

**NOVI METODOLOŠKI PRISTUP REVIZIJE FINANSIJSKIH
IZVEŠTAJA****Primena mašinskog učenja tehnikom samoučeće multivarijantne
- diskriminacione analize -****Ljubiša V. Stanojević**

Univerzitet Singidunum, Beograd

Tel. +381112451-665, 063-222158

e-mail. ljstanojevic@singidunum.ac.rs

Sažetak: Živimo u vremenima koje karakteriše izrazita aleatornost poslovnog okruženja, koja se posebno izražava u domenu korisnosti finansijskih informacija - izveštaja. Klasičan pristup analize finansijskih izveštaja je već duži niz desetleća na marginama tzv. klasičnih „površinskih analiza” (horizontalne, vertikalne, racio brojeva, itl.). Novi metodološki pristup baziran na primeni mašinskog učenja zasigurno širi metodološke okvire koji će svakako zameniti klasičan pristup jer ima sušastvenu prednost, a to je pronalaženje skrivenih znanja sadržanih u finansijskim izveštajima, posebno kada se radi o postupku obavljanja revizije. Metode, koje se zasnivaju na simulaciji načina rezonovanja ljudskog mozga (učenju) je dalo iznenadjujuće dobre rezultate. Ovaj prilog je pravo posvećen ilustraciji primene ovog novog metodološkog aparata na primeru kontrole tačnosti unosa podataka u kompjuterskoj obradi u postupku validacije internih kontrol.

Ključne reči: finansijski izveštaj, mašinsko učenje, veštačka inteligencija, multivarijantna diskriminaciona analiza, samoučenje, efikasnost internih kontrola.

Abstract: We live in times that are dominantly characterized by an aleatory quality of the business environment, that is especially expressed in the realm of financial information utility – reports. A classical approach to financial reports analysis (horizontal, vertical, ratio numbers, etc.) has already been called “cursory analyses” for decades. A new methodological approach based on the application of AI learning would surely expand the methodological framework and definitely substitute the classical approach because it contains an essential advantage, and that is to uncover hidden knowledge embodied in the financial reports, especially when those financial reports are used in a financial audit. The methods, which are based on the simulation of a human brain reasoning (learning), have given surprisingly good results. This article is dedicated to the illustration of applying this new methodological apparatus by showing an example of the accuracy control of data entry in computer processing in the procedure of validating internal controls.

Key words: financial report, AI learning, artificial intelligence, multivariant discriminant analysis, self-learning, internal controls efficiency.

1. Uvod

Istraživanje podataka (*data mining*) obuhvata primenu metoda mašinskog učenja, kao posebna oblast veštačke inteligencije³¹ (*Artificial Intelligence*) odnosi se na razvoj algoritama i tehnika koji omogućavaju kompjuterima da "uče", kao i drugih metoda, za pronalaženje uzoraka u

³¹ **Artificial intelligence (AI) ili Veštačka inteligencija** je oblast kompjuterskih nauka koja proučava inteligentna ponašanja, učenja i adaptaciju mašina. Istraživanja u oblasti Veštačke inteligencije su u vezi sa proizvodnjom mašina u cilju automatizacije zadataka zahtevanog inteligentnog ponašanja.

razmatranim podacima. Data mining poznat i kao proces otkrivanja znanja u bazama podataka (*Knowledge-Discovery in Databases (KDD) or Knowledge-Discovery and Data Mining*).

Svi metodi data mining-a koriste učenje zasnovano na indukciji. To je proces definisanja opštih konceptnih definicija posmatranjem specifičnih primera iz kojih se uči. Induktivno učenje uz pomoć učitelja (*Nadzirano učenje*) je učenje koje se vrši na osnovu primera klasifikovanih primera, odnosno donesenih odluka.

Indukcija ili induktivno zaključivanje, ponekad se naziva i induktivna logika je proces donošenja zaključaka u kome premise nekog argumenta podržavaju zaključak, ali ga ne garantuju. *Induktivno učenje konceptata* je forma mašinskog učenja čiji je cilj indukcija takvih opisa konceptata (kategorija, klasa objekata), koji su razumljivi ljudima koji se bave određenom problematikom, odnosno koji odgovaraju konceptima koje bi proizveo čovek prilikom razmatranja istih entiteta.

Kao osnovne tehnike data mining-a možemo izdvojiti stabla odlučivanja, pravila asocijalacije, algoritam najbliže tačke i genetičko učenje. Produkciona pravila, liste odlučivanja i stabla odlučivanja su primeri razumljivog načina predstavljanja empirijskog znanja.

Metode induktivnog učenja konceptata se koriste u istraživanju podataka jer daju razumljive rezultate, koji se mogu direktno interpretirati. Tačnost predviđanja ovih metoda takođe visoka i može se uporediti sa najuspešnijim metodama induktivnog učenja".

Ovde treba napomenuti precizno šta kompjuter može da nauči. Učenje je složen proces. Možemo definisati četiri nivoa učenja (Merril i Tennzson, 1977): Činjenice - Činjenice su jednostavno iznošenje istine, Koncepti. - Koncepti su skup objekata, simbola, ili događaja grupisani zajedno zato što dele određene karakteristike, Procedure. - Procedure su tokovi akcija za dostizanje cilja, i Principi. - Principi ili pravila ponašanja predstavljaju najveći nivo učenja. To su opšte istine ili zakoni koje su osnova za ostale istine.

Napred pomenuta procedura učenja može biti sprovedena u dva slučaja i to:

učenja sa nadgledanjem i

učenje bez nadgledanja.

Učenje sa nadgledanjem (supervised learning) je tehnika mašinskog učenja na osnovu skupa podataka koji nazivamo trening setom (training data). Trening set se sastoji od ulaznih objekata i željenih izlaza. Izlazna funkcija može imati vrednost u kontinuitetu (naziva se tada regresija) ili može predvideti klasu ulaznog objekta (tada to nazivamo klasifikacijom). Osnovni cilj ovakvog učenja je predvideti vrednost izlazne funkcije za svaki valjani ulazni objekat nakon što se vidi određeni broj primera takvih instanci (sa određenim ulaznim i izlaznim vrednostima).

Najčešće, učenje sa nadgledanjem generiše globalni model koji mapira ulazne objekte na željene izlaze. U nekim slučajevima, štaviše, mapa je implementirana kao set lokalnih modela. U cilju rešavanja problema učenja sa nadgledanjem postoje sledeći koraci:

- Utvrđiti tip primera za trening. Utvrđivanja tipa podataka u našem slučaju kvantitativni podaci o internim kontrola u IT okruženju.
- Prikupljanje obučavajućeg skupa (trening skupa). Podaci o kopletnosti, tačnosti i ažurnosti podataka u IT obradi.
- Utvrđiti ulazne karakteristike. Broj karakteristika koje opisuju objekat.
- Odrediti odgovarajući algoritam učenja. Na primer, neuronske mreže, stabla odlučivanja, itsl.

- Kompletirati plan. Pokrenuti izabrani algoritam učenja na prikupljenom obučavajućem skupu. Parametri algoritma učenja mogu biti podešeni optimizacijom performansi na podskupu trening seta (validacioni set), ili putem krosvalidacije (cross-validation³²).

Učenje bez nadgledanja se razlikuje od učenja sa nadgledanjem u tom smislu što kod učenja bez nadgledanja nema *a priori*³³ izlaza. Prikupljen je skup ulaznih objekata, i učenje bez nadgledanja tretira skup ulaznih objekata kao skup slučajnih (random) varijabli. U narednom delu priloga pokušaćemo da napredne premise sublimiramo na primeru primene multivarijatne diskrecione analize.

2. Metodologija istraživanja: primena multivarijantne diskriminacione metode

Statistička tehnika definicijski opredeljena kao multivarijantna linearna diskriminaciona analiza je razvijena davnih tridesetih godine prošlog veka, kao statistička kategorija, koja formalno-kvantitativno vrši diskriminaciju između klasa bioloških i drugih fenomena i njima asociranih observacija. Ubrzo su ova istraživanja počela da se primenjuju i u drugim oblastima posebno u ekonomiji. Tako beležimo niz radova, primer radi, u domenu analize rangiranja kvaliteta državnih obveznica, određivanja kreditnog potencijala banaka i njihovih klijenata, klasifikacije grešaka u domenu računovodstveno - poslovnih transakcija, itsl. Vremenom ova tehnika je postala standardno sredstvo ne samo u prirodnim naukama, nego već i u ekonomiji, kao i u računodovostvenim i revizorskim ekspertizama. Medjutim, primena diskriminacione analize u kontekstu pronalaženja skrivenih (implicitnih) znanja (**data mining**) ima posebnu dimenziju sa razloga što se analiza ne svodi na primeu modela već se vrši nadzirano učenje na uzorku (**Nadzirano učenje**) modela diskriminacione analize u više navrata da bi rešenje bilo što tačnije.

Svrha ovog priloga je da se na jednostavan način prezentira aplikacija metode multivarijantne diskriminacione analize, a u svrhu ilustracije načina primene ove metode u domenu analize efikasnosti sistema internih kontrola.

2.1. Formalne osnove multivarijantne diskriminacione analize

Izračunavanje diskriminacione funkcije, primereno našem ispitivanju, za dve grupe u putanju, i n-nezavisnih varijabli se vrši putem izračunavanja funkcije, kako sledi.

$$Z = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

Vrednost Z u funkciji diskriminacije je opredeljena vrednošću diskriminacionog skora ili Z vrednost, koja se izračunava za svaku observaciju. Vrednost Z-a predstavlja kompoziciju od

³² **Cross-validation** - praksa particionisanja skupa podataka na podskupove tako da se analiza izvršava na pojedinačnim podskupovima, dok se jedan deo primera koristi radi potvrde i validacije inicijalnih analiza. Teorija krosvalidacije je utvrđena od strane Seymour Geisser-a.

³³ **a priori** - lat. izraz, ono što dolazi pre.

nekoliko mera koje se dobijaju iz jednostruke-jedne observacije. Relacija svake Z-vrednosti, izračunate za svaku observaciju u odnosu na prelomnu vrednost Z-a, se naziva **kritična diskriminantna vrednost** (KDV), koja se koristi da bi se utvrdilo koja od dve³⁴ grupe - e observacija je verovatnija.

Varijable X_i u diskriminantnoj funkciji su multivarijantne mere asignirane iz jednostruke observacije. Konstante A_i su koeficijenti referirani na varijablama X_i .

Koeficijenti A_i se mogu izračunati iz sledeće matrične jednakosti i to:

$$\alpha = \Sigma^{-1}(\mu_1 - \mu_2)$$

gde je Σ^{-1} = inverzna kovarijaciona matrica dimenzija (nxn),

μ_1 = je vektor sredina mera uzetih iz observacija koje stvarno pripadaju Grupi Jedan, dimenzije (nx1), i

μ_2 = je vektor sredina mera tako uzet iz observacija koje aktualno pripadaju Grupi Dva, dimenzija (nx1).

Pošto višestruka linearna diskriminaciona analiza (VLDA) predpostavlja da su kovariacione matrice iste za sve gupe, tada se koristi formula za estimaciju, kako sledi:

$$S_{ij} = \frac{\left[\sum_{a=1}^{N_1} (X_{ai} - \bar{X}_{.i})(X_{aj} - \bar{X}_{.j}) + \sum_{b=1}^{N_2} (X_{bi} - X_{bj} - \bar{X}_{.j}) \right]}{N_1 + N_2 + N - 2} \quad (1)$$

gde je

S_{ij} = ocena elementa i-tog reda i j-te kolone kovariacione matrice,

N_1 = broj observacija u Prvoj Grupi,

N_2 = broj observacija u Drugoj Grupi,

a = pojedinačne observacije iz prve grupe (a=1,...,N₁),

b = pojedinačne observacije iz druge grupe (b=1,...,N₂),

X_i = sredina i-te varijable za indiciranu grupu,

X_j = sredina j-te varijable za indiciranu grupu.

Varijable i podaci

Na osnovu kvantitativne analize podataka u 34 preduzeća utvrđen su sledeći inspekcijom naredne podaci o efikasnosti sistema internih kontrole i to u domenu, kako sledi:

- Kompletnost,

³⁴ Generalizacija problema se opredeljuje na tri ili više dimenzija, tj. N dimenzionalni slučaj.

- Tačnost,
- Greške u obradi, i
- Ažurnost.

Kao rezultata ove kvantitativne obrade utvrđeno je da **sistem efikasnosti internih kontrola** ima više varijacija vrednovanja i to: slabe, srednje i dobre interne kontrole. naš zadatak je utvrdimo koji od pomenutih atributa, gore pomenutih, dominantni u utvrđivanju efikasnosti internih kontrola i to na osnovu podataka rada u računskom centru klijenta revizije (Tabele 1 u Prilogu):

Eksperiment i rezultati analize

Ovaj primer je obradjen, kako smo napred rekli višestrukom linearnom diskriminacionom analizom sa svrhom da se utvrdi koji su faktori dominantni u opredeljenju kvalifikacije za efikasnost internih kontrola i to: između faktora : Kompletnost, Tačnost, Greške u obradi, i Ažurnost. U dalje, delu teksta opriimenićemo modele diskriminacione analize u sistemu nadziranog učenja.

Nadzirano učenje 1 (Linearna diskriminaciona analiza)

Rezultati				
Performanse klasifikovanja				
Stopa greške		0,1765		
Prediktovana vrednost		Matrica neodređenosti		
Vrednost	Ponavljanje obrade	1-Preciznost		
srednji	0,7273	0,2727	srednji	8
slab	0,9167	0,0833	slab	1
dobar	0,8182	0,1818	dobar	2
			Sum	11
			Sum	34

Sumarni prikaz LDA

Attribute	Klasifikacione funkcije			Statistička vrednovanja			
	srednji	slab	dobar	Wilks L.	Partial L.	F(2,28)	p-vrednost
Kompletnost %	26,786628	26,181011	27,315519	0,247511	0,841204	2,64281	0,088842
Tačnost %	-0,574329	-0,805217	-0,408257	0,251574	0,827618	2,91601	0,070733
Greške u obradi %	-2,241464	-2,198536	-2,215486	0,219069	0,950420	0,73034	0,490699
Ažurnost %	0,050860	0,162535	0,003938	0,248508	0,837829	2,70986	0,083980
constant	-1144,981548	-1080,738473	-1207,590253	-			

Poboljšanje procesa učenja 1

Ponavljanje : 25

Rezultati			
Poboljšanje procesa učenja estimacija greške			
Detalji			
Ponavljanja	Testna greška	Poboljšanje procesa učenja	Poboljšanje procesa učenja+
1	0,3300	0,2735	0,2862

Korak učenja 1

Rezultati

Rezultati selekcije
[2] selektirani atributi [4]
Selektirani podskup atributa

N°	Selektirani atributi
1	Kompletnost%
2	Tačnost %

Detaljni rezultati

N°	d.f	Najbolje	Reš..1	Reš..2	Reš..3	Reš..4	Reš..5
1	(2, 31)	Kompletnost% L : 0,361 F : 27,40 p : 0,0000	Kompletnost% L : 0,361 F : 27,40 p : 0,0000	Tačnost % L : 0,382 F : 25,05 p : 0,0000	Ažurnost % L : 0,647 F : 8,44 p : 0,0012	Greške u obradi % L : 0,649 F : 8,39 p : 0,0012	-
2	(2, 30)	Tačnost % L : 0,261 F : 5,80 p : 0,0074	Tačnost % L : 0,261 F : 5,80 p : 0,0074	Ažurnost % L : 0,280 F : 4,36 p : 0,0217	Greške u obradi % L : 0,322 F : 1,82 p : 0,1790	-	-
3	(2, 29)	-	Ažurnost % L : 0,219 F : 2,74 p : 0,0810	Greške u obradi % L : 0,249 F : 0,70 p : 0,5039	-	-	-

Nadzirano učenje 2 (Linearna diskriminaciona analiza)
Rezultati
Performanse klasifikovanja

Stopa greške			0,2941				
Prediktovana vrednost			Matrica neodredjenosti				
Vrednost	Recall	1-Preciznost		srednji	slab	dobar	Sum
srednji	0,3636	0,4286	srednji	4	4	3	11
slab	0,9167	0,2667	slab	1	11	0	12
dobar	0,8182	0,2500	dobar	2	0	9	11
			Sum	7	15	12	34

Sumarni prikaz LDA

Atributi	Klasifikacione funkcije			Statistička Evaluacija			
	srednji	slab	dobar	Wilks L.	Partial L.	F(2,30)	p-vrednost
Kompletnost%	13,560450	13,246824	14,226717	0,382223	0,681630	7,00607	0,003186
Tačnost %	1,226724	0,950406	1,376420	0,361342	0,721021	5,80384	0,007400
konstanta	-664,798921	-614,965602	-739,588990	-			

Poboljšanje procesa učenja 2
Ponavljjanja 25 puta
Rezultati
Detalji

Ponavljjanja	Testna greška	Poboljšanje procesa učenja	Poboljšanje procesa učenja+
1	0,3967	0,3589	0,3663

3. Zaključak

Prezentirani primer generiše diskriminacionu funkciju na bazi observacija sistem internih kontrola u preduzećima (privrednim društvima), dao je ohrabrujuće rezultate. Naravno, ovo je samo jedna uzorak, u slučaju više preduzeća razložno je da veći broj observacija daje signifikatnije rezultate.

Praktičan aspekt ove metode se u našem primeru bazira na uzimanjem uzoraka iz preduzeća na bazi obavljenih inspekcija internih kontrola od strane revizora.

Tačnost MLDA se može signifikatno povećati stratifikacijom uzoraka i drugim metodama uzorkovanja, čime, bez sumnje, ova metoda ima značajnu primenu u domenu onih problema koji imaju suštinsku karakteristiku, diskriminacije dve ili više grupa observacija, čime je ova metoda postala neophodna alatka za ispitivanje ovih klasa fenomene.

Precizno vrednujući dobijene rezultate možemo bez rezerve da zaključimo da je revizija primenom ovih rezultata dobila nadmoćan metodološki aparat, čija će dalja i dublja implementacija sigurno dati rezultate, koji će umnogome promeniti metodološki aparatikus revizije, i to u dva pravca: skraćivanje vremena obavljanja revizije i povećanje preciznosti dobijenih rezultata, kao i iskazanih revizorskih mišljenja o finansijskih izveštaja. Krajnji cilj primene ovih metoda jeste da se na bazi finansijskih izveštaja, primenom mašinskog učenja (data mining- pronalažanje skrivenih ili implicitnih znanja) neposredno obradom finansijskih izveštaja pruži precizna osnova za davanje revizorskog mišljenje.

LITERATURA

1. Wesley E. Harder, *"Multiple linear discriminant analysis"*, School of Business, Ca, USA, *Management Datamatics*, Vol. 4 (1975).
2. Edward I. Altman, *"Financial Ratios Discriminates Analysis, and Prediction of Corporation Bankruptcy"*, *The Journal of Finance*, 1968.
3. E.Kirkos, C. Spathis, Y. Manolopoulos, "Data Mining for the detection of fraudulent financial statements", *Expert Systems with Application*, 2007 str. 995-1003.
4. Koskivaara, E. (2004). "Artificial neural networks in analytical review procedures". *Managerial Auditing Journal*, 192-223.