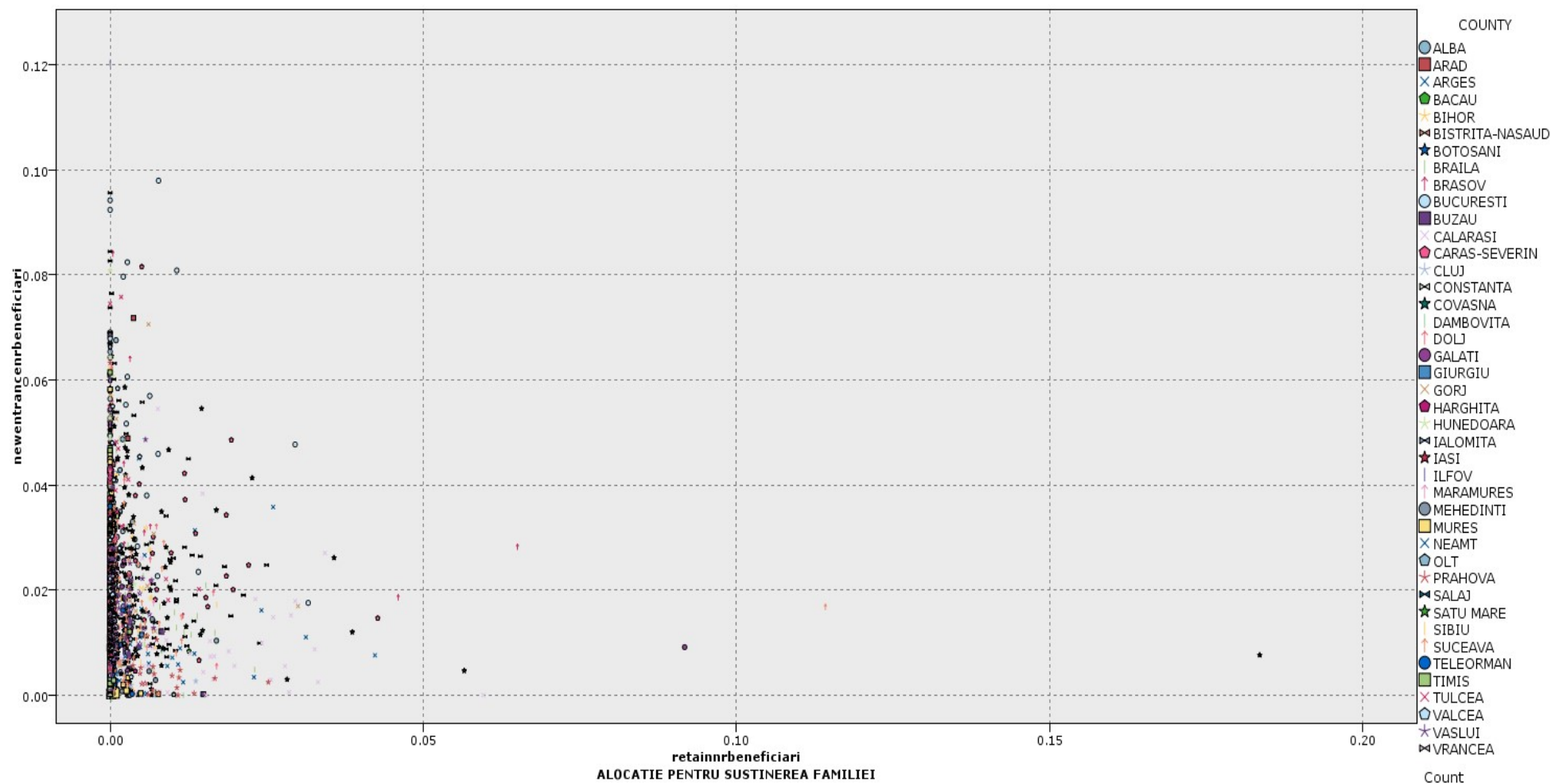
Graficul de mai sus relevă faptul că, la ASF, reținerile sunt zero sau în procente foarte mici, la ponderi foarte mari sau foarte mici ale indicatorului pondere a suspendărilor . Nu putem spune că există o legătură liniară directă între ponderea reținerilor și ponderea suspendărilor la plată, ci că mai degrabă poate fi vorba de o distribuție statistică apropiată de cea normală, care poate indica o legătură nonliniară.



Graficul nr.2: Rețineri la plată ASC pentru intervalul ianuarie 2012 – mai 2022

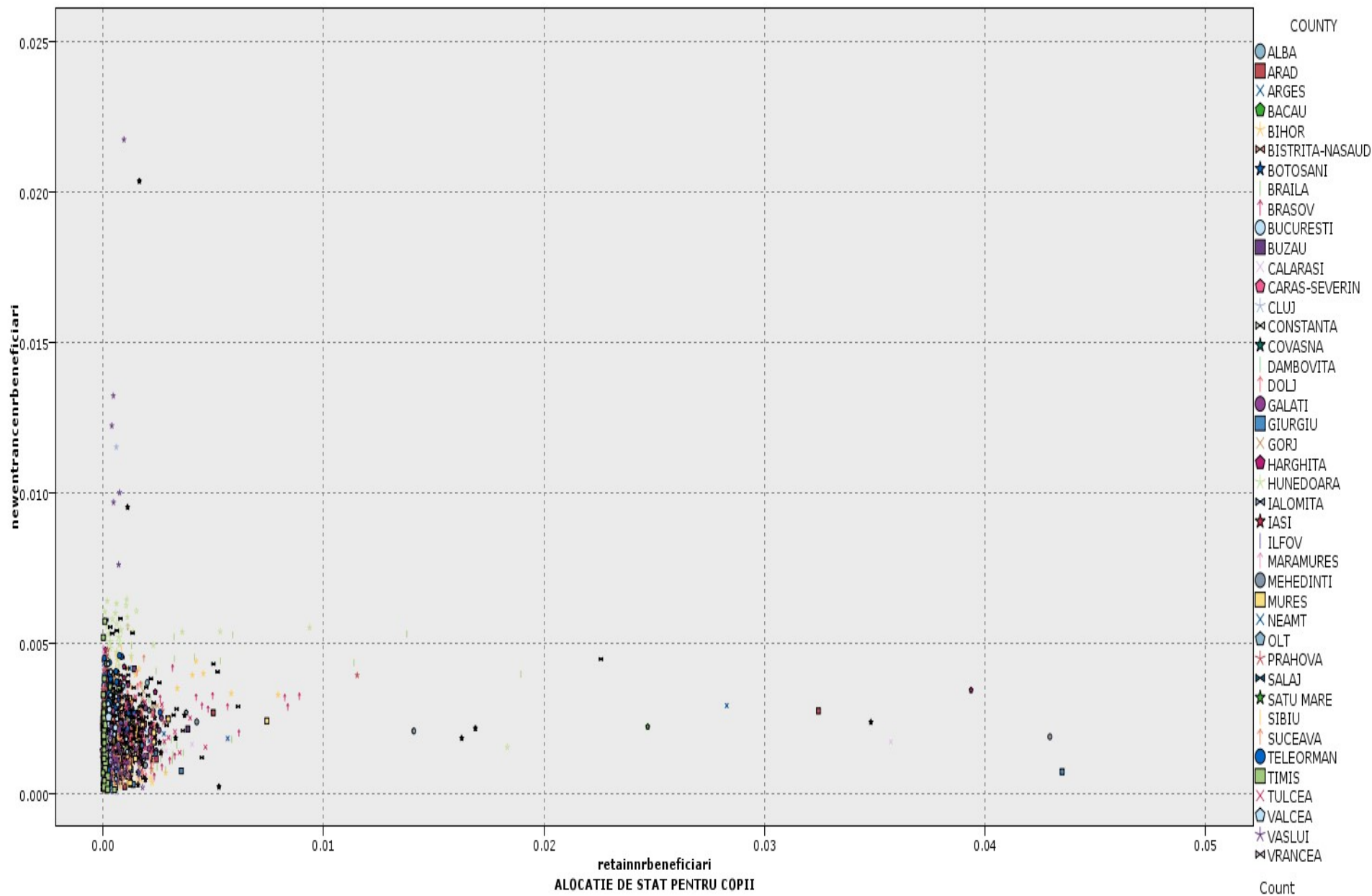
Similar, la alocația de stat pentru copii, se constată o repartizare apropiată de cea normală a distribuției valorilor, dar cu alți parametri. Pe de o parte, cu cât ponderea suspendărilor este aproape de 0 sau are valori foarte mari, cu atât ponderea reținerilor are valori scăzute .

În continuare, s-a analizat în ce mod ponderea numărului de beneficiari reținuți de la plată este influențată de cealaltă variabilă explicativă, respectiv ponderea beneficiarilor nou intraților în plată.



Graficul nr.3: Ponderea numărului de beneficiari reținuți de la plată, față de ponderea beneficiarilor nou intrați în plată la ASF

Din graficul de mai sus nu s-a observat o legătură între ponderea reținerilor și ponderea noilor intrați la ASF .



Graficul nr.4: Ponderea numărului de beneficiari reținuți de la plată, față de ponderea beneficiarilor nou intrați în plată la ASC.

Din graficul de mai sus, la ASC se observă o distribuție apropiată de cea normală, în jurul valorii 0,0025, dar se remarcă și excepții. Excepțiile au putut de altfel sta la baza dezvoltării raționamentului profesional al auditorilor interni în cadrul misiunii .

În general, ponderea reținuțiilor este foarte mică, dar se constată că depinde non liniar de ponderea numărului de beneficiari noi intrați în plată.

Evaluarea modului cum au fost implementate politicile publice aferente beneficiilor sociale, precum și dezvoltarea raționamentului profesional al auditorilor, poate fi completată prin analize cantitative avansate , specifice tehnicilor de inteligență artificială . În acest caz , s-a utilizat în primul rând metoda arborilor decizionali .

Arborii decizionali se definesc ca o tehnică de explorare a datelor cu potențial, atât predictiv cât și descriptiv. Originea sa rezidă din faptul că, rezultatul este expus utilizatorului sub forma unui grafic de tip arbore. În general, algoritmi care formează arbori decizionali pot fi aplicate fără restricții legate de tipul datelor.

Algoritmi de construire a arborilor decizionali impun un număr mic de parcurgeri a setului de date utilizat. Forma output-ului permite nu numai realizarea de previziuni și clasificări, ci și descrierea relațiilor existente și analiza datelor cantitative.

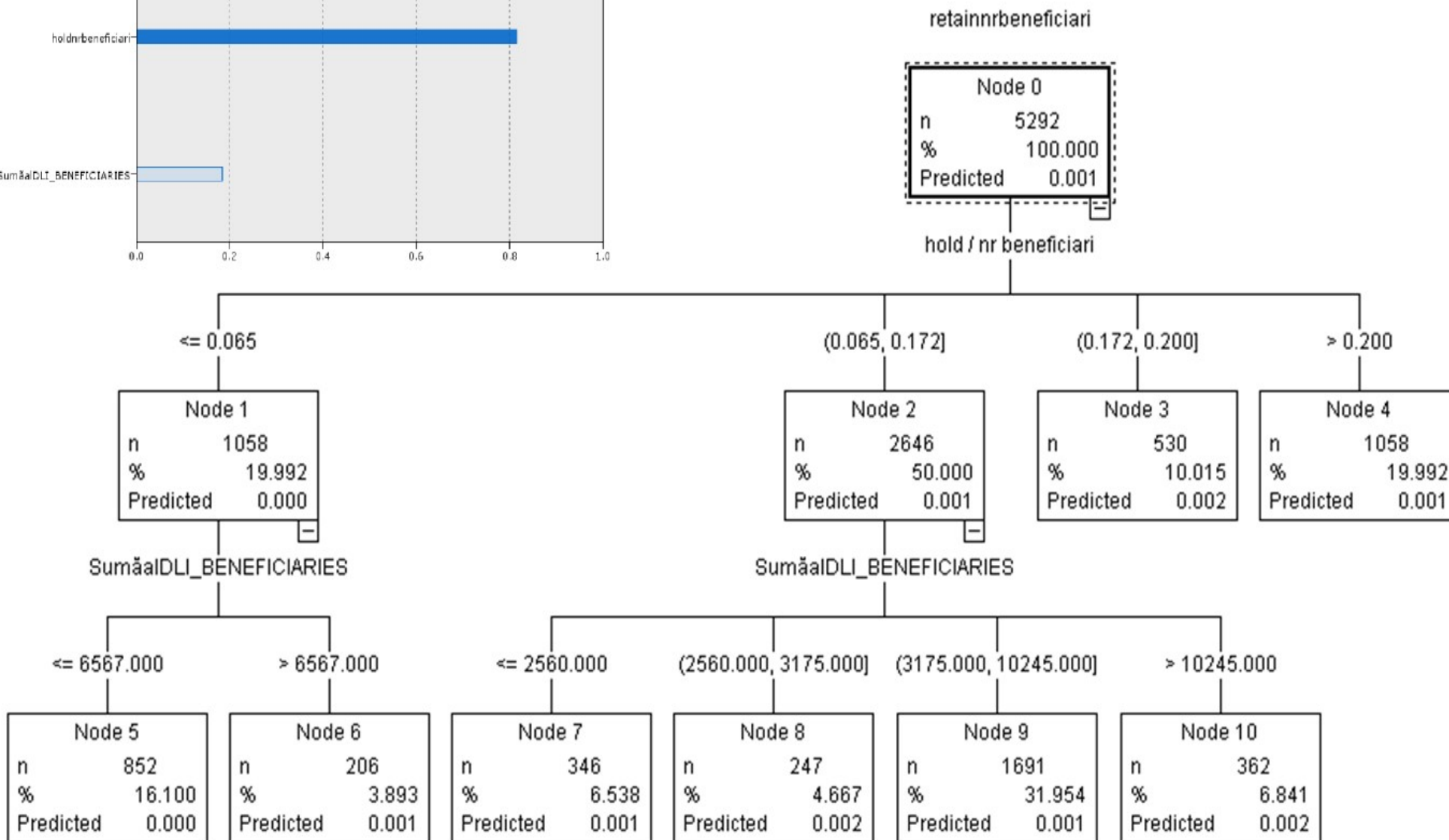
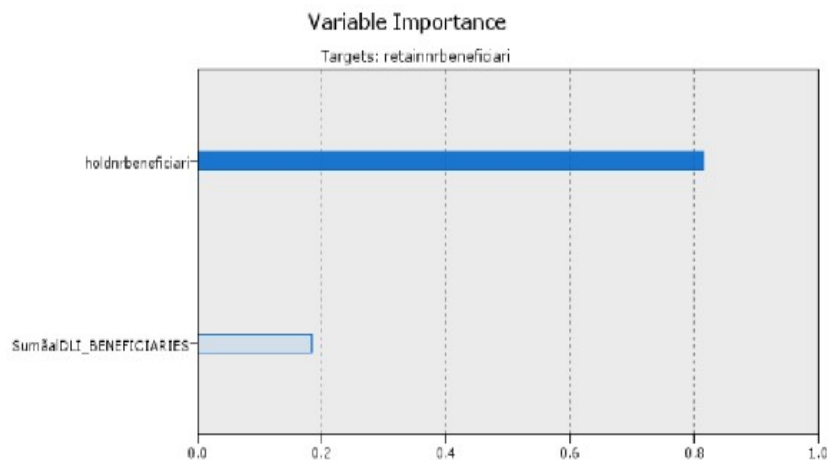
Arborii decizionali CHAID care au fost utilizați în prezentul studiu (Chi Square Automatic Interaction Detection) se bazează pe testul din tabelele de contingență.

Concret, în studiu, s-au avut în vedere ponderea reținerilor la plată, față de ponderea suspendărilor și ponderea noilor intrați în plată (față de numărul total de beneficiari).

A fost introdus în studiu și un indicator de volum, dat de numărul total de beneficiari, în luna de plată. Analiza CHAID este folosită de obicei pentru a construi un model predictiv și pentru a contura anumite grupuri. Analiza CHAID utilizează variabile predictoare explicative (de ex. ponderea suspendațiilor, față de nr. de beneficiari, ponderea noilor intrați, față de nr. de beneficiari) pentru a împărți baza de date într-o serie de subgrupuri care împart caracteristici similare numite „*arborele de decizie*”. Aceste subgrupuri permit predicția apartenenței la grup pe baza caracteristicilor unui tip.

Deoarece eșantionul este împărțit în mod repetat, tehnica funcționează cel mai bine folosind dimensiuni mari ale eșantionului.

Pentru alocația de sprijin familial, conform figurii de mai jos, s-a constatat că este importantă variabila pondere a suspendărilor și în secundar numărul de beneficiari. La valori ale ponderii suspendărilor mari, variabila explicativă: număr de beneficiari nu are o importanță foarte mare. Variabila de pondere a noilor intrați nu este semnificativă.

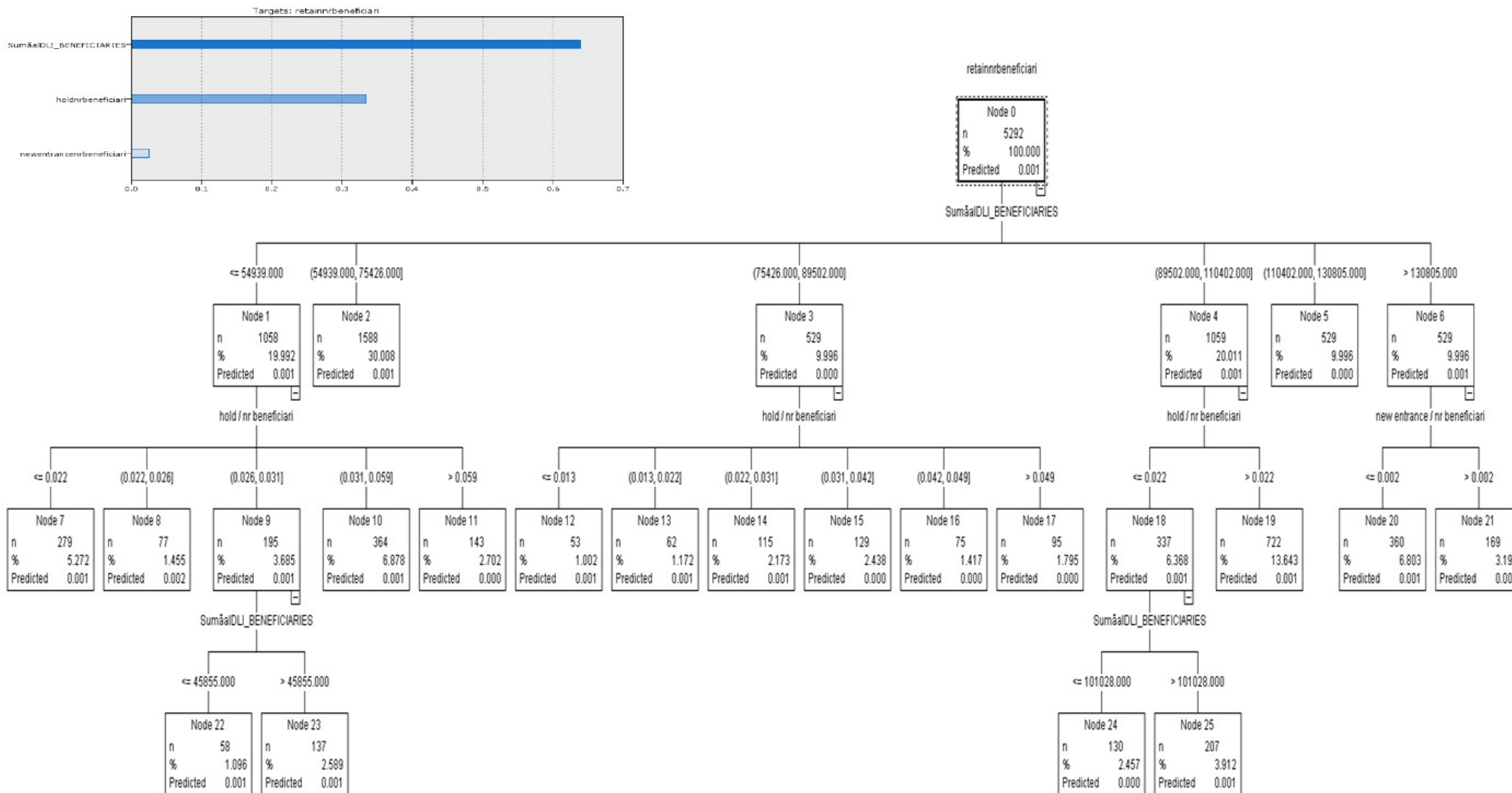


Graficul nr.5: Ponderea reținerilor la plată, față de ponderea suspendărilor și ponderea noilor intrați în plată (față de numărul total de beneficiari) la ASF



Pentru alocația de stat pentru copii, toate variabilele considerate au prezentat importanță și în principal, cea dată de numărul de beneficiari. Ponderea noilor intrați în plată are o importanță mică și doar la valori mari ale numărului de beneficiari. La diverse nivele ale numărului de beneficiari conform figurii de mai jos pot fi identificate în pofilele de administrare .

4

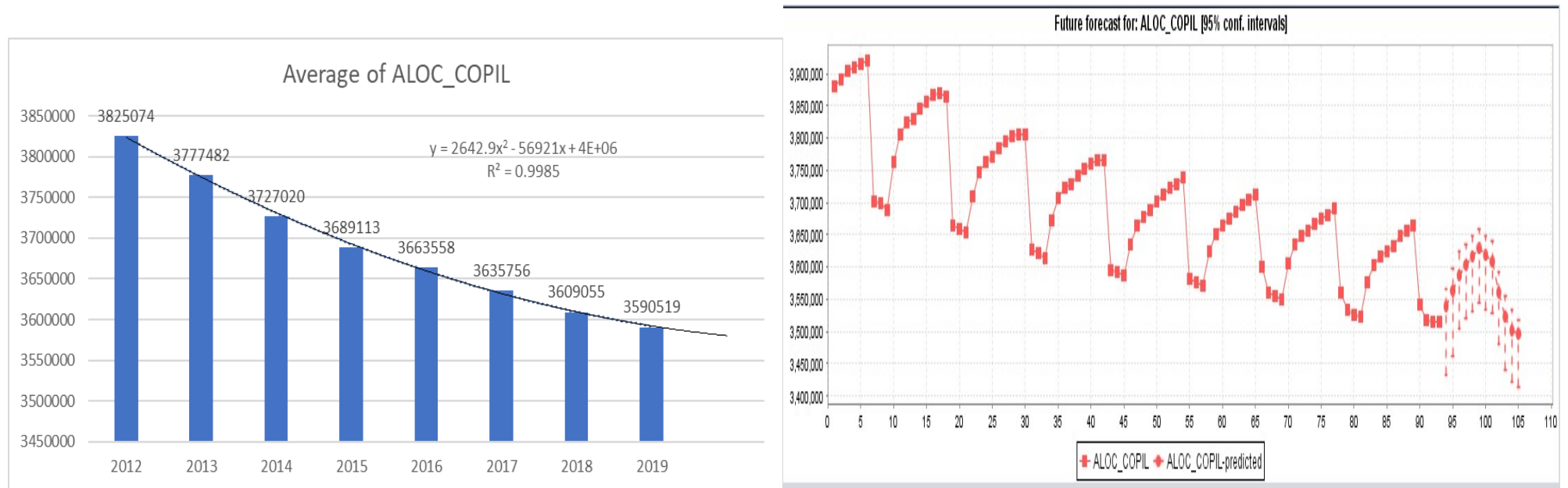


Graficul nr.6: Ponderea reținerilor la plată, față de ponderea suspendărilor și ponderea noilor intrați în plată (față de numărul total de beneficiari) la ASC

<sup>4</sup> nivele ale ponderii suspendărilor în total număr de beneficiari.

O altă metodă ce a fost avută în vedere de auditori a fost dată de predicția numărului de beneficiari la alocația de stat pentru copii . Prognoza seriilor de timp este procesul de utilizare a unui model pentru a genera predicții (prognoze) pentru evenimente viitoare bazate pe evenimente trecute cunoscute.

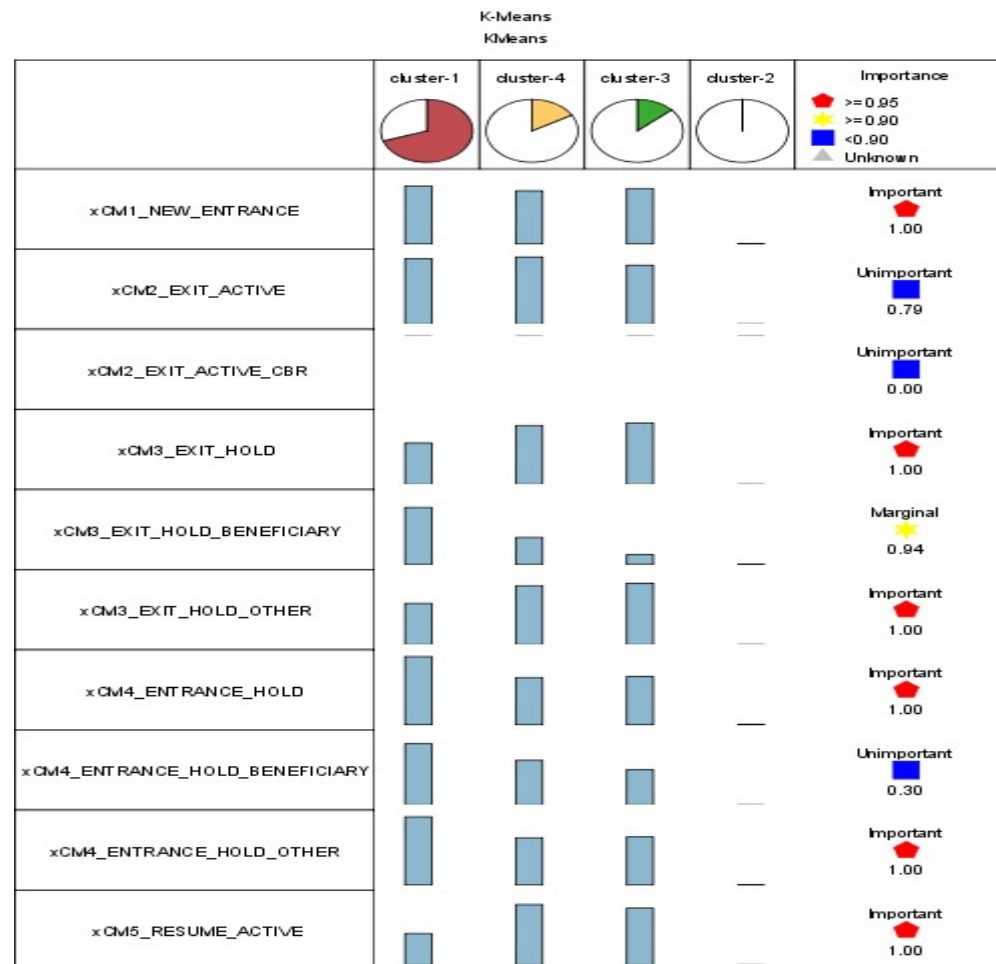
În graficul următor de mai jos am reprezentat mediile anuale ale numărului beneficiarilor de ASC , care a fost urmată de aplicarea unui model de prognoză care include și algoritmi de inteligență artificială în predicția seriilor de timp . În cadrul misiunii au fost efectuate predicții similare cu privire la alți parametri de administrare . Aceste analize au putut sta și ele la baza dezvoltării raționamentului profesional al auditorilor .



Graficul 7 . Model de predicție nesupervizată a tendinței utilizând algoritmi de analiză a seriilor de timp și modele de AI Weka

Se remarcă din model tendința de atenuare a scăderii numărului de beneficiari de ASC .

O altă analiză specifică metodologiilor de inteligență artificială utilizată în evaluarea preliminară a fost realizată la alocația de stat pentru copii prin aplicarea unui algoritm de clusterizare care a cuprins indicatori derivați din componentele de administrare raportate la numărul de beneficiari ( pe județ și pe luni ) . Analiza cluster este un mod de analiză nesupervizată de grupare care are obiectivul de a identifica date cât mai asemănătoare între ele care să fie cuprinse în categorii cât mai diferite .



Graficul 8 . Analiza cluster

Pe baza acestui tip de analiză, așa cum rezultă de altfel din diagrama de mai sus, se remarcă importanța unor indicatori care pot fi utilizați la alegerea unor esantioane. În contextul construcției unor esantioane bazate pe raționamentul profesional la alocația de stat pentru copii a putut fi fost utilizată pentru prima oară în cadrul unei misiuni (modelul a mai fost rulat formal în faza de completare a analizei riscurilor specifică planurilor de audit) o altă tehnologie de inteligență artificială bazată pe identificarea unor anomalii în date. Metoda este bazată pe un index compus din trei câmpuri care determină includerea în clase de excepții.

Modelele de detecție a anomaliei sunt folosite pentru a identifica acele cazuri neobișnuite din mulțimea de date și, spre deosebire de alte metode de modelare care înmagazinează reguli despre cazurile neuzuale, modelele de detecție a anomaliei stochează informații despre cum arată datele identificate ca anormale în procesul de analiză.

Detectarea anomaliei este o metodă nesupervizată, ceea ce înseamnă că nu necesită un set de date prealabil care să conțină cazurile cunoscute ca punct de începere.

În timp ce metodele de identificare privesc, în general, una sau două variabile în același timp, detectarea anomaliilor poate examina un număr mare de câmpuri pentru a identifica clusterse sau grupuri pereche pentru care s-au făcut înregistrări similare.

Apoi, fiecare înregistrare este comparată cu altele din grupul pereche pentru a fi identificate posibile anomalii. De exemplu, algoritmul poate să considere înregistrări în doua clusterse diferite și să le marcheze pe acelea care sânt departe de centrul oricărui cluster.

Camp	Contor de \$O-FieldImpact-1	Medie de \$O-FieldImpact-1	Max de \$O-FieldImpact-1
xCM1_NEW_ENTRANCE	6	0.27121067	0.760340793
xCM2_EXIT_ACTIVE	11	0.246729116	0.326668918
xCM3_EXIT_HOLD	6	0.179847206	0.479114077
xCM3_EXIT_HOLD_BENEFICIARY	6	0.468370839	0.754554018
xCM3_EXIT_HOLD_OTHER	4	0.137390144	0.478761633
xCM4_ENTRANCE_HOLD	3	0.385645046	0.397727494
xCM4_ENTRANCE_HOLD_BENEFICIARY	8	0.301953775	0.767660229
xCM4_ENTRANCE_HOLD_OTHER	3	0.385925086	0.398014463
xCM5_RESUME_ACTIVE	2	0.42692414	0.826357641
xDIF_BENEFICIARIES_OUT_PAID	5	0.317668778	0.467664411
xDLI_BENEFICIARIES_HOLD	1	0.019435847	0.019435847
xDLI_BENEFICIARIES_RETAIN	6	0.663603337	0.8419653
xDLI_DENOMITOR_ALL	17	0.089259535	0.099107561
xDLI_DENOMITOR_CBR	8	0.081180791	0.094424074
xDLI_DENOMITOR_FA_GMI	9	0.073463245	0.094405803
xDLI_INDICATOR_ALL	1	0.010397735	0.010397735
xDLI_INDICATOR_FA_GMI	5	0.011902	0.024208289
xDLI_OTHER_AMOUNT_ARR_RES_COM	8	0.020462904	0.031242704
xDLI_OTHER_AMOUNT_DEBTS_COMP	5	0.683702111	0.974242405
xIN1_ACTIVE	12	0.228039137	0.326771278
xIN2_HOLD	12	0.077043222	0.099591334
xIN2_HOLD_BENEFICIARY	10	0.28004748	0.687812393
xIN2_HOLD_OTHER	13	0.070577552	0.099567649
xOUT1_ACTIVE	5	0.326392007	0.467664411
xOUT2_HOLD_BENEFICIARY	11	0.234476018	0.444045331

Graficul 8 Analiza valorilor outliers

În tabelul de mai sus sunt reliefate astfel de tipologii care au putut sta la baza construirii unor eşantioane bazate pe raționamentul profesional .

## Concluzii

În acest studiu a fost analizată legătura dintre indicatorii de administrare aferenți alocației de stat pentru copii și alocației pentru sprijin familial, prin utilizarea arborilor de decizie și alte tehnici de inteligență artificială și data mining. Studiul a condus la creșterea calității în exercitarea raționamentului profesional, ce implică și cunoașterea unor profile de risc de administrare utilizând astfel de tehnologii de date . Studiul indică o dependență nonliniară între ponderea reținerilor la plată, față de ponderea suspendărilor la plată, aferente ASC și ASF.

Limitele cercetării sunt date de specificul metodei utilizate. Direcții viitoare de cercetare vor fi reprezentate de includerea în analiză și a altor componente sau prestații sociale și în determinarea funcției de distribuție non liniare și utilizarea ( după finalizarea misiunilor de audit ) a unei alte tehnici de inteligență artificială ( deep learning ) care este dată de rețelele neuronale .

## Bibliografie

1. \*\*\* Manual de audit de conformitate – Curtea de Conturi, disponibil on line:  
[https://www.curteadeconturi.ro/uploads/ba799d10/00f73f99/0b553817/efbebc95/dcafc673/e1638e13/9edcf166/e54c9a73/Manual\\_de\\_audiot\\_de\\_conformitate.pdf](https://www.curteadeconturi.ro/uploads/ba799d10/00f73f99/0b553817/efbebc95/dcafc673/e1638e13/9edcf166/e54c9a73/Manual_de_audiot_de_conformitate.pdf) ;
2. Jiawei Han, Micheline Kamber, Jian Pei - Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition, 2015.