

ANALIZA INTELIGENTĂ A DATELOR ÎN AUDITUL FINANCIAR

CONF. DR. GHEORGHE MIRELA
ASIST. DRD. BOLDEANU DANA
ASIST. DRD. ALECA OFELIA

Academia de Studii Economice București, Facultatea Contabilitate și Informatică de Gestiune

Adresa: Aleea Botorani, nr.10, bl.v57, sc.a, ap.11, sect.5, Cod Postal 050812

Telefon: 0723858611/021-4117625

Email: mirelaghe@gmail.com

The Data Mining techniques can support the auditor's judgment, providing him an automatic analysis of an important and complex set of data with the purpose of finding models with specific „patterns” or that otherwise will pass unnoticed. Choosing samples will stand out as an essential functionality that can be approach through this technology because a clustering analysis will allow auditors to identify the samples from several representatives groups classified in a less noticeable way using other tools.

1. Repere privind analiza inteligentă a datelor

Creșterea complexității sistemelor de contabilitate informatizată cât și a procesului de audit în sine au marcat o nouă tendință în dezvoltarea software-ului de audit, prin integrarea tehnicilor OLAP, Data Mining – ce pot asigura o analiză multidimensională a datelor, detectarea fraudelor, descoperirea cunoștințelor ascunse în spatele datelor clasice, oferind astfel noi oportunități auditorilor.

Ultimii ani au marcat o reorientare în utilizarea volumelor de date acumulate, de la un proces de cercetare retrospectiv spre unul care abordează mai mult aspecte legate de viitor, schimbare ce s-a putut impune ca urmare a maturizării tehnologiilor legate de Data Mining.

Literatura de specialitate abundă în definiții pentru conceptul de Data Mining, dar în esență ele exprimă același lucru: un proces de extragere a noi informații din colecțiile de date existente, informații ce pot răspunde nu numai la întrebarea ce se întâmplă, ci și de ce se întâmplă ? Un Data Mining are, în general, două obiective: unul de descriere ce constă în determinarea variabilelor semnificative și a influențelor acestora, cât și unul de predicție.

Analiza și exploatarea datelor unei baze de date multidimensionale se realizează folosind tehnici de analiză avansată ce includ algoritmi statistici, inteligență artificială, rețele neuronale, logică fuzzy, algoritmi genetici și vizualizarea datelor. Data Mining nu este, în fapt o “rețetă universală” pentru orice problemă de gestiune, ci am putea spune că aportul său se rezumă la următoarele categorii de operații:

- Descrierea datelor
- Analiza dependențelor
- Clasificări și predicții
- Analiza cluster-elor
- Analiza excepțiilor

Aceste operații pot asigura o analiză complexă a tranzacțiilor clientului, identificarea tranzacțiilor neobișnuite, sau chiar frauduloase, definirea unor trenduri, previziuni, fapt pentru care, integrarea acestei tehnologii poate reprezenta un instrument extrem de util îndeosebi auditorilor interni din cadrul oricărei organizații.

Detalierea analizei procesului de audit și a metodelor Data Mining au condus la identificarea unor posibile arii de integrare¹⁴⁵:

Procesul de audit	Metode Data Mining	Posibile arii de integrare
1. Acceptarea clientului		Plecând de la declarațiile financiare ale anilor anteriori, media ramurii de activitate respective (rating-ul industrial), clientul poate fi apreciat „favorabil” sau „nefavorabil”, prin utilizarea metodelor de predicție și clasificare.
	Clasificarea și predicția Analize de evaluare	
2. Planificarea		
Evaluarea riscului	Analiza dependențelor Clasificarea și predicția	Analiza dependențelor poate fi considerată ca fiind operația Data Mining oportună identificării factorilor de risc. Setul de date cerute a fi analizate poate fi alcătuit din declarațiile financiare ale anilor anteriori, media ramurii economice, indicele inflației, diagramele de flux ale sistemului. În acest context, evaluarea riscurilor la nivelul fiecărui domeniu semnificativ poate fi analizată utilizând clasificarea și metodele de predicție.
3. Evaluarea controlului intern		
Testarea controalelor	Analiza cluster-elor Analiza excepțiilor	<i>Testarea controalelor</i> are ca punct de plecare selectarea eșantioanelor, urmată apoi de o analiză a excepțiilor. Rezultatele finale pot fi clasificate ca fiind satisfăcătoare sau nesatisfăcătoare, situație care impune aplicarea testelor de detaliu.
4. Aplicarea testelor de detaliu		
Proceduri analitice	Clasificarea și predicția Analiza excepțiilor	Analiza tranzacțiilor contabile ale anilor anteriori, asociate cu mediul economic în ansamblul său poate conduce la o previzionare a datelor curente.
Teste de detaliu asupra tranzacțiilor	Analiza cluster-elor Analiza	Tranzacțiile contabile cu caracteristici similare sunt grupate prin tehnica „clustering”, ceea ce va permite selectarea eșantioanelor din fiecare cluster, împreună cu date neobișnuite identificate

¹⁴⁵ Elemente prelucrate după sursa: Sirikulvadhana S. - Data Mining As A Financial Auditing Tool
683

	excepțiilor	prin metoda analizei excepțiilor.
5. Raportul de audit		
Formularea concluziilor	Analiza dependențelor Clasificarea și predicția	Utilizând analiza dependențelor, pot fi colectate probele legate de împrejurările ce vor afecta diferitele tipuri de opinii formulate. Apoi, analizând rezultatele testelor întreprinse și alte circumstanțe determinate, pot fi raportate concluziile și definitivată opinia auditorului.

Tabelul 1.2 Posibile arii de integrare a tehnicilor Data Mining în cadrul procesului de audit

Studiul realizării unui model Data Mining subliniază necesitatea utilizării seturilor de date bazate pe informații istorice, fapt pentru care în opinia noastră, anumite etape ale procesului de audit, precum *acceptarea clientului și planificarea* nu pot fi abordate întotdeauna prin prisma tehnologiei Data Mining. În schimb, realizarea etapelor de *evaluare a controlului intern și aplicarea testelor de detaliu*, permit o analiză complexă a datelor ce conduce adeseori la detectarea fraudelor.

2. Microsoft SQL Server 2005 instrument pentru implementarea tehnicilor Data Mining

Oferind o platformă integrată și robustă de Business Intelligence ce presupune prelucrarea datelor, realizarea modelelor, validarea și previziunea prin folosirea tuturor etapelor CRISP-DM¹⁴⁶, Microsoft® SQL Server™ 2005, este un instrument util în luarea deciziilor de business documentate.

Business Intelligence Development Studio oferă posibilitatea de: pregătire și curățare a datelor, vizualizarea modelelor existente, editarea de noi modele, testarea acurateții modelului și comparații între modelele, realizarea de previziuni pe baza modelelor folosind limbajul DMX (Data Mining Extensions) sau cu ajutorul instrumentului Prediction Query Builder. SQL Server 2005 este un instrument ușor de utilizat (cu ajutorul interfaței vizuale), integrând în cadrul unui SGBD componenta Data Mining și OLAP, fapt ce permite utilizatorilor reducerea costurilor materiale, financiare și umane.

3. Studiu de caz

În etapa de realizare a testelor de detaliu, auditorul trebuie să analizeze un volum considerabil de date și să extragă eşantioane de date, pentru a verifica completitudinea, integritatea și realitatea acestora. Tehnicile data mining pot sprijini raționamentul auditorului, oferindu-i acestuia o analiză automată a unui set larg și complex de date cu scopul de a descoperi modele, „șabloane” deosebite sau tendințe care, altfel, ar trece neobservate.

Considerăm următoarea situație furnizată auditorilor, situație ce include tranzacțiile realizate de SC ALFA SA, pe o perioadă de 2 luni (numărul de tranzacții este de cel puțin 2000):

¹⁴⁶ Cross Industry Standard Process for Data Mining - standard asociat procesului de Data Mining

idOperatie	cod_validare	explicatie	cont_d	cont_c	suma	Nr Doc	tip operatie	eronata
284	00019	dobanda con	5121.02	766	0.16	dob	Transfer	0
282	00019	TVA	4426	401.00009	21.18	4005215	Transfer	0
256	00019	COMISION E	627	5121.02	9.1	C2	Transfer	0
283	00019	NIR	371	401.00009	111.47	4005215	Transfer	0
36	00008	Plata fact.	401.00008	5121.02	3969.19	2	Transfer	0
39	00009	Plata fact.	401.00009	5311	132.65	8611957	Transfer	0
246	00019	Factura de ie	607	371	1086.8	4830309	Transfer	1
244	00019	Iesiri	4111.00002	707	504	4830308	Transfer	0

Tabelul 3.1. Situația testelor efectuate asupra tranzacțiilor în perioada 01-11-05...31-12-05, în cadrul societății S.C. ALFA S.A.

Pentru a putea identifica anumite „artificii” ale clientului, cum ar fi:

- tranzacții mici ce sunt executate în mod repetat într-o anumită perioadă a lunii
- tranzacții mari care se execută cu o frecvență regulată fie la începutul/sfârșitul lunii
- tranzacții ce nu se încadrează în clustere (excepțiile)
- tranzacții neobișnuite, de valori similare, ce au loc repetat.

s-au realizat diferite analize, prin tehnici data mining cu ajutorul instrumentului Microsoft SQL Server 2005. În acest sens, s-a construit un model de analiză a numărului de tranzacții și a valorilor asociate acestora pe baza seriilor temporare.

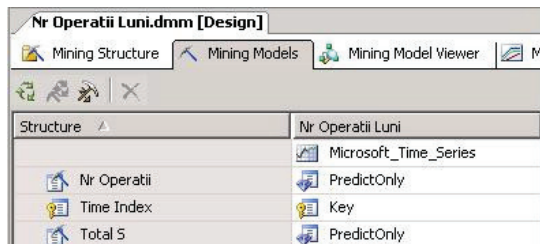


Fig 3.1. Model de analiză a numărului de tranzacții și a cuantumului valoric al acestora.

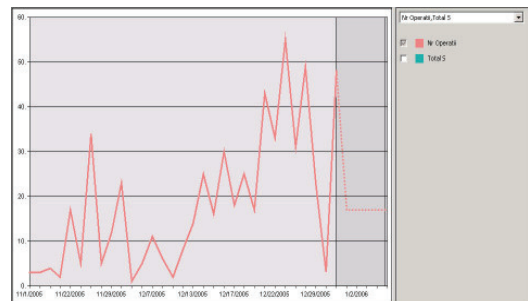


Fig 3.2. Intensitatea numărului de tranzacții

Din grafic se observă o ușoară tendință de creștere a tranzacțiilor în a doua lună, frecvența acestora încadrându-se într-un anumit șablon, ce atinge un vârf (valoare maximă) la mijlocul perioadei.

Pentru auditor această imagine este relevantă, deoarece îl va determina să verifice cu precădere perioada cea mai solicitată. Alte analize efectuate precum analiza clustering, arbore de decizie au condus la selecția eșantionului.

Cercetare s-a extins și asupra eșantionul extras și verificat, auditorul detectând erori în datele clientului. Astfel, pentru fiecare operație testată atributul „eronată” va lua valoarea 0 dacă nu s-au depistat erori și valoarea 1 în caz contrar. Identificarea tranzacțiilor eronate pentru următoarele perioade a necesitat realizarea unui model de analiză și previziune a tranzacțiilor eronate pe baza arborilor de decizie.

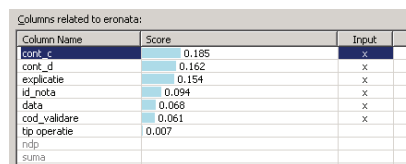
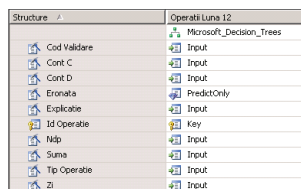


Fig 3.4. Intensitatea legăturilor

Fig 3.3. Model de analiză și previziune a tranzacțiilor.

dintre atribute

Pe baza legăturilor dintre atribute (variabila ce se dorește să fie previzionată este eroarea) se va obține diagrama arborelui de decizie și diagrama dependențelor.

Previziunea realizată ne relevă probabilitatea ca o anumită tranzacție să fie sau nu eronată. Din grafic, se observă că numărul relativ redus al operațiilor eronate corespund preponderent sumelor peste 5484 (încadrate în categoria sumelor mari).

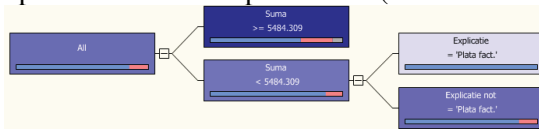


Fig 3.5. Arbore de decizie pentru luna decembrie

idOperatie	Eronata	Probabilitate
1	0	0.99972491575...
2	0	0.99972491575...
3	0	0.67549668874...
4	0	0.99972491575...

Fig 3.6. Probabilitate de apariție a erorilor atasate operațiilor din luna ianuarie

Pentru luna ianuarie, se preconizează că operația identificată prin nr 1 are probabilitatea de 99% să nu fie eronată, în timp ce operația numărul 3 are șanse de numai 67%. Aceste analize, previziuni realizate îi oferă auditorului o imagine complexă, reală asupra datelor clientului oferindu-i un real ajutor în selecția eșantioanelor. În afara interfeței grafice se pot realiza predicții pe baza unui model și prin intermediul limbajului DMX (**Data Mining Extensions**). De exemplu, pentru a se afișa probabilitatea ca operația “Plată factură” încadrată în tipul operației „Transfer”, având suma de 7000 RON, validată de operatorul cu codul 00018, sintaxa DMX este următoarea:

```
SELECT 1 as Eronata, PredictProbability([Eronata],1) as [Probabilitate]
FROM [Operatii Luna 12] NATURAL PREDICT JOIN
(Select 7000 as Suma,00018 as [Cod Validare],’Transfer’ as [tip operatie], 16 as zi, „Plata fact.” As explicatie) as t
```

Rezultatul rulării acestei cereri în cadrul Microsoft SQL Server 2005 este prezentat în figura alăturată.

Eronata	Probabilitate
1	0.24503311258...

Bibliografie

1. Sirikulvadhana S. – „Data Mining As A Financial Auditing Tool, Thesis in Accounting”
2. Seth P., MacLennan P.- “Microsoft SQL Server 2005 - Data Mining Tutorial “
3. Han J., Kamber M. – „Data Mining – Concepts and Techniques „ (www.cs.sfu.ca)
4. www.microsoft.com