

*Prof. dr. sc. Mirjana Pejić-Bach*  
*Mr. sc. Mirjana Demonja*

**OTKRIVANJE POREZNE UTAJE IZ INFORMACIJSKOG  
SUSTAVA POREZNE UPRAVE METODOM OTKRIVANJA  
ZNAJNA IZ BAZA PODATAKA**

**DATA MINING APPROACH TO DISCOVERING TAX EVASION  
FROM THE TAXATION OFFICE INFORMATION SYSTEM**

---

**SAŽETAK:** Pod pojmom porezna utaja obuhvaćeno je nezakonito izbjegavanje poreza kod čega porezni obveznik prema vlastitom sustavu vrijednosti određuje koliko će poreza platiti državi. Procesi otkrivanja porezne utaje često nisu dovoljno efikasni, a temelje se na subjektivnim procjenama zaposlenika poreznih uprava. Otkrivanje znanja iz baza podataka može se koristiti za efikasnije rješavanje problema koji se tradicionalno rješavaju isključivo temeljem znanja stručnjaka. Cilj rada je razviti model otkrivanja porezne evazije na temelju podataka iz informacijskog sustava porezne uprave. Model je razvijen korištenjem metode linearne regresije uz korištenje iznosa utajenog poreza pronađenog nadzorom u razdoblju od 2001. do 2005. godine kao zavisne varijable i nezavisnih varijabli: dohodak vlasnika, ukupan prihod, kupljene poslovne i stambene nekretnine.

**KLJUČNE RIJEČI:** porezna utaja, otkrivanje znanja iz baza podataka, linearna regresija, informacijski sustav.

**ABSTRACT:** Tax evasion includes illegal avoidance of paying tax, where tax tributary determines by herself own value system how much tax will she pay to the government. Discovering tax evasion processes are not often efficient, and are based on subjective estimation of the taxation office employees. Data mining approach could be used for more efficient solving of the problems that are traditionally solved based solely on the expert knowledge. Goal of the paper is to develop the model of discovering tax evasion based on the taxation office information system. Linear regression model is developed that could be used for discovering tax evasion based on the data collected from the year 2001 till year 2005. Amount of the tax evaded is used as the dependent variable, while independent variables are: owner income, total revenue, and business and residential real estate bought.

**KEY WORDS:** tax evasion, data mining, linear regression, information system

---

## 1. UVOD

Saznanje o golemoj količini podataka spremljenih u bazama podataka dovodi do potrebe njihova korištenja i povezivanja s izradom modela za otkrivanje znanja iz baza podataka, posebice u financijskim institucijama (Zhang et al., 2004). Ova nova metoda stvaranja znanja često se koristi upravo za otkrivanje prijevара u različitim područjima (Weatherford, 2002). Yang i Hwang (2006) koriste otkrivanje znanja iz baza podataka za detekciju prijevара i zloraba u zdravstvenom osiguranju. Kirkos (2007) je razvio ekspertni sustav za otkrivanje krivotvorenih financijskih izvješća. Otkrivanje prijevара korištenjem kreditnih kartica posebice je kompleksno zbog korištenja kreditnih kartica na brojnim prodajnim mjestima. Chan i drugi (1999) razvili su on-line sustav za otkrivanje takvih prijevара, a sustav se dalje razrađuje i nalazi brojne primjene u praksi (Hand et al, 2008). Metode otkrivanja znanja iz baza podataka koriste se također u telekomunikacijama (Xing i Girolami, 2007; Daskalaki, 2003), za otkrivanje prijevара osiguravajućih društava (Vieaene et al, 2004; Morley i drugi, 2006).

Qifengs et al. (2008) istražuju primjenu sustava za prevenciju porezne utaje u on-line plaćanju. Pisani i DeSisti (2008) za otkrivanje porezne utaje koriste Bayesove mreže. Utaja poreza na dobit također se otkriva korištenjem inteligentnih metoda (Chun-Nan et al., 2007). U Sloveniji su simulacijski modeli korišteni za poreznu reformu (Čok et al., 2008).

Prema izvorima literature, inteligentne metode nisu korištene u svrhu otkrivanja porezne utaje u hrvatskim državnim institucijama. Međutim, Informacijski sustav Porezne uprave Republike Hrvatske prikuplja brojne podatke koji bi se mogli koristiti upravo u te svrhe. Primjerice, u Informacijskom sustavu Porezne uprave postoje podatci o ukupnim prihodima, ukupnim troškovima, iskazanim dohodcima i dobiti, odnosno gubitcima. Osim toga, prikupljaju se podatci o broju zaposlenih, isplaćenim plaćama za koje su obračunati i uplaćeni porezi i doprinosi. Međutim, dostupni su i podatci o kupljenim automobilima, jahtama, zemljištu, stambenim i poslovnim prostorima. Pri tome za sve nabave mora postojati transparentan izvor financiranja ili u prikazanoj dobiti, odnosno dohotku ili u kreditima. Nerazmjernost između prikazanog dohotka ili dobiti i trošenja uočljiv je za mnoge porezne obveznike.

Ovakvi su podatci poglavito korisni za otkrivanje porezne utaje iz mnogih razloga. Korištenjem postojećih podataka i njihovim sagledavanjem izbjeglo bi se narušavanje privatnosti poreznih obveznika, izbjeglo bi se uznemiravanje poreznih obveznika uvođenjem novih obrazaca te bi se izbjegao povećani opseg inspeksijskog nadzora. Posebice bi bilo korisno pregledavati podatke kroz dulje razdoblje. Drugačija se slika stječe analizom podataka o poreznim obveznicima kroz jednu, a drugačija kroz više godina, primjerice petogodišnje razdoblje.

Iz svega navedenog proizlazi cilj rada koji se očituje u ispitivanju pretpostavke o korištenju metoda otkrivanja znanja iz baza podataka u Informacijskom sustavu Porezne uprave Republike Hrvatske. Model je razvijen korištenjem podataka o dvjesto obveznika – pravnih osoba. Rad se sastoji od pet dijelova. Nakon uvodnog dijela rada, u drugom poglavlju razmatra se problematika porezne evazije, a u trećem se dijelu ukratko prikazuje metodologija otkrivanja znanja iz baza podataka. Četvrti dio sadrži prikaz razvijenog modela, pri čemu su opisane karakteristike uzorka, te je opisan razvoj modela metodom jednostavne linearne regresije. U posljednjem dijelu rada sadržana su zaključna razmatranja.

## 2. POREZNA UTAJA

Porezna utaja (engl. *tax evasion*) obuhvaća nezakonite oblike smanjenja poreznih obveza, a to je nezakonito izbjegavanje poreza kod čega porezni obveznik prema vlastitom sustavu vrijednosti određuje koliko će poreza platiti državi (Mađarević-Šujster, 2002). Siva ekonomija tako guta dio društvenog proizvoda, odnosno dio poslovnih aktivnosti odvija se izvan službenog gospodarstva što opterećuje porezne obveznike koji plaćaju porez (Tanzi, 1983).

Otpor plaćanju poreza proizlazi iz mnogo činjenica i ne ovisi samo o karakteristikama poreznih obveznika nego ovisi i o poreznom opterećenju i o pojedinim vrstama poreza. Plaćanje poreza smanjuje ekonomsku snagu poreznog obveznika (Vukšić, 2003). Međutim, savjest o obvezi plaćanja ili neplaćanja poreza ovisi o tradiciji, obrazovnoj strukturi stanovništva, izvršavanju sankcija prema neplatišama poreza u nekoj sredini, stručnosti poreznih tijela i slično. Nezakonito izbjegavanje plaćanja poreza može se pojaviti kao aktivno izbjegavanje plaćanja poreza kad porezni obveznik namjerno netočno prikazuje ili prikriva važne činjenice za utvrđivanje poreza. Nezakonita porezna utaja kod koje se na nezakonit način izbjegava plaćanje neposrednih poreza obilježava se terminom defraudacija poreza (*defraudare* lat. varati).

Pojavni oblici nezakonitog izbjegavanja poreza su mnogobrojni i ovise kako o poreznom opterećenju tako i o vrsti poreza. Razlikujemo nezakonitu poreznu utaju neposrednih i posrednih poreza. Neposredni porezi su, na primjer, porez na dohodak, porez na dobit, općinski, gradski i županijski porezi, a porezna utaja je postupak kojim se na nezakonit način izbjegava plaćanje posrednih poreza i poznata je po nazivima krijumčarenje, šverc, kontra-banda. Posredni porezi su na primjer porez na dodanu vrijednost i trošarine.

Porezna evazija razlikuje se od izbjegavanja poreza (engl. *tax avoidance*) što podrazumijeva zakonom dopušteno smanjenje porezne obveze upotrebom svih zakonom dopuštenih mogućnosti. Drugim riječima, svatko mora platiti porez samo onoliko koliko određuje zakon i ne može biti kažnjavan dok koristi zakonske mogućnosti, pa makar to bili i nedostatci sustava. Prepoznavanjem problema i poduzimanjem odgovarajućih mjera smanjilo bi se porezno opterećenje za većinu poreznih obveznika koji nisu u mogućnosti izbjegavati porezne obveze, a to su poglavito građani koji nisu u mogućnosti manipulirati svojom poreznom osnovicom.

## 3. OTKRIVANJE ZNANJA IZ BAZA PODATAKA

Ključni resursi postindustrijskog društva su informacija i znanje. Informatizacija za pojedinca i poslovnoorganizacijski sustav mora zadovoljiti potrebe za informacijama i znanjima zbog podizanje kvalitete odlučivanja i upravljanja. Baze podataka postaju važan faktor u donošenju odluka. Tako i baze podataka državne uprave postaju statistički izvor i važan čimbenik odlučivanja.

Proces rudarenja podacima odvija se u nekoliko faza (koraka): (1) definiranje problema, odnosno razumijevanje problema, (2) priprema podataka, (3) modeliranje i (4) evaluacija i implementacija (Pejić Bach, 2005).

### *Definiranje problema*

U procesu obrade podataka uvijek se mora misliti na cilj obrade podataka, na problem koji se želi riješiti. U rješavanju problema treba razumjeti ograničenja u rješavanju problema i otkriti čimbenike važne za rezultat obrade podataka. Uz definiranje problema važno je odrediti kriterij uspješnosti obrade podataka. Rezultati moraju sadržavati novi i korisni uvid u razumijevanje problema. Potrebno je ocijeniti praktične aspekte procesa obrade podataka: (1) odgovarajuće znanje o problemu, (2) sadrže li podatci dovoljno informacija, (3) definiranje specifične terminologije i (4) procjenu trajanja, koštanja i procjenu opravdanosti projekta

Važno je odrediti ciljeve obrade podataka, jer oni se razlikuju od ciljeva iz perspektive domene problema. Cilj rada iz domene problema je skrenuti pozornost na mogućnosti korištenja podataka koji već postoje u "spremišta podataka" i koje je zbog osjetljivosti teme ili zbog većih troškova nepoželjno posebno prikupljati. Cilj rada iz domene obrade podataka je odrediti bitna svojstva iz već postojećih podataka iz "spremišta podataka".

### *Priprema podataka*

Centralni dio procesa obrade podataka su podatci (Inmon, 2002). Uvodnim, odnosno pripremnim prikupljanjem podataka stvara se popis dostupnih podataka, njihova lokacija, metode prikupljanja i problemi pri prikupljanju podataka. Kao osnovna karakteristika podataka, stvara se opis podataka koji sadrži popis varijabli i atributa, identitet i značenje varijabli, format, odnosno vrijednost atributa. Kod istraživanja osnovnih karakteristika podataka cilj je provjeriti sadrže li podatci u svojoj strukturi kvalitetne informacije pomoću kojih bi se pristupilo rješavanju problema. Verifikacija kvalitete podataka provodi se preko određivanja konzistentnosti podataka prema vrijednosti i tipu atributa (nominalni, kategorički, numerički), pomoću određivanja količine i raspodjele primjera s neodređenim vrijednostima pojedinih atributa, preko otkrivanja primjera koji odskaču, koji su posebni (ili grješke ili rijedak fenomen), koje treba ili ne treba isključiti iz podataka prije procesa modeliranja, netipične primjere, odnosno prljave, netočne ili nedostatne podatke.

Priprema podataka za proces modeliranja odvija se u četiri faze (Han i Kamber, 2000): **odabir podataka** - odabir podskupa prikupljenih podataka na bazi kriterija kvalitete podataka najčešće metodom slučajnog uzorka; **čišćenje podataka** - pomoću mnogih tehnika i može biti vremenski dugotrajno, odnosi se na normaliziranje podataka, "glađenje" podataka, **formiranje novih atributa ili primjera** - konstrukcija novih podataka iz postojećih podataka, transformacijom podataka i spajanjem podataka iz različitih tablica stvaranjem novih atributa i **formatiranje podataka** u format prema odabranoj tehnici modeliranja, što znači promjene u redosljedu atributa (primjera) i promjene koje se odnose na sređivanje podataka s obzirom na sintaksu. Podatke je potrebno transformirati u standardni format koji može biti **binarni** (1-znači da atribut poprima neku vrijednost, a 0-znači da atribut ne poprima neku vrijednost) ili **numerički** (atribut ima brojčanu vrijednost).

### *Modeliranje*

Inteligentna obrada podataka podrazumijeva mnoštvo tehnika modeliranja. Tehnika modeliranja odabire se nakon definiranja skupa podataka, ovisno o zadatku obrade podataka, odnosno o tipu problema. Odabir tehnike ovisi o cilju analize i uobičajeno se provjerava više metoda. Metode rudarenja primjenjuju se pomoću specijaliziranog softvera, kao što je Statistica Data Miner, SAS Enterprise Miner ili SPSS. Tehnike modeliranja su na primjer **stabla odlučivanja** (deskriptivna tehnika modeliranja podataka kojom se može

opisivati određena grupa primjera), **neuronska mreža** (kompleksna nelinearna tehnika modeliranja stvorena po uzoru na ljudske neurone kojom je, upotrebljavajući velik broj parametara linearnim kombinacijama i transformacijama, moguće aproksimirati buduće vrijednosti zavisne varijable), te **klaster analiza** (postupak kod čega se segmentiraju heterogeni podatci u nekoliko homogenih skupina ili razreda).

#### *Evaluacija i implementacija*

Evaluacija modela, odnosno ocjena otkrivenog znanja ključna je komponenta procesa modeliranja (Berry i Linoff, 1997). Podatci u uzorku se dijele na dva dijela, podatke za izradu modela (podatci o poduzećima kod kojih nije bilo nadzora) i podatke za testiranje modela (podatci o poduzećima kod kojih je izvršen nadzor) koji se koriste za vrjednovanje efikasnosti metode.

U inteligentnoj analizi podataka je klasifikacija najčešći problem. Pri evaluaciji klasifikacijskih modela osnovni pojam je pojam grješke. Važno je razlikovanje grješaka po tipovima, što se utvrđuje pomoću "matrice grješaka". Matrica grješaka daje broj točno klasificiranih primjera. Senzitivnost i specifičnost opisuju kvalitetu modela. Mjera kvalitete modela provodi se metodom kojom se određuju granične vrijednosti pouzdanosti određenog klasifikatora.

Za evaluaciju modela također je važna jednostavnost rezultata modela za interpretaciju, kao i mogućnost korištenja dobivenih rezultata u rješavanju problema kojim se bavi (npr. generiranjem pravila koja se implementiraju u informacijski sustav, a koja ukazuju poreznom referentu da za pojedinog poreznog obveznika postoji veća vjerojatnost porezne utaje).

## **4. EMPIRIJSKO ISTRAŽIVANJE**

### **4.1. Opis problema, metodologija i opis uzorka**

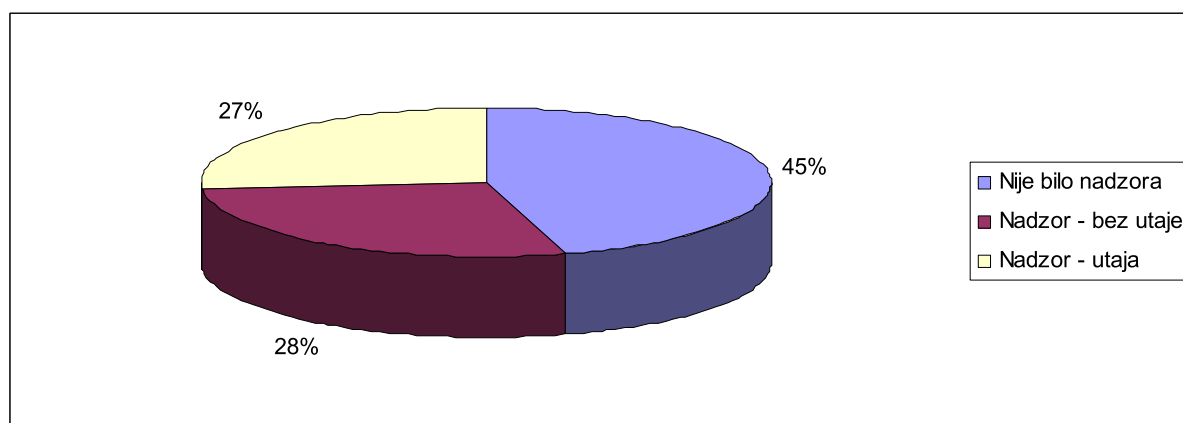
Cilj rada je razviti model otkrivanja porezne evazije na temelju podataka iz informacijskog sustava porezne uprave. Baze podataka porezne uprave vrijedan su izvor podataka na temelju kojih bi se brže i učinkovitije otkrivala porezne evazije, pri čemu bi se koristile metode otkrivanja znanja iz baza podataka. Slučajnim izborom odabrano je u uzorak 200 poduzeća. Podatci o poduzećima prikupljeni su za razdoblje od 2000. do 2005. godine. Analizirane su sljedeće varijable: (1) vlasnik pravna osoba ili nije, (2) vlasnik zaposlen u poduzeću ili nije, (3) vlasnik stranac ili nije, (4) spol vlasnika, (5) dohodak vlasnika od 2000. do 2005. godine, (6) ukupan prihod poduzeća od 2001. do 2005. godine, (7) dobit/gubitak poduzeća od 2001. do 2005. godine, (8) kupljeni automobili od 2000. do 2005. godine za pravnu osobu, (9) kupljene stambene nekretnine od 2001. do 2005. godine, pri čemu se pod stambenim nekretninama razumijevaju stanovi, kuće i zemljišta, (10) kupljene poslovne nekretnine od 2001. do 2005. godine i (11) iznos utajenog poreza pronađen nadzorom u razdoblju od 2001. do 2005. godine.

Na temelju ovih varijabli izvedene su sljedeće pomoćne varijable: (1) gubitak – označuje je li u barem jednoj od promatranih godina ostvaren gubitak, (2) ukupan broj kupljenih automobila u razdoblju od 2001. do 2005. godine, (3) ukupan broj stambenih nekretnina kupljenih u razdoblju od 2001. do 2005. godine, (4) ukupan broj poslovnih nekretnina kup-

ljenih u razdoblju od 2001. do 2005. godine, (5) ukupan broj stambenih i poslovnih nekretnina zajedno kupljenih u razdoblju od 2001. do 2005. godine, (6) ukupan iznos naplaćenog utajenog poreza u razdoblju od 2001. do 2005. godine.

U uzorak je odabrano 91 poduzeće nad kojima nije vršen nadzor s ciljem pronalaženja porezne utaje te 109 poduzeća nad kojima je vršen nadzor u barem jednoj od promatranih godina, što znači da je u većem dijelu poduzeća iz uzorka bar jednom u promatranom razdoblju izvršen nadzor. Od 109 poduzeća nad kojima je izvršen nadzor u 53 poduzeća, pronađena je porezna utaja (slika 1), odnosno utaja je pronađena u 53 poduzeća od 200 poduzeća u uzorku. U većem broju poduzeća kod kojih je izvršen nadzor nije bilo utaje.

**Slika 1.** Struktura poduzeća u uzorku s obzirom na nadzor i utaju



*Izvor: Izvorno istraživanje*

Nakon odabira i čišćenja podataka obrađeni su podatci s obzirom na karakteristike poduzeća u odnosu na poreznu utaju. Prihod poduzeća je bio u porastu iz godine u godinu, da bi se u promatranom razdoblju od pet godina gotovo udvostručio. Dobit nije tako linearno rasla, ali je u promatranom razdoblju više nego udvostručena. Nabava automobila u promatranom razdoblju nije povezana niti s porastom ukupnih prihoda niti s iskazivanjem dobiti, odnosno gubitka. Pravilan rast ukupnog prihoda ne prati rast kupljenih poslovnih nekretnina, niti je kupovanje poslovnih nekretnina utjecalo na ne-iskazivanje dobiti prema rastu ukupnih prihoda. U promatranom razdoblju je kupovano više stambenih nekretnina, nego poslovnih. Kupovanje stambenih nekretnina je u određenim godinama višestruko povećano. Korištenjem hi-kvadrata i Anova analize izdvojene su one varijable koje su statistički povezane s činjenicom je li nadzorom pronađena porezna utaja kod poreznog obveznika. Te varijable korištene su u razvoju modela metodom linearne regresije, koji će biti prikazani u nastavku.

## 4.2. Razvoj modela metodom jednostavne linearne regresije

Metoda jednostavne linearne regresije prikladna je za izradu takvih prognostičkih metoda zbog nekoliko svojih karakteristika: jednostavno i brzo izvođenje algoritma, lako tumačenje rezultata te mogućnost korištenja regresijske varijable za prognožiranje zavisne varijable.

Da bi se model razvio, korištena je regresijska analiza "postupnog razvoja modela unazad" (engl. *backward regression*). Takva analiza započinje korištenjem jedne zavisne varijable i skupa nezavisnih varijabli, pri čemu se u svakom koraku izbacuju one varijable koje ne prolaze test tolerancije (engl. *tolerance test*). U regresijskoj analizi se test tolerancije provodi temeljem pokazatelja tolerancije koji određuje u kojoj su mjeri nezavisne varijable međusobno povezane, tj. u kojoj mjeri je prisutna multikolinearnost. Računa se kao razlika jedinične vrijednosti i koeficijenta determinante modela u kojem je testirana nezavisna varijabla postavljena kao zavisna, a sve ostale nezavisne varijable tretirane su kao nezavisne. U sljedećem koraku se ponavlja regresijska analiza korištenjem preostalih varijabli. Postupak se ponavlja dok se model ne sastavi od onog skupa nezavisnih varijabli koje sve prolaze test tolerancije.

Iznos utajenog poreza pronađen nadzorom u razdoblju od 2001. do 2005. godine odabran je za zavisnu varijablu. Nezavisne varijable odabrane su na temelju njihove statističke značajnosti pronađene korištenjem hi-kvadrat testa i Anova analize pri čemu je postojanje porezne utaje (nije bilo nadzora, nadzor bez utaje, nadzor sa utajom) korišteno kao varijabla grupiranja. Analiza je započeta korištenjem sljedećih nezavisnih varijabli: (1) dohodak vlasnika 2005., (2) ukupan prihod 2002.-2005., (3) poslovne nekretnine 2005., (4) stambene nekretnine 2001.-2002., 2004.-2005., (5) ukupne stambene nekretnine, (6) nekretnine kupljene ukupno. Procedura "postupnog razvoja modela unazad" opisuje se u nastavku ovoga dijela rada.

#### 4.2.1. Prvi korak "postupnog razvoja modela unazad"

U prvom koraku "postupnog razvoja modela unazad" korištene su sve prethodno navedene nezavisne varijable. Test tolerancije pokazao je da varijabla "ukupan prihod prosječno" ne prolazi test tolerancije, jer ima razinu značajnosti veću od 0,05 (tablica 2), stoga je navedena varijabla isključena iz daljnje analize. Preostale varijable korištene su za procjenu linearnog regresijskog modela "Model 1".

**Tablica 1.** Test značajnosti nezavisnih varijabli regresijskog modela "Model 1"

Model 1	Nestandardizirani koeficijenti	Standardna pogreška	Standardizirani koeficijenti	t	P-vrijednost
Konstanta	23001,251	238661,715		0,096	0,924
Dohodak vlasnika 2005	-2,138	2,795	-,091	-,765	0,452
Ukupan prihod 2002	5,938E-02	0,045	1,584	1,331	0,197
Ukupan prihod 2003	0,109	0,065	2,379	1,678	0,107
Ukupan prihod 2004	1,528E-04	0,038	0,003	0,004	0,997
Ukupan prihod 2005	0,126	0,027	3,967	4,680	0,000**
Poslovne nekretnine 2005	330199,907	1000266,553	0,032	0,330	0,744
Stambene nekretnine 2001	4478671,255	1331196,800	0,729	3,364	0,003**
Stambene nekretnine 2002	287360,059	559636,921	0,053	0,513	0,613
Stambene nekretnine 2004	3419304,088	949668,999	2,037	3,601	0,002**
Stambene nekretnine 2005	4640479,870	801218,881	1,625	5,792	0,000**
Stambene nekretnine ukupne	4653755,699	1424193,913	4,067	3,268	0,004**
Nekretnine kupljene ukupno	1430013,988	526285,334	1,824	2,717	0,013*

Izvor: izvorno istraživanje

a Zavisna varijabla: Ukupno naplaćena porezna utaja

\* statistički značajno uz 5% vjerojatnosti

\*\* statistički značajno uz 1% vjerojatnosti

Tablica 1 prikazuje rezultate testa značajnosti nezavisnih varijabli navedenog regresijskog modela, te se pokazalo da su statistički značajne varijable "ukupan prihod 2005", "stambene nekretnine 2001", "stambene nekretnine 2004", "stambene nekretnine 2005", "stambene nekretnine ukupne", te "nekretnine kupljene ukupno". Analiza se nastavlja korištenjem svih varijabli, osim onih isključenih temeljem testa tolerancije.

**Tablica 2.** Test tolerancije za isključenu varijablu regresijskog modela "Model 2"

	Beta In	t	Značajnost	Parcijalna korelacija	Kolinearnost (tolerancija)
Ukupan prihod prosječno	25,873	3,021	0,007	0,550	6,019E-05

*Izvor: Izvorno istraživanje.*

Ukupan prihod prosječno ne prolazi test tolerancije, jer ima razinu značajnosti veću od 0,05.

#### 4.2.2. Drugi korak "postupnog razvoja modela unazad"

U drugom koraku korištene su sve prethodno navedene nezavisne varijable, osim "prosječnog ukupnog prihoda", koji je izostavljen iz analize zbog prethodno navedenih razloga. Ponovo je proveden test tolerancije koji je pokazao da varijabla "ukupan prihod 2004" ne prolazi test tolerancije, jer ima razinu značajnosti veću od 0,05 (tablica 4). Navedena varijabla isključena je iz daljnje analize, a preostale varijable korištene su za procjenu linearnog regresijskog modela "Model 2". Tablica 3 prikazuje rezultate testa značajnosti nezavisnih varijabli navedenog regresijskog modela, te se pokazalo da su statistički značajne varijable "ukupan prihod 2003", "ukupan prihod 2005", "stambene nekretnine 2001", "stambene nekretnine 2004", "stambene nekretnine 2005", "stambene nekretnine ukupne", te "nekretnine kupljene ukupno". Analiza će se nastaviti korištenjem svih varijabli osim onih isključenih temeljem testa tolerancije.

**Tablica 3.** Test značajnosti nezavisnih varijabli regresijskog modela "Model 2"

Model 2	Nestandardizirani koeficijenti	Standardna pogreška	Standardizirani koeficijenti	t	P-vrijednost
Konstanta	23234,350	226410,662		0,103	0,919
Dohodak vlasnika 2005	-2,138	2,733	-0,091	-0,782	0,442
Ukupan prihod 2002	5,928E-02	,036	1,581	1,629	0,117
Ukupan prihod 2003	,109	,048	2,383	2,283	0,032*
Ukupan prihod 2005	,126	,022	3,965	5,757	0,000**
Poslovne nekretnine 2005	329411,513	959262,278	0,032	0,343	0,734
Stambene nekretnine 2001	4479201,700	1295514,700	0,730	3,457	0,002**
Stambene nekretnine 2002	287027,977	541330,719	0,053	0,530	0,601
Stambene nekretnine 2004	3419059,955	926890,880	2,037	3,689	0,001**
Stambene nekretnine 2005	4640077,575	777461,730	1,625	5,968	0,000**
Stambene nekretnine ukupne	4653265,090	1387757,999	4,067	3,353	0,003**
Nekretnine kupljene ukupno	1429769,016	511249,805	1,823	2,797	0,010*

*Izvor: Izvorno istraživanje.*

a Zavisna varijabla: Ukupno naplaćena porezna utaja

\* statistički značajno uz 5% vjerojatnosti

\*\* statistički značajno uz 1% vjerojatnosti



**Tablica 4.** Test tolerancije za isključenu varijablu regresijskog modela "Model 2"

	Beta In	t	Značajnost	Parcijalna korelacija	Kolinearnost (tolerancija)
Ukupan prihod prosječno	3,188	0,907	0,373	0,182	4,407E-04
Ukupan prihod 2004	0,089	0,124	0,902	0,025	1,095E-02

*Izvor: Izvorno istraživanje.*

"Ukupan prihod prosječno" i "ukupan prihod 2004" ne prolaze test tolerancije jer imaju razinu značajnosti veću od 0,05.

#### 4.2.3. Treći korak "postupnog razvoja modela unazad"

U trećem koraku korištene su sve prethodno navedene nezavisne varijable, osim varijabla "prosječni ukupni prihod" i "ukupan prihod 2004" koje su izostavljene iz analize zbog testa tolerancije. Ponovo je proveden test tolerancije koji je pokazao da varijabla "poslovne nekretnine 2005" ne prolazi test tolerancije, jer ima razinu značajnosti veću od 0,05 (tablica 6). Navedena varijabla isključena je iz daljnje analize, a preostale varijable korištene su za procjenu linearnog regresijskog modela "Model 3". Tablica 5 prikazuje rezultate testa značajnosti nezavisnih varijabli navedenog regresijskog modela te se pokazalo da su statistički značajne varijable "ukupan prihod 2003", "ukupan prihod 2005", "stambene nekretnine 2001", "stambene nekretnine 2004", "stambene nekretnine 2005", "stambene nekretnine ukupne" te "nekretnine kupljene ukupno". Analiza se nastavlja korištenjem svih varijabli osim onih isključenih temeljem testa tolerancije.

**Tablica 5.** Test značajnosti nezavisnih varijabli regresijskog modela "Model 3"

Model 3	Nestandardizirani koeficijenti	Standardna pogreška	Standardizirani koeficijenti	t	P-vrijednost
Konstanta	17339,328	221571,420		0,078	0,938
Dohodak vlasnika 2005	-2,179	2,680	-0,093	-0,813	0,424
Ukupan prihod 2002	6,082E-02	0,035	1,622	1,716	0,099
Ukupan prihod 2003	0,109	0,047	2,372	2,317	0,029*
Ukupan prihod 2005	0,127	0,021	3,996	5,963	0,000**
Stambene nekretnine 2001	4335933,361	1203748,786	0,706	3,602	0,001**
Stambene nekretnine 2002	263549,616	527035,262	0,049	0,500	0,622
Stambene nekretnine 2004	3272001,176	806783,640	1,949	4,056	0,000**
Stambene nekretnine 2005	4521192,004	683214,201	1,584	6,618	0,000**
Stambene nekretnine ukupne	4410375,764	1171834,561	3,854	3,764	0,001**
Nekretnine kupljene ukupno	1336111,757	424405,127	1,704	3,148	0,004**

*Izvor: Izvorno istraživanje.*

a Zavisna varijabla: Ukupno naplaćena porezna utaja

\* statistički značajno uz 5% vjerojatnosti

\*\* statistički značajno uz 1% vjerojatnosti

**Tablica 6.** Test tolerancije za isključenu varijablu regresijskog modela "Model 3"

	Beta In	t	Značajnost	Parcijalna korelacija	Kolinearnost (tolerancija)
Ukupan prihod prosječno	3,188	0,907	0,373	0,182	4,407E-04
Ukupan prihod 2004	0,089	0,124	0,902	0,025	1,095E-02
Poslovne nekretnine 2005	0,026	0,283	0,780	0,058	0,680

*Izvor: izvorno istraživanje*

"Ukupan prihod prosječno", "ukupan prihod 2004" i "poslovne nekretnine 2005" ne prolaze test tolerancije, jer imaju razinu značajnosti veću od 0,05.

#### 4.2.4. Četvrti korak "postupnog razvoja modela unazad"

U četvrtom koraku korištene su sve prethodno navedene nezavisne varijable, osim varijabla "prosječni ukupni prihod", "ukupan prihod 2004" i "poslovne nekretnine 2005". Ponovo je proveden test tolerancije koji je pokazao da varijabla "stambene nekretnine 2002" ne prolazi test tolerancije jer ima razinu značajnosti veću od 0,05 (Tablica 8). Navedena varijabla isključena je iz daljnje analize, a preostale varijable korištene su za procjenu linearnog regresijskog modela "Model 4". Tablica 7 prikazuje rezultate testa značajnosti nezavisnih varijabli navedenog regresijskog modela te se pokazalo da su statistički značajne varijable "ukupan prihod 2003", "ukupan prihod 2005", "stambene nekretnine 2001", "stambene nekretnine 2004", "stambene nekretnine 2005", "stambene nekretnine ukupne" te "nekretnine kupljene ukupno". Analiza se nastavlja korištenjem svih varijabli, osim onih isključenih temeljem testa tolerancije.

**Tablica 7.** Test značajnosti nezavisnih varijabli regresijskog modela "Model 4"

Model 4	Nestandardizirani koeficijenti	Standardna pogriješka	Standardizirani koeficijenti	t	P-vrijednost
Konstanta	28620,817	217088,691		0,132	0,896
Dohodak vlasnika 2005	-2,017	2,620	-0,086	-0,770	0,449
Ukupan prihod 2002	6,258E-02	0,035	1,669	1,802	0,084
Ukupan prihod 2003	0,112	0,046	2,439	2,441	0,022*
Ukupan prihod 2005	0,131	0,020	4,111	6,625	0,000**
Stambene nekretnine 2001	4258155,706	1175618,318	0,694	3,622	0,001**
Stambene nekretnine 2004	3154838,009	760345,747	1,879	4,149	0,000**
Stambene nekretnine 2005	4382514,768	614963,018	1,535	7,126	0,000**
Stambene nekretnine ukupne	4254452,473	1112517,642	3,718	3,824	0,001**
Nekretnine kupljene ukupno	1290840,737	408370,595	1,646	3,161	0,004**

*Izvor: Izvorno istraživanje.*

a Zavisna varijabla: Ukupno naplaćena porezna utaja

\* statistički značajno uz 5% vjerojatnosti

\*\* statistički značajno uz 1% vjerojatnosti

**Tablica 8.** Test tolerancije za isključenu varijablu regresijskog modela "Model 4"

	Beta In	t	Značajnost	Parcijalna korelacija	Kolinearnost (tolerancija)
Ukupan prihod prosječno	3,188	0,907	0,373	0,182	4,407E-04
Ukupan prihod 2004	0,089	0,124	0,902	0,025	1,095E-02
Poslovne nekretnine 2005	0,026	0,283	0,780	0,058	0,680
Stambene nekretnine 2002	0,049	0,500	0,622	0,102	0,585

*Izvor: Izvorno istraživanje.*

"Ukupan prihod prosječno", "ukupan prihod 2004", "poslovne nekretnine 2005" i "stambene nekretnine 2002" ne prolaze test tolerancije jer imaju razinu značajnosti veću od 0,05.

#### 4.2.3. Peti korak "postupnog razvoja modela unazad"

U posljednjem, petom koraku korištene su sve prethodno navedene nezavisne varijable, osim "prosječnog ukupnog prihoda" te varijabli "ukupan prihod 2004", "poslovne nekretnine 2005" i "stambene nekretnine 2002". Ponovo je proveden test tolerancije koji je pokazao da varijabla "Dohodak vlasnika" ne prolazi test tolerancije jer ima razinu značajnosti veću od 0,05 (tablica 10). Navedena varijabla isključena je iz daljnje analize, a preostale varijable korištene su za procjenu linearnog regresijskog modela "Model 5". Tablica 9 prikazuje rezultate testa značajnosti nezavisnih varijabli navedenog regresijskog modela te se pokazalo da su statistički značajne varijable: "ukupan prihod 2002", "ukupan prihod 2003", "ukupan prihod 2005", "stambene nekretnine 2001", "stambene nekretnine 2004", "stambene nekretnine 2005", "stambene nekretnine ukupne" te "nekretnine kupljene ukupno".

**Tablica 9.** Test značajnosti nezavisnih varijabli regresijskog modela "Model 5"

Model 5	Nestandardizirani koeficijenti	Standardna pogriješka	Standardizirani koeficijenti	t	P-vrijednost
Konstanta	-88100,618	154126,899		-0,572	0,572
Ukupan prihod 2002	8,024E-02	0,026	2,140	3,103	0,005**
Ukupan prihod 2003	8,810E-02	0,034	1,925	2,610	0,015*
Ukupan prihod 2005	0,129	0,019	4,067	6,634	0,000**
Stambene nekretnine 2001	4113802,749	1151433,152	0,670	3,573	0,001**
Stambene nekretnine 2004	3011132,280	731270,038	1,794	4,118	0,000**
Stambene nekretnine 2005	4407270,468	609289,793	1,544	7,233	0,000**
Stambene nekretnine ukupne	4107003,694	1087279,885	3,589	3,777	0,001**
Nekretnine kupljene ukupno	1252724,744	402168,041	1,598	3,115	0,004**

*Izvor: Izvorno istraživanje.*

a Zavisna varijabla: Ukupno naplaćena porezna utaja

\* statistički značajno uz 5% vjerojatnosti

\*\* statistički značajno uz 1% vjerojatnosti

**Tablica 10.** Test tolerancije za isključenu varijablu regresijskog modela "Model 5"

	Beta In	t	Značajnost	Parcijalna korelacija	Kolinearnost (tolerancija)
Ukupan prihod prosječno	3,240	0,930	0,361	0,183	4,409E-04
Ukupan prihod 2004	0,081	0,114	0,910	0,023	1,095E-02
Poslovne nekretnine 2005	0,030	0,330	0,744	0,066	0,682
Stambene nekretnine 2002	0,039	0,407	0,687	0,081	0,594
Dohodak vlasnika 2005	0,086	0,770	0,449	0,152	0,435

*Izvor: Izvorno istraživanje.*

"Ukupan prihod prosječno", "ukupan prihod 2004", "poslovne nekretnine 2005", "stambene nekretnine 2002" i "dohodak vlasnika 2005" ne prolaze test tolerancije jer imaju razinu značajnosti veću od 0,05.

Tabela 11. prikazuje koeficijente korelacije i determinacije regresijskih modela "Model 1" do "Model 5". Premda "Model 1" ima najviši koeficijent korelacije te koeficijent determinacije, "Model 5" ima najviši prilagođen koeficijent determinacije koji korigira koeficijent determinacije za broj nezavisnih varijabli, čime se umanjuje tendencija metode najmanjih kvadrata da se koeficijent determinacije povećava brojem nezavisnih varijabli<sup>1</sup>. Koeficijent korelacije neznatno je manji za "Model 5" u odnosu na "Model 1". Može se zaključiti da je "Model 5" prema većini odabranih kriterija (tolerancija, koeficijent determinacije) najbolji te se predlaže njegovo korištenje za otkrivanje poreznih evazija na temelju podataka iz Informacijskog sustava Porezne uprave u Republici Hrvatskoj.

Informacijski sustav porezne uprave raspolaže podacima o kupnji nekretnina na cijelom području Republike Hrvatske. U podacima o kupnji nekretnine nalazi se i podatak o izvoru financiranja za određenu nekretninu (kredit, osobna sredstva – uštedjevina, nasljedstvo).

Preko podataka u ISPU lako je povezati koliko je poreza na dohodak i poreza na dobit, kao vlasnik poduzeća ili kao građanin, netko prikazao i platio u nekom razdoblju. Svatko tko je prikazao takve obveze ne bi trebao biti izložen problemu dokazivanja izvora sredstava za kupnju nekretnine. Međutim netko tko godinama nije plaćao porez ili je plaćao neodgovarajuće iznose poreza, a ima sredstva za kupnju nekretnina, morao bi dokazati izvore financiranja. To znači da bi se takav postupak vodio samo izuzetno, samo kod onih kod kojih se u informacijskom sustavu porezne uprave ne bi moglo utvrditi izvor financiranja za nabavljene nekretnine.

**Tablica 11.** Koeficijenti korelacije i determinacije regresijskih modela "Model 1" do "Model 5"

Model	Koeficijent korelacije	Koeficijent determinacije	Prilagođen koeficijent determinacije
1	0,931	0,867	0,794
2	0,931	0,867	0,803
3	0,931	0,866	0,811
4	0,930	0,865	0,816
5	0,928	0,862	0,819

*Izvor: Izvorno istraživanje.*

<sup>1</sup> Kako je koeficijent višestruke korelacije monotono neopadajuća funkcija broja nezavisnih varijabli, njegova se vrijednost može povećati dodavanjem više varijabli, čime model postaje reprezentativniji Šošić, I. (2004). *Uvod u statistiku*, Školska knjiga, Zagreb.

Jednadžba linearne regresije za "Model 5" je slijedeća:

Ukupno naplaćena porezna utaja =  $-88100,618 + 0,0080204$  Ukupan prihod 2002 +  $0,00810$  Ukupan prihod 2003 +  $0,129$  Ukupan prihod 2005 +  $4113802,749$  Stambene nekretnine 2001 +  $3011132,280$  Stambene nekretnine 2004 +  $4407270,468$  Stambene nekretnine 2005 +  $4107003,694$  Stambene nekretnine ukupne +  $1252724,744$  Nekretnine kupljene ukupno.

Analizom nezavisnih varijabli uključenih u konačnu verziju pokazalo se da je ukupan naplaćeni iznos porezne utaje slabo povezan s ukupnim prihodom poduzeća u posljednjih nekoliko godina, ali je veza između porezne utaje i kupljenih stambenih nekretnina vrlo snažna. Ovaj je nalaz u potpunosti sukladan s općepoznatom praksom prelijevanja sredstava nezakonito stečenih poreznom utajom u privatnu imovinu poduzetnika. Korištenjem prikazanog modela može se prognozirati iznos porezne utaje. Međutim, operativna vrijednost navedenog modela pokazala bi se tek njegovim inkorporiranjem u ekspertni sustav dostupan stručnjacima zaduženim za otkrivanje porezne evazije.

## 5. ZAKLJUČAK

Uvijek treba imati na umu kako izbjegavanje plaćanja poreza stvara veće porezno opterećenje za sve porezne obveznike kojima se povećava porezna stopa. Izbjegavanjem plaćanja poreza opterećuju se drugi porezni obveznici i smanjuju se prihodi mirovinskih i zdravstvenih fondova, što pogađa cijelo stanovništvo. Osim nepravednog poreznog opterećenja, porezna utaja utječe na kvalitetu života cijelog stanovništva, jer umanjuje sredstva ne samo za državni aparat nego i sredstva za mirovinske i zdravstvene fondove, obrazovanje i socijalnu zaštitu.

U radu je metodom jednostavne linearne regresije razvijen model koji temeljem podataka prikupljenih iz porezne uprave o poreznim obveznicima – pravnim osobama (dobit/gubitak, kupljene nekretnine i pokretnine) otkriva kod kojih poreznih obveznika bi bilo korisno izvršiti nadzor nad poslovanjem s ciljem otkrivanja porezne utaje. Istraživanjem otkrivanja znanja iz baza podataka informacijskog sustava porezne uprave, obradom podataka i razvijanjem modela pomoću regresijske analize numeričkih varijabli na uzorku od dvjesto poduzeća došlo se do dokaza da je porezna utaja u korelaciji s kupnjom stambenih nekretnina, što znači da je kupnja stambene nekretnine podatak koji treba pratiti.

Moguće je da postoje izvori sredstava za takvu kupovinu koji se mogu očitovati i opravdati ostvarenom dobiti poduzeća, kreditnim izvorima ili naslijeđenim sredstvima. U slučaju da ne postoji evidencija o takvim izvorima sredstava, potrebno je izvršiti nadzor, utvrditi izvore financiranja, utvrditi porezne obveze i poduzeti ostale mjere.

Zaključak rada je da bi se razvijanjem modela za otkrivanje porezne evazije iz podataka u informacijskom sustavu porezne uprave došlo do identifikacije poreznih obveznika kod kojih je prioritet izvršiti nadzor u svrhu utvrđivanja ispravne porezne osnovice. Kad bi postalo vidljivo da se nadzori obavljaju ciljano, da porezna uprava zna gdje treba tražiti utaje, potakle bi se i promjene u ponašanju poreznih obveznika te bi takva praksa utaje poreza postala manje učestala.

## LITERATURA

1. Chan, PK., Fan, W., Prodromidis, AL., Stolfo, SJ. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, Vol. 14, No. 6, 67-74.
2. Chun-Nan, L., Ching-Hui, W., JinSheng, R. (2007). Application of Data Mining in Tax Evasion of Profit-Seeking Enterprise Income Tax – A Case Study. *Proceedings of the 13th Asia Pacific Management Conference*, Melbourne, Australia, 2007, 1081-1084.
3. Daskalaki, S., Kopanas, I., Goudara, M., Avouris, N., (2003). Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research*, Vol. 145, No. 2, 239-255.
4. Han, J., Kamber, M., "Data Mining: Concept and Techniques", Morgan Kaufmann, 2000.
5. Hand, DJ., Withrow, C., Adams, NM., Juszczak, P., Weston, D., (2008). Performance criteria for plastic card fraud detection tools. *Journal of the Operational Research Society*, Vol. 59, No. 7, 956-962.
6. Inmon, W. H., "Building The Data Warehouse", John Wiley & Sons, 2002.
7. J. A., Berry, M., Linoff, G., "Data Mining Techniques", Wiley Computer Publishing, Canada, 1997.
8. Kirkos, E. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* Vol.32, No.4, 995-1003.
9. Madžarević – Šujster, S., "Procjena porezne evazije u Hrvatskoj", *Financijska teorija i praksa* 26 (1), 2002.
10. Morley, NJ., Ball, LJ., Ormerod, TC. (2006). How the detection of insurance fraud succeeds and fails. *Psychology, Crime and Law*, Vol. 12, No. 2, 163-180.
11. Pejić Bach, M., "Rudarenje podataka u bankarstvu", *Zbornik radova Ekonomskog fakulteta*, 1/2005.
12. Pisani, S., DeSisti, P. (2008). Risk Analysis applied to Tax Evasion using Data Mining Methodology. Dostupno na: [http://www1.agenziaentrate.it/ufficiostudi/documenti/risk\\_analysis\\_tax\\_evasion.pdf](http://www1.agenziaentrate.it/ufficiostudi/documenti/risk_analysis_tax_evasion.pdf) (Pregledano: 25.10.2008.)
13. Qifeng, Y., Bin, F., Ping, S. (2008). *Research on Anti-tax Evasion System based on Union-Bank Online Payment mode*. Springer, Boston.
14. Tanzi, V., "The underground economy in the United States: annual estimates: 1930-1980", *IMF Staff Papers*, Vol. 30 No.2, 1983.
15. Viaene, S., Derrig, RA., Dedene, G. (2004). A case study of applying boosting naive Bayes to claim fraud diagnosis. *IEEE Transactions on Knowledge & Data Engineering*, Vol. 16, No. 5, 612-620.
16. Vukšić, Z., "Porezna utaja", *Hrvatska pravna revija*, 3/2003.
17. Weatherford, M. (2002). Mining for fraud. *IEEE Intelligent Systems*, Vol. 17, No.4, 4-6.
18. Xing, DS., Girolami, M. (2007). Employing Latent Dirichlet Allocation for fraud detection in telecommunications. *Pattern Recognition Letters*, Vol. 28, No. 13, 1727-1734.
19. Yang, WS., Hwang, SY. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, Vol. 31, No. 1, 56-68.
20. Zhang, DS., Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man & Cybernetics Part C: Applications & Reviews*, Vol. 34, No. 4, 513-522.