UDK: 368:371.694:61:659.24:681.3.064:004.4:004.021:679.058:681.322.2:681.3.06
(4-672EEZ) (73)(479.11)
doi: 10.5937/tokosig1903007P

Branko R. Pavlović¹

IZAZOVI U PRIMENI MAŠINSKOG UČENJA U DELATNOSTI OSIGURANJA

PREGLEDNI RAD

Apstrakt

Internet, inteligentni uređaji, Blockchain tehnologija,² softverski roboti³ i razni aspekti veštačke inteligencije, kao što je mašinsko učenje, vrlo često se pominju u savremenoj literaturi kao veliki potencijal za unapređenje procesa u oblasti osiguranja. Ovaj rad biće posvećen mašinskom učenju.

Mašinsko učenje je podskup veštačke inteligencije namenjen za proučavanje i prepoznavanje obrazaca u ponašanju, korišćenjem statističkih metoda obrade dostupnih podataka. Drugim rečima, mašinsko učenje je softver koji može samostalno da zaključuje na osnovu prethodnog iskustva. Ključna korist koju kompanije mogu imati od mašinskog učenja jeste predviđanje budućih trendova, na osnovu toga što softver samostalno uočava obrasce u dostupnim podacima.

Mašinsko učenje može se koristiti u osiguravajućim kućama za optimalnije određivanje tarifa, efikasnije rešavanje i adekvatnije rezervisanje šteta, te moćno oružje u borbi protiv prevara.

Ključne reči: *mašinsko učenje, osiguranje*

I. Uvod

Mašinsko učenje je metoda analize podataka, koja pomoću specifičnih algoritama omogućava pronalaženje skrivenih veza u podacima, bez eksplicitnog programiranja postupka pronalaska tih veza. Drugim rečima, mašinsko učenje je

¹ Predsednik Udruženja aktuara Srbije
I-mejl: pavlovic.branko@gmail.com
Rad je primljen: 25. juna 2019.
Rad je prihvaćen: 16. jula 2019.

² Pavlović, B. (2018). Blockchain Technology in Insurance and Reinsurance, XVI međunarodni simpozijum „Novi izazovi na tržištu osiguranja“, Arandelovac

³ Pavlović, B. (2019). Robot Usage in Insurance, XVII međunarodni simpozijum „Osiguranje na pragu IV industrijske revolucije“, Zlatibor

sposobnost kompjutera da pronađe nove informacije u odnosu na one što su bile poznate u trenutku učenja.

Primenjuje se u prepoznavanju lica, prepoznavanju objekata na slikama ili na video-klipu (npr. prepoznavanju anomalija na rendgenskim snimcima), zatim u autonomnoj vožnji automobila, igranju igara na tabli (npr. šah, dame, tavlja i go), igranju računarskih igara (npr. Doom i Counter Strike), u kvizovima, klasifikaciji teksta, prevodenju, analizi društvenih mreža, prepoznavanju govora itd.

Predmet rada jeste analiza mogućnosti za unapređenje procesa u delatnosti osiguranja korišćenjem tehnika mašinskog učenja, kao i izazova koji se pritom javljaju. Cilj rada jeste da ukaže na koristi od uključivanja srpskih osiguravajućih kuća u jedan od trendova koji donosi IV industrijska revolucija.

1. Podela metoda mašinskog učenja

Mašinsko učenje zasniva se na mogućnosti specijalizovanog softvera da napravi sopstvenu logiku i samostalno uoči veze između podataka. Koriste se tri metoda⁴ rešavanja problema.

1.1. Nadgledano učenje

Podrazumeva formiranje trening skupa podataka i njegovo korišćenje za učenje softvera. Uočene veze između podataka u trening skupu softver primenjuje na novi skup podataka, koji do tada nije video. Primer je klasifikacija tkiva na maligna i benigna.

Problemi koji se rešavanju metodom nadgledanog učenja dele se na regresivne i klasifikacione.

Kod regresivnih problema cilj je da se unete promenljive povežu i da se rezultat predstavi neprekidnom funkcijom. Primer je predviđanje cena nekretnina na osnovu prethodnog iskustva praćenja transakcija na tržištu nekretnina.

Kod klasifikacionih problema, cilj je da se rezultat predstavi u obliku diskretne izlazne vrednosti. Primer je prepoznavanje rukopisa.

Sličnosti i razlike regresivnih i klasifikacionih problema mogu se objasniti na sledećem primeru:⁵ date su realne veličine kuća zajedno s njihovim cenama u trening skupu i cilj je da se na osnovu veličine kuće predvidi njena cena. Cena kao funkcija u zavisnosti od veličine kuće predstavlja neprekidni izlazni parametar. Dakle, to je regresivni problem. Ali ukoliko je cilj da se sazna da li je kuća prodana za manju ili veću cenu od predviđene, umesto da se predvidi cena kuće, to postaje klasifikacioni problem. Tada se kuće klasifikuju u dve diskretne kategorije u zavisnosti od toga da li su prodate po predviđenoj ceni (jedna kategorija) ili nisu (druga kategorija).

⁴ Nikolić, M. i Zečević, A. (2019). Mašinsko učenje, Matematički fakultet Univerziteta u Beogradu

⁵ Milovanović, A. (2015). Seminarski rad „Mašinsko učenje“, Matematički fakultet Univerziteta u Beogradu

1.2. Nenadgledano učenje

Ne koristi se trening skup, nego samo tehnike pronalaska skrivenog znanja na skupu koji se analizira. Primer⁶ je klasifikacija tkiva u grupe sa sličnim karakteristikama.

Kod nenadgledanog učenja mogu da se formiraju strukture podataka (grupe ili klasteri). Klasteri se prave na osnovu povezanosti promenljivih u podacima. Cilj je da se uoče pravilnosti među podacima i nema povratne informacije zasnovane na predviđenim rezultatima, to jest ne zna se šta je tačno, a šta nije. Zbog toga se potencijalno rešenje problema ovim metodom ne može optimizovati. Klasterovanje predstavlja dobar metod za, na primer, tagovanje dokumenata.

1.3. Učenje potkrepljivanjem⁷

Teorijski okvir učenja potkrepljivanjem opisuje se Markovljevim procesima odlučivanja (engl. Markov Decision Processes). Neka agent i okruženje imaju interakciju samo u diskretnim trenucima $t = 0, 1, \dots$. U svakom trenutku t , agent opaža stanje okruženja S_t iz konačnog skupa stanja S i preduzima akciju A_t iz konačnog skupa dopustivih akcija $A(s)$ u konkretnom stanju „ s “, dobija nagradu R_{t+1} iz konačnog skupa nagrada R i prelazi u novo stanje S_{t+1} . Osnovno je Markovljevo svojstvo da novo stanje i nagrada zavise samo od prethodnog stanja i preduzete akcije, a ne od cele istorije procesa.

Nagrade implicitno definišu cilj agenta. Određuju se tako da pokušavajući da maksimira nagradu, agent obavlja posao koji treba da se obavi. Klasična greška je da se nagrade koriste za usmeravanje agenta kako da nešto uradi, umesto šta da uradi. Na primer, ukoliko agent igra šah, nagradu treba da dobije ukoliko pobedi u partiji, a ne ukoliko pojede protivničku figuru.

II. Pregled literature

Veliki broj autora u svetu bavio se mašinskim učenjem u poslednjih sedamdeset godina, dok se su primenom ove metode analize podataka u delatnosti osiguranja autori počeli baviti relativno skoro. Artur Samjuel dao je neformalnu definiciju mašinskog učenja⁸ 1959. godine, na sledeći način: „Oblast nauke koja daje računaru mogućnost da uči a da prethodno ne bude eksplicitno programiran.“ Savremenu naučnu definiciju⁹ dao je Tom Mičel 1997. godine: „Kaže se da računarski program uči iz iskustva E , uzimajući u obzir određenu klasu zadataka T i merilo učinka P , ako se njegov učinak u zadacima T , meren sa P , poboljšava sa iskustvom E .“

U Tabeli 1 su prikazani najznačajniji događaji u razvoju mašinskog učenja u prethodnom periodu.

⁶ Novaković, J. (2013). Rešavanje klasifikacionih problema mašinskog učenja, Reinženjering poslovnih procesa, vol. 4, Fakultet tehničkih nauka u Čačku

⁷ Nikolić, M. i Zečević, A. (2019). *Mašinsko učenje*, Matematički fakultet Univerziteta u Beogradu

⁸ Samuel, L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of research and development*, 3(3), p. 210–229.

⁹ Mitchell, T. (1997). *Machine Learning*, McGraw Hill and MIT Press, USA

Tabela 1. Pregled istorijskog razvoja mašinskog učenja

Godina	Događaj
1943.	Mekaloh i Pits su formulisali „threshold logic“, preteču neuronskih mreža
1950.	Alen Tjuring se bavio mašinama koje uče
1952.	Artur Semjuel je napisao prvi program koji uči da igra igru dame
1957.	Frenk Rozenblat je napravio hardverski „perceptron“
1963.	Vapnik i Červonenkis su predložili prvu varijantu metoda potpornih vektora
1967.	Kaver i Hart su predložili algoritam „k najbližih suseda“ s primenom u problemu trgovačkog putnika
1975.	Verbos je formulisao algoritam propagacije unazad (engl. back-propagation) za izračunavanje gradijenta neuronske mreže
1981.	Dedžong je uveo učenje zasnovano na objašnjavanju kojim se omogućuje izvođenje pravila iz podataka
1985.	Sejnovski i Rozenberg napravili su softver koji može da uči da izgovara engleske reči
1992.	Bozer, Gijon i Vapnik predložili su upotrebu kernela s metodom potpornih vektora, što je dovelo do toga da ovaj metod dominira oblašću mašinskog učenja tokom devedesetih
1992.	Tezauro je napravio TD-Gammon, sistem koji igra igru tavlja (engl. Backgammon)
2006.	Hinton je uveo izraz duboko učenje (engl. deep learning) za algoritme za trening višeslojnih neuronskih mreža koje od tada dominiraju oblašću mašinskog učenja
2011.	IBM-ov sistem Watson pobedio je prvake u američkom popularnom televizijskom kvizu Jeopardy!
2012.	Google X je razvio sistem koji je u stanju da sam pregleda video zapise na YouTube-u i prepozna je mačke

Godina	Događaj
2014. ¹⁰	Facebook sistem za prepoznavanje lica pomoću neuronskih mreža, DeepFace, dostigao je uspešnost prepoznavanja oko 97%
2016.	Googleov sistem Alfa Go pobedio je svetskog prvaka u igri go rezultatom 4:1
2017. ¹¹	Alphabetov sistem Jigsaw, zasnovan na mašinskom učenju, uspeo je da prepozna i zaustavi poplavu nerelevantnih, automatskih komentara koje su ostavljali roboti na sajtovima (engl. on-line trolling)
2018. ¹²	Unapređen model prepoznavanja teksta govornih jezika pomoću neuronskih mreža uvođenjem prethodnog treniranja (engl. pretrained language model)

Izvor: Nikolić, M. (2018). *Uvod u nadgledano mašinsko učenje*, Matematički fakultet Univerziteta u Beogradu

U delatnosti osiguranja autori su se pretežno bavili primenom mašinskog učenja u određivanju tarifa, rezervisanju šteta i borbi protiv prevara. Grupa američkih autora¹⁰ 2019. godine bavila se sledećim tehnikama mašinskog učenja: *simple decision tree*, *random forest* i *boosted trees* na primeru formiranja tarifa u osiguranju na osnovu podataka o štetama. Švajcarski profesor Wuthrich istražio je 2016. godine rezervisanje šteta pomoću Chain Ladder metode analiziranjem ponašanja pojedinačnih šteta metodom mašinskog učenja *regression trees*.¹¹ Grupa australijskih autora¹² je 2018. godine, korišćenjem LASSO modela, opisala potpuno automatizovano rezervisanje šteta na primeru nematerijalnih šteta osiguranja od auto-odgovornosti. Grupa turskih autora¹³ je 2015. godine dala doprinos u formiranju sistema za prevenciju prevara u zdravstvenom osiguranju pomoću metoda mašinskog učenja.

Domaći autori se nisu do sada bavili primenom mašinskog učenja u delatnosti osiguranja u Srbiji.

¹⁰ Henckaerts, R., Cote, M.-P., Antonio, K. and Verbelen, R. (2019). *Boosting Insights in Insurance Tariff Plans with Tree-based Machine Learning*, Cornell University arXiv:1904.10890, New York, USA

¹¹ Wuthrich, M. (2016). *Machine Learning in Individual Claims reserving*, *Swiss Finance Institute Research Paper Series No. 16–67*, Geneva, Swiss

¹² McGuire, G., Taylor, G. and Miller, H. (2018). *Self-assembling Insurance Claim Models Using Regularized Regression and Machine Learning*, <http://dx.doi.org/10.2139/ssrn.3241906>, Sydney, Australia

¹³ Kose, I., Gokturk, M. and Kilic, K. (2015). *An Interactive Machine-learning-based Electronic Fraud and Abuse Detection System in Healthcare Insurance*, *Applied Soft Computing* 36, p. 283–299

III. Koncept

1. Princip rešavanja problema mašinskim učenjem

Veliki broj problema iz prakse može se svesti na neku vrstu funkcije koja zavisi od određenog broja parametara. Npr. cene nekretnine zavise od površine, broja soba, lokacije, godine gradnje, vrste grejanja, sprata itd. Ako softver nekako uspe da poveže sve te podatke ili samo podatke koje on izabere, tako da može odrediti cenu nove nekretnine, smatra se da je softver naučio da procenjuje cenu. Kod mašinskog učenja, čovek ne mora da zna ništa o tome kako je softver odredio funkciju međuzavisnosti podataka. Uloga čoveka je da obezbedi dovoljno podataka iz prošlosti npr. o prodatim nekretninama, a uloga softvera je da te podatke nekako poveže.

Ipak, bitno je razumeti princip po kome to softver radi. Jedan od jednostavnijih metoda jeste da prvo pokuša uspostaviti uprošćenu funkciju cene nekretnine pomoću npr. 3 promenljive, na sledeći način:

$$\text{cena}(x_1, x_2, x_3, x_4) = \text{površina} * x_1 + \text{lokacija} * x_2 + \text{starost} * x_3 + x_4$$

gde su x_1 - x_4 tzv. težinski faktori koje softver bira tako da se dobije optimalno rešenje, tj. da se dobije najmanja greška. Na početku svakom težinskom faktoru može da dodeli vrednost 1 i pusti sve primere prodatih nekretnina koje ima. Naravno, dobijene cene na osnovu ove funkcije mnogo će se razlikovati od originalnih. Kada se saberu greške na kvadrat za sve postojeće podatke o prodatim nekretninama, dobija se ukupna greška, koja se može zapisati na sledeći način:

$$E = \sum (\text{stvarna_cena} - \text{cena}(x_1, x_2, x_3, x_4))^2$$

Sledeći korak jeste menjanje težinskih faktora s ciljem da ukupna greška bude minimalna. Ukoliko bi se postiglo to da ukupna greška bude jednaka nuli, to bi značilo da je funkcija savršena i da se svi primeri iz prošlosti mogu potpuno opisati datom funkcijom i izabranim težinskim faktorima, odnosno da se s velikom verovatnoćom, na osnovu pomenutih parametara (površina, lokacija i starost), može pogoditi i cena za koju će se prodati sledeća nekretnina. Naravno, menjanje težinskih faktora ne vrši se nasumičnim pogađanjem, nego tako što se minimizira funkcija ukupne greške pomoću parcijalnih izvoda po svakom težinskom faktoru:

$$\frac{\partial E}{\partial x} = 0$$

za $x_1, \dots, x_4$.

U praksi, standardni softver za mašinsko učenje sve to sam uradi, isprobavajući mnogo komplikovanije međuzavisnosti promenljivih nego što je linearna funkcija data u primeru, ali je princip sličan.

2. Osobine

Iako je koncept mašinskog učenja prilično jednostavan, potrebno je poprilično iskustvo da bi se uspešno primenio na određeni problem iz prakse. Istraživanja su pokazala da softver za mašinsko učenje koji konfiguriše iskusni informatičar daje mnogo bolje međuzavisnosti podatka nego kada stručnjak iz određene oblasti pokušava formulisati pravila na osnovu znanja i iskustva.

Softver daje funkciju, koja je za čoveka što ga je konfigurisao crna kutija, ali iako čovek ne razume međuzavisnosti između podataka, može da pokaže na primerima da su rezultati koje daje softver ispravni.

Mašinsko učenje funkcioniše samo u slučajevima kada u realnosti postoji veza između podataka. Ukoliko se u softver ubace podaci o tome koje je knjige pročitao vlasnik nekretnine i cena za koju je nekretnina prodana, nema te magije koja na osnovu knjiga može da odredi cenu nekretnine. Najbolji rezultati dobijaju se kada se uključi mašinsko učenje u probleme koje bi i čovek mogao rešiti, jer ih softver za mašinsko učenje rešava znatno brže i kvalitetnije.

3. Primena mašinskog učenja

Mašinsko učenje u praksi se svodi na interaktivni proces sa sledećim koracima:

- Obezbeđenje relevantnih podataka za rešavanje datog problema
- Priprema podataka za analizu mašinskim učenjem, s ciljem da se radi samo kvalitetnim i pouzdanim podacima
- Izbor odgovarajućeg algoritma mašinskog učenja
- Trening algoritma, koji može biti nadgledan, nenadgledan i potkrepljen neophodan je za formiranje kvalitetnog modela
- Evaluacija modela, da bi se izabrao algoritam s najboljim performansama na konkretnom problemu
- Kreirani model distribuiran se korisnicima u vidu aplikacije
- Korisnici rešavaju problem na osnovu sopstvenih podataka, koje model do tada nije video
- Procena validnosti rešenja problema i zatim vraćanje na početak procesa dok se ne postigne dovoljno dobro rešenje.

4. Algoritmi¹⁴

Algoritmi mašinskog učenja drugačiji su od standardnih algoritama koji se sreću u informatici, jer umesto da algoritam počinje unosom podatka, kod

¹⁴ Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*, IBM Limited Edition, John Wiley & Son, New York, USA

mašinskog učenja podaci kreiraju model. Glavna prednost tog pristupa jeste ta što eliminiše predrasude i pogrešne pretpostavke iz modela. Izbor odgovarajućeg algoritma nije lak i potrebno je veliko iskustvo kako bi se napravio dovoljno dobar izbor, koji će dati prilagođen model konkretnom problemu. Najčešće su algoritmi napisani na jednom od sledećih programskih jezika: Java, Python i R.

Najinteresantniji algoritmi mašinskog učenja su sledeći:

- Linearna regresija – to je algoritam koji se najviše koristi u mašinskom učenju. Koristi se za statističke analize tako što uspevaju da uspostave veze između podataka. Algoritmi koji koriste regresiju mogu da kvantifikuju jačinu korelacije između promenljivih u datom skupu podataka, kao i da predvide vrednosti promenljivih u budućnosti, na osnovu podatka o njima iz prošlosti.
- Klasterovanje – objekti sa sličnim osobinama grupišu se u klustere tako da svi objekti u klasteru budu međusobno sličniji u odnosu na objekte iz drugih klastera. Algoritam prvo prepoznaje parametre objekta, a zatim na osnovu parametara razvrstava objekte.
- Stablo odlučivanja – koristi se struktura stabla da se prikažu rezultati odlučivanja. Svaki krajnji čvor stabla predstavlja mogući ishod algoritma, dok iz čvorova odluke izlaze grane u zavisnosti od vrednosti parametra koji se nalazi u čvoru. Osim navedenih, koriste se i algoritmi bazirani na instancama, regularizacija, mašinsko učenje bazirano na pravilima, neuronske mreže, itd.

5. Preduslovi za intenzivan razvoj primene mašinskog učenja

Iako su osnovni principi mašinskog učenja razvijeni pre više od pedeset godina, primena u praksi naročito je intenzivirana u poslednjih nekoliko godina, jer su tek tada ispunjeni potrebni preduslovi.

Moderni računarski procesori dostigli su potrebnu snagu da rade dovoljno brzo s velikom količinom podataka. Troškovi skladištenja podataka dramatično su pali. Pojavile su se relativno jednostavne mogućnosti za povezivanje velikog broja kompjutera u klustere koji mogu analizirati velike količine podataka zadovoljavajućom brzinom. Zahvaljujući opštoj digitalizaciji, mnogo različitih podataka (meteorološki, medicinski, društvene mreže itd.) postaju komercijalno upotrebljivi putem usluga u oblaku i API interfejsima. Algoritmi mašinskog učenja postali su besplatni i lako dostupni zainteresovanim preko open-source zajednice (npr. Googleov TensorFlow).

Zahvaljujući ispunjenim navedenim preduslovima, intenzivna primena mašinskog učenja nije zaobišla ni delatnost osiguranja.

IV. Primena mašinskog učenja u osiguranju

U delatnosti osiguranja u svetu su već realizovani brojni projekti bazirani na mašinskom učenju.

Oko trećina evropskih osiguravajućih kuća koristi mašinsko učenje u

svom poslovanju, pokazala je analiza¹⁵ Evropskog nadzornog tela za osiguranje i profesionalna penzijska osiguranja (engl. *European Insurance and Occupational Pensions Authority* = EIOPA) o upotrebi Big Data Analytics na 28 evropskih tržišta osiguranja u osiguranju od auto-odgovornosti i zdravstvenom osiguranju, u kome su učestvovala 222 entiteta iz delatnosti osiguranja. Metode i softver mašinskog učenja već koristi 31 odsto osiguravajućih kuća, dok još 24 odsto radi na njihovom razvoju. Pomenute napredne metode omogućavaju precizno procenjivanje raznih trendova, sa ljudskom intervencijom ili bez nje, i na taj način povećavaju efikasnost u donošenju poslovnih odluka i smanjuju troškove.

Analiza koju je sproveda EIOPA pokazala je da se tradicionalni izvori podataka (demografski podaci, podaci o izloženosti rizicima i podaci o štetama) češće koriste u kombinaciji s tehnologijom mašinskog učenja nego novi izvori podatka (npr. podaci iz telematike, genetski podaci, podaci o korišćenju kreditnih kartica). Najvažniji rezultat korišćenja tradicionalnih podataka na nov način jesu usluge bolje prilagođene pojedinačnim osiguranicima i valjanija procena rizika. Takođe, osiguravači sve više koriste podatke koje dobijaju od trećih strana za izračunavanje kreditnih bodova, rejtinga na osnovu sigurnosti vožnje, bodovanja višestrukih štetnika itd.

1. Oblasti primene mašinskog učenja u osiguranju

Rezultat istraživanja EIOPA jeste da se mašinsko učenje može koristiti u sledećim oblastima¹⁶ u osiguranju:

- Razvoj novih usluga – uvođenje usluga koje se baziraju na upotrebi određenog resursa (npr. vožnja vozila), usluge skrojene po potrebama pojedinačnog osiguranika, nove vrste osiguranja (npr. cyber osiguranje) itd.
- Tarifiranje i preuzimanje rizika – poboljšana procena rizika, uvođenje novih dodatnih faktora za obračun cene polise, prediktivni modeli za štete, optimizacija cena, modeliranje raskida polisa itd.
- Prodaja – automatizovano savetovanje, izbacivanje posrednika iz procesa prodaje polisa, sofisticirani sistem upravljanja odnosom s klijentima (engl. *Customer Relationship Management*), povećanje broja interakcija s osiguranicima itd.
- Postprodaja – pametne aplikacije na mobilnim uređajima, usluga klijentima tokom svih 24 sata dnevno, *Chat botovi*, automatska obaveštenja o opasnostima od npr. poplave, oluje itd.
- Rešavanje šteta – napredna analitika u prevenciji prevara u osiguranju, prepoznavanje oblika na fotografijama za podršku proceni šteta na vozilima, automatsko prepoznavanje faktura npr. iz servisa za popravku vozila i njihovo plaćanje itd.

¹⁵ EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu

¹⁶ EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu

2. Izazovi u primeni mašinskog učenja u osiguranju

Uvođenje mašinskog učenja u delatnost osiguranja može da donese više štete nego koristi ukoliko se ne obrati pažnja na sledeće izazove:

- Diskriminacija – budući da se učenje obično izvodi nad podacima iz prošlosti, kao i da se ne može zabraniti algoritmima mašinskog učenja da pronađu očigledne veze koje su zabranjene za korišćenje prilikom određivanja tarifa po regulativi Evropske unije (npr. da žene žive duže), mora se na kraju proveriti da li je ono što preporučuje softver u skladu s propisima.
- Netransparentnost modela – kao što je već objašnjeno, rad softvera koji primenjuje mašinsko učenje predstavlja crnu kutiju za korisnike, što može izazvati probleme u režimu Solventnost II, koji zahteva da modeli budu transparentni i da ih mogu proveravati revizori.
- Nefer poslovanje – optimizacija cena može rezultovati time da softver u obračun uključi ne samo faktore rizika nego i faktore o osetljivosti pojedinih grupa na promenu cene ili nesklonost drugih grupa da pretražuju tržište u cilju pronalaska najbolje ponude. Takođe, softver bi mogao da pronađe koliki je procenat iznosa pune štete većina oštećenih spremna da prihvati i da odustane od sudskog postupka za potraživanje ostatka iznosa. Ovakva praksa ne bi bila etična i takav poslovni odnos s osiguranicima ne bi se mogao uzeti kao fer odnos.
- Zloupotreba genetskih podataka – s obzirom na to da genetski podaci osiguranika mogu postati dostupni preko eksternih kompanija, softver ih može upotrebiti za određivanje cena zdravstvenog ili životnog osiguranja, što nije u skladu s osiguravajućom praksom u Evropskoj uniji.
- Neusaglašenost sa GDPR-om – kompanije koje koriste crnu kutiju mašinskog učenja teško mogu dokazati da su usaglašene sa Opštom uredbom o zaštiti podataka o ličnosti Evropske unije (engl. *General Data Protection Regulation* = GDPR), odnosno da na pravi način rukuju ličnim podacima, što u Evropskoj uniji može dovesti do velikih novčanih kazni.
- Kompromitovanost podataka – cyber rizici su sve veći u savremenom poslovanju, što može ishodovati situacijom da softver mašinskog učenja koristi podatke koji su kompromitovani, pa izvede neodgovarajuće zaključke, a da ljudi koji upravljaju njime toga uopšte nisu svesni.
- Neadekvatnost klasičnih polisa¹⁷ – ako softver sa sposobnošću mašinskog učenja prouzrokuje štetu, postavlja se pitanje ko je za to odgovoran. Pitanje osiguranja od odgovornosti za upotrebu veštačke inteligencije je komplikovano, jer potencijalni krivci mogu biti od programera algoritama, preko koda i integratora, do vlasnika skupa podataka i proizvođača koji sve to koristi u svom proizvodu. Skriveni rizici koje nosi veštačka inteligencija nateraću osiguravače da u budućnosti unaprede svoje polise.

¹⁷ Papović, B. (2019) Prikaz članka „Veštačka inteligencija – skriveni rizici“, izvor *Insurance Post*, novembar 2018, 37–38 str, Tokovi osiguranja 1/2019, p. 107–108

V. Primeri primene mašinskog učenja u osiguranju u praksi

Jedan od realizovanih primera mašinskog učenja u praksi, koji se koristi i u osiguravajućim kućama jeste klasifikacija poruka elektronske pošte na spam ili regularnu poštu.

Ipak, postoji i nekoliko realizovanih primera koji su specifični samo za oblast osiguranja u kojima je ova tehnologija primenjena. Četiri najveće osiguravajuće kuće u SAD, po zvaničnim podacima¹⁸ iz 2016. godine, koriste mašinsko učenje u različitim poslovnim procesima¹⁹, dok u Srbiji Udruženje osiguravača Srbije koristi algoritme mašinskog učenja u borbi protiv prevara.

1. State Farm²⁰

Najveća američka osiguravajuća kuća u 2016. godini „State Farm“ koristi mašinsko učenje u klasifikaciji sigurnosti vožnje svakog vozača koji je osiguran kod njih, a zatim i u pripremi ponude odgovarajućih usluga za svaku klasu vozača.

Pre tri godine „State Farm“ otvorila je konkurs za najuspešnije rešenje za klasifikaciju vozača, na osnovu fotografija iz saobraćaja. Vozači su klasifikovani u deset kategorija: sigurni vozači, oni koji šalju SMS poruke u vožnji, vozači koji podešavaju radio tokom vožnje, oni koji razgovaraju preko telefona itd. Prvo mesto je osvojilo rešenje koje je koristilo mašinsko učenje i metodu dveju neuronskih mreža.

2. Liberty Mutual²¹

Jedna od najvećih američkih osiguravajućih kuća, „Liberty Mutual“, osnovala je „Solaria Labs“, specijalizovanu kompaniju za inovacije. Jedan od projekata koji je „Solaria Labs“ razvila 2017. godine jeste razvoj portala s otvorenim aplikacionim programskim interfejsom ili API interfejsom (skraćeno od engleskog naziva *Application Programming Interface*). Svrha portala jeste integrisanje sopstvenih informatičkih projekata „Solaria Labsa“ s javnim podacima, s ciljem razvoja aplikacije bazirane na mašinskom učenju za poboljšanje sigurnosti u saobraćaju.

Takođe, razvili su aplikaciju za mobilne uređaje koja pomaže vozačima što su imali saobraćajnu nesreću da brzo procene štetu na vozilu korišćenjem kamere mobilnog telefona. Trening skup aplikacije bazirane na mašinskom učenju sastojao se od više hiljada fotografija saobraćajnih nesreća uparenih s podacima o troškovima popravki vozila sa fotografija.

¹⁸ National Association of Insurance Commissioners web site www.naic.org

¹⁹ Sennaar, K. (2019). How America's Top 4 Insurance Companies are Using Machine Learning, Emerj Artificial Intelligence Research, <http://emerj.com>, Boston, USA

²⁰ www.statefarm.com

²¹ www.libertymutual.com

3. Allstate²²

Osiguravajuća kuća „Allstate“ razvila je virtuelnog asistenta, *Chat-bot* ABIE (skraćeno od engl. *Allstate Business Insurance Expert*), zasnovanog na metodama mašinskog učenja, za pomoć svojim agentima u prodaji proizvoda za korporativne klijente. Ranije su njihovi agenti bili obučeni samo za prodaju osiguranja kuća i stanova i zdravstvenog osiguranja, tako da im je bila potrebna pomoć kada je kompanija odlučila da im poveri prodaju složenijih usluga imovinskih osiguranja. ABIE je obezbedio agentima vodič „korak po korak“ na engleskom jeziku za primenu tarifa, određivanje cena i pripremu ponude raznih usluga osiguranja za korporativne klijente. Budući da je ABIE pružao kvalitetnu pomoć, agenti su koristili njegove usluge u oko 25.000 slučajeva mesečno.

4. Progressive²³

Velika američka osiguravajuća kuća „Progressive“ koristi algoritme prediktivne analitike mašinskog učenja za proučavanje podataka o vozačima s ciljem razumevanja tržišnih trendova i mogućnosti za dalji razvoj usluga osiguranja vozila. Njihova aplikacija za telematiku, Snapshot, u 2016. godini prikupila je podatke o vožnji na preko dvadeset milijardi kilometara.

5. Udruženje osiguravača Srbije²⁴

Skoro sve veće srpske osiguravajuće kuće imaju formirane organizacione delove za sprečavanje prevara, koje se sa većim ili manjim uspehom bore protiv prevara. U cilju što efikasnije borbe s prevarantima iz oblasti osiguranja vozila, Udruženje osiguravača Srbije obezbedilo je 2015. godine osiguravajućim kompanijama moćan alat za borbu protiv prevara. FROPS (engl. *Fraud Risk Operational Performance Solution*) softver britanske kompanije „Salviol“²⁵ za prevenciju prevara, analitički je alat koji se koristi u osiguranju i drugim finansijskim sektorima. Glavni cilj FROPS-a jeste da sakuplja, upoređuje, istražuje i analizira velike količine podataka. On traži anomalije odstupanja ili nekoherentnosti u podacima s ciljem otkrivanja prevarnih radnji koje dovode do gubitka prihoda. FROPS analizira najširi spektar informacija kako bi obezbedio najefikasnije analitičko okruženje s maksimalnom preciznošću ključnih indikatora prevara. On vrši indeksiranje podataka iz kojih obračunava rizik od prevare. Funkcioniše nezavisno od drugih programskih okruženja. Koristi prediktivnu analitiku i mašinsko učenje za identifikovanje novih tipova prevara u osiguranju.

²² www.allstate.com

²³ www.progressive.com

²⁴ www.uos.rs

²⁵ www.salviol.com

Ključni indikatori prevara jesu parametri štete koji s velikom verovatnoćom ukazuju na postojanje prevarne radnje. Oni predstavljaju kritičnu tačku za uspostavljanje sistema za otkrivanje i sprečavanje prevara. Pravilnim kombinovanjem tih indikatora, sužava se skup šteta koje su potencijalno prevare. Sektor za borbu protiv prevara fokusira se na taj skup u daljoj kontroli i analizi, što povećava njegovu efikasnost. Prepoznavanjem nestrukturiranih podataka, tekstualnih datoteka, skeniranih dokumenta itd. FROPS omogućava njihovo pretraživanje, analizu i kategorizaciju. Tako pripremljeni podaci daju analitičaru neophodnu širinu za proces istrage.

Takođe, FROPS analizira socijalne mreže kvantitativnom tehnikom koja kombinuje organizacionu teoriju s matematičkim modelima. Na osnovu tih algoritma, analitičar može bolje razumeti dinamiku grupa, mreža i organizacija. Može da identifikuje lica ili organizaciju koja krije informaciju, da prati i analizira kako se mreže razvijaju u toku vremena i da velikom brzinom utvrdi važne karakteristike posmatrane grupe.

VI. Zaključak

U delatnosti osiguranja u svetu su već realizovani brojni projekti bazirani na mašinskom učenju u borbi protiv prevara, određivanju tarifa, rešavanju šteta, podršci korisnicima i agentima, naročito u zdravstvenom osiguranju i osiguranju vozila. U Srbiji je najpoznatiji primer upotrebe mašinskog učenja u osiguranju softver FROPS, koji koristi Udruženje osiguravača Srbije.

Doprinos ovog rada je demistifikacija primene mašinskog učenja u osiguranju, uz skretanje pažnje na izazove koje korišćenje mašinskog učenja donosi – sve kao podsticaj domaćim osiguravajućim kuća da unapređenje i osavremenjivanje pojedinih internih poslovnih procesa počnu da vrše korišćenjem tehnika mašinskog učenja.

S obzirom na brojne koristi od uvođenja savremenih tehnologija u delatnost osiguranja, koje su opisane u radu i potkrepljene primerima iz strane i domaće prakse, očekuje se masovnije korišćenje mašinskog učenja u srpskim osiguravajućim kućama u narednih nekoliko godina.

Literatura

- EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu
- Henckaerts, R., Cote, M.-P., Antonio, K. and Verbelen, R. (2019). *Boosting Insights in Insurance Tariff Plans with Tree-based Machine Learning*, Cornell University arXiv:1904.10890, New York, USA
- Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*, IBM Limited Edition, John Wiley & Son, New York, USA
- Kose, I., Gokturk, M. and Kilic, K. (2015). An Interactive Machine-learning-based Electronic Fraud and Abuse Detection System in Healthcare Insurance, *Applied Soft Computing* 36, p. 283–299

- McGuire, G., Taylor, G. and Miller, H. (2018). Self-assembling Insurance Claim Models Using Regularized Regression and Machine Learning, <http://dx.doi.org/10.2139/ssrn.3241906>, Sydney, Australia
- Milovanović, A. (2015). Seminarski rad „Mašinsko učenje“, Matematički fakultet Univerziteta u Beogradu
- Mitchell, T., (1997). *Machine Learning*, McGraw Hill and MIT Press, USA
- National Association of Insurance Commissioners web site www.naic.org
- Nikolić, M. i Zečević, A. (2019). *Mašinsko učenje*, Matematički fakultet Univerziteta u Beogradu
- Novaković, J. (2013). Rešavanje klasifikacionih problema mašinskog učenja, *Reinženjering poslovnih procesa*, vol. 4, Fakultet tehničkih nauka u Čačku
- Papović, B. (2019). Prikaz članka „Veštačka inteligencija – skriveni rizici“, izvor *Insurance Post*, November 2018, 37–38 str, Tokovi osiguranja 1/2019, p. 107–108
- Pavlović, B. (2018). Blockchain Technology in Insurance and Reinsurance, XVI međunarodni simpozijum „Novi izazovi na tržištu osiguranja“, Arandelovac
- Pavlović, B. (2019). Robot Usage in Insurance, XVII međunarodni simpozijum „Osiguranje na pragu IV industrijske revolucije“, Zlatibor
- Ruder, S. (2018). A Review of the Neural History of Natural Language Processing, www.ruder.io
- Samuel, L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of research and development*, 3(3), p. 210–229
- Sennaar, K. (2019). How America's Top 4 Insurance Companies are Using Machine Learning, Emerj Artificial Intelligence Research, <http://emerj.com>, Boston, USA
- Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014). *DeepFace: Closing the Gap to Human-Level in Face Verification*, Facebook AI Research, Menlo Park, USA
- Wuthrich, M. (2016). Machine Learning in Individual Claims reserving, *Swiss Finance Institute Research Paper Series No. 16–67*, Geneva, Swiss
- www.allstate.com
- www.jigsaw.google.com
- www.libertymutual.com
- www.progressive.com
- www.salviol.com
- www.statefarm.com
- www.uos.rs

UDK: 368:371.694:61:659.24:681.3.064:004.4:004.021:679.058:681.322.2:681.3.06
(4-672EEZ) (73)(479.11)
doi: 10.5937/tokosig1903007P

Branko R. Pavlović¹

CHALLENGES IN APPLICATION OF MACHINE LEARNING IN INSURANCE INDUSTRY

REVIEW ARTICLE

The Internet of Intelligent Things, Blockchain technology², software robots³ and various aspects of artificial intelligence such as machine learning are often referenced in modern literature as having great potential to improve the processes in insurance business. This paper will focus on machine learning.

Machine learning is a subset of artificial intelligence intended for study and recognition of behaviour patterns by using statistical methods for available data processing. In other words, machine learning is a software able to make its own decisions based on previous experience. The key benefit companies can draw from machine learning is prediction of future trends, where software independently discovers patterns in the available data.

Insurance companies may use machine learning to optimize their tariffs, settle claims more efficiently, and improve the quality of loss reserving, and as a powerful tool in combating frauds.

Key words: *machine learning, insurance*

I. Introduction

Machine learning is a method of data analysis that uses specific algorithms to uncover hidden connections in data, without being explicitly programmed to act in order to find these connections. In other words, machine learning is the ability of a computer to find the information that is new to what was known at the time of learning.

¹ President of the Serbian Actuarial Association

E-mail: pavlovic.branko@gmail.com

Paper received on: June 25, 2019

Paper admitted on: July 16, 2019

² Pavlović, B. (2018). Blockchain Technology in Insurance and Reinsurance, the 16th International Symposium *New Challenges in Insurance Market*, Arandelovac

³ Pavlović, B. (2019). Robot Usage in Insurance, the 17th International Symposium *Insurance on the Eve of the Fourth Industrial Revolution*, Zlatibor

It is used in face recognition, recognition of objects in images or videos (e.g. X-ray detection of abnormalities), self-driving cars, board games (e.g. chess, checkers, backgammon and go), computer games (e.g. Doom and Counter Strike), quizzes, text classification, translation, social network analysis, speech recognition, etc.

The subject of this paper is the analysis of possibilities for improving the processes in insurance industry by using machine learning techniques and the challenges arising therefrom. The aim is to highlight the benefits that Serbian insurance companies can derive from taking part in one of the trends that the fourth industrial revolution brings upon.

1. Classification of Machine Learning Methods

Machine learning is based on the ability of specialized software to create its own logic and independently notice the connections between data. Three problem-solving methods⁴ are used.

1.1. Supervised Learning

It involves forming of a training dataset and using it to teach the software. The links between the data identified in the training set are applied to a new dataset, which the software has not seen before. An example is the classification of tissue into malignant and benign.

The problems solved by the supervised learning method are divided into regressive and classification problems.

In regressive problems, the goal is to connect the entered variables and present the result with a continuous function. One example of the aforementioned is the prediction of real estate prices based on previous experience in monitoring the transactions on the property market.

In classification problems, the goal is to present the result in the form of a discrete output value. An example is handwriting recognition.

The similarities and differences of regressive and classification problems can be explained by the following example:⁵ the training set provides the actual sizes of houses with their prices in the training set, and the goal is to predict the house price based on its size. Price, as a function of the size of a home, is a continuous output parameter. Therefore, it represents a regressive problem. However, if the goal is to find out if the house was sold for less or greater than the predicted price, instead of predicting the price of the house, it becomes a classification problem. Subsequently, the houses are classified into two discrete categories, depending on whether they were sold at the predicted price (one category) or not (the other category).

⁴ Nikolić, M. and Zečević, A. (2019) *Mašinsko učenje*, Faculty of Mathematics, University of Belgrade

⁵ Milovanović, A. (2015). Seminar Paper *Mašinsko učenje*, Faculty of Mathematics of the University of Belgrade

1.2. Unsupervised Learning

Rather than using a training set, employed are the techniques for uncovering hidden knowledge in the analysed set. The classification of tissues into the groups with similar characteristics may serve as an example⁶.

In unsupervised learning, data structures (groups or clusters) may be formed. Clusters are made based on the correlation of variables in the data. The goal is to see the correctness of the data, and there is no feedback based on predicted results, that is, it is not known what is true and what is not. Therefore, a potential solution to the problem cannot be optimized with this method. Clustering is a good method for, for example, tagging documents.

1.3. Reinforcement Learning⁷

Theoretical framework of reinforcement learning is described by Markov Decision Processes. Let us assume that the agent and environment interact only at discrete time $t = 0, 1, \dots$. At any time t , the agent perceives a finite set S of distinct states S_t in its environment and has a finite set of distinct actions A_t that it can perform from the finite set of permissible actions $A(s)$ in the particular state „ s “, receives a reward R_{t+1} from the finite set of rewards R and moves forward to the new state S_{t+1} . The main Markov's property is the property that the new state and reward depend only on the previous state and action taken and not on the entire process history.

Rewards implicitly define the agent's goal. They are determined so that in trying to maximize the reward, the agent does the work that needs to be done. The classic mistake is that rewards are used to direct the agent toward how to do something instead of what to do. For example, if an agent plays chess, he should be rewarded if he wins the game, not if he captures an opponent's piece.

II. Literature Review

In the past seventy years, a large number of authors around the world have been exploring machine learning and yet, when it comes to insurance industry, it was not until quite recently that they have begun to address this method of data analysis. In 1959, Arthur Samuel provided an informal definition of machine learning⁸, as follows: "The field of study that gives computers the ability to learn without being explicitly programmed." A modern scientific definition⁹ was formulated by Tom Mitchell in 1997: „A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .“

⁶ Novaković, J. (2013). Rešavanje klasifikacionih problema mašinskog učenja, Reinženjering poslovnih procesa, vol. 4, Faculty of Technical Sciences in Čačak

⁷ Nikolić, M. and Zečević, A. (2019). *Mašinsko učenje*, Faculty of Mathematics of the University of Belgrade

⁸ Samuel, L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of research and development*, 3(3), p. 210-229.

⁹ Mitchell, T., (1997). *Machine Learning*, McGraw Hill and MIT Press, USA

The Table 1 shows the most significant past events in the development of machine learning.

Table 1 Brief History of Machine Learning

Year	Event
1943	McCulloch and Pitts formulated „threshold logic“, a predecessor of neural networks
1950	Alan Turing dealt with learning machines
1952	Arthur Samuel developed the first computer checkers-playing programme
1957	Frank Rosenblatt developed hardware „perceptron“
1963	Vapnik and Chervonenkis proposed the first variant of support vectors
1967	Cover and Hart proposed „k-nearest neighbours“ algorithm applied to the travelling salesmen problem
1975	Werbos formulated backpropagation algorithm for calculating the gradient of a neural network
1981	DeJong introduced explanation-based learning that enabled deriving rules from data
1985	Sejnowski and Rosenberg developed a software able to learn how to pronounce English words
1992	Boser, Guyon and Vapnik proposed the use of kernel with support vector method, making this method predominant in the machine learning of the 1990s
1992	Tesauro developed TD-Gammon, the system for playing the game of backgammon
2006	Hinton introduced the term deep learning for training algorithms for multi-layered neural networks that have been predominant in the machine learning ever since
2011	IBM Watson system faced off against champions of the popular American television quiz show Jeopardy and won!
2012	Google X developed a system able to independently view video recordings on YouTube and identify cats

B. Pavlović: Challenges in Application of Machine Learning in Insurance Industry

Year	Event
2014. ¹⁰	Facebook system for facial recognition that uses neural networks, DeepFace, scored a 97% success rate.
2016.	Google AlphaGo system beat a world champion in the game series 4-1
2017. ¹¹	Alphabet's Jigsaw, based on machine learning, managed to recognise and stop on-line trolling
2018. ¹²	Improved neural network language model for recognition of spoken languages by introduction of pretrained language model

Source: Nikolić, M. (2018). *Uvod u nadgledano mašinsko učenje*, Faculty of Mathematics, University of Belgrade

In the insurance industry, the authors have mainly addressed the application of machine learning in defining insurance rates, claims reserving, and combating frauds. In 2019, a group of American authors¹⁰ dealt with the following machine learning techniques: simple decision tree, random forest and boosted trees, on the example of developing tariff plans based on the claims data. In 2016, a Swiss professor Wüthrich explored claims reserving using Chain Ladder method and analysing the behaviour of individual claims by machine learning method and tree-based regression techniques¹¹. In 2018, a group of Australian authors¹² used LASSO model to describe fully automated loss reserving on the example of non-material damages in motor third party liability insurance. In 2015, a group of Turkish authors¹³ contributed to setting a system for prevention of frauds in health insurance by using machine learning method.

So far, the domestic authors have not explored the use of machine learning in Serbian insurance business.

¹⁰ Henckaerts, R., Cote, M.-P., Antonio, K. and Verbelen, R. (2019). *Boosting Insights in Insurance Tariff Plans with Tree-based Machine Learning*, Cornell University arXiv:1904.10890, New York, USA

¹¹ Wuthrich, M. (2016). *Machine Learning in Individual Claims Reserving*, *Swiss Finance Institute Research Paper Series No. 16-67*, Geneva, Swiss

¹² McGuire, G., Taylor, G. and Miller, H. (2018). *Self-assembling Insurance Claim Models Using Regularized Regression and Machine Learning*, <http://dx.doi.org/10.2139/ssrn.3241906>, Sydney, Australia

¹³ Kose, I., Gokturk, M. and Kilic, K. (2015). *An Interactive Machine-learning-based Electronic Fraud and Abuse Detection System in Healthcare Insurance*, *Applied Soft Computing* 36, pp. 283-299

III. Concept

1. Problem-Solving with Machine Learning

A large number of problems encountered in practice can be reduced to some kind of a function that depends on a number of parameters. For example, real estate prices depend on the area, number of rooms, location, year of construction, type of heating, floor, etc. If the software somehow manages to connect all that information or just the information it chooses so it can determine the price of the new property, the software is considered to have “learned” to do the pricing. In machine learning, one does not need to know anything about the method that software uses to determine the function of data interdependence. The role of a human is to provide sufficient information from the past e.g. about real estate sold, whereas the role of the software is to somehow connect that data.

However, it is important to understand the principle according to which software works. One of the simpler methods is to first try to establish a simplified property price function with, for example, 3 variables, as follows:

$$\text{price}(x_1, x_2, x_3, x_4) = \text{area} * x_1 + \text{location} * x_2 + \text{age} * x_3 + x_4$$

where x_1 - x_4 are so-called weighting factors selected by the software to obtain the optimal solution, i.e. to obtain the smallest error. In the beginning, it can assign each weighting factor a value of 1 and release any examples of the sold real property it has available. Of course, the prices obtained through this feature will be much different from the original ones.

When squaring errors for all existing data on the sales of real estate, a total error is obtained, which can be presented as follows:

$$E = \sum (\text{actual_price} - \text{price}(x_1, x_2, x_3, x_4))^2$$

The next step is to change the weighting factors so as to reduce the total error to a minimum. If the total error could be made equal to zero, this would mean that the function is perfect and that all examples from the past can be fully described by the given function and the selected weighting factors, that is, the price at which the next property will be sold can be predicted with high probability, based on the mentioned parameters (surface, location and age). Naturally, weighting factors are not changed by random guessing, but by minimizing the total error function with the use of partial derivatives for each weighting factor:

$$\frac{\partial E}{\partial x} = 0$$

for $x_1, \dots, x_4$.

In practice, standard machine learning software does it all on its own, trying out interdependencies of variables that are much more complicated than the linear function given in the example, but the principle is similar.

2. Features

Although the concept of machine learning is quite simple, it takes considerable experience to successfully apply it to a particular problem in practice. Researches have shown that machine learning software configured by an experienced developer provides much better data interdependence than the attempts of an expert in a particular field who uses his knowledge and experience to formulate the rules.

Software provides a function that is a black box for the person who configured it, but even though one does not understand the interdependencies between the data, one can show that the results provided by the software are correct by using examples.

Machine learning works only when in reality there is a connection between the data. If the software includes the information about the books that the property owner has read and about the price for which the property was sold, there is no chance in the world to use the books to determine the price of property. The best results are obtained when machine learning is engaged in solving the problems that could be solved by humans, because machine learning software solves them in a much faster and better way.

3. Application of Machine Learning

In practice, machine learning is an iterative process with the following steps:

- Providing relevant data to solve a particular problem
- Preparing the data for analysis by machine learning with the aim to process only good-quality and reliable data
- Selecting appropriate machine learning algorithm
- Training the algorithm, which may be supervised, unsupervised and reinforced, which is necessary for the creation of a good model
- Evaluating the model to select the algorithm with the best performances in relation to a particular problem
- Distributing the created model to the users in the form of application
- Problem-solving by users based on their own data which the model has not seen before
- Assessing the validity of problem solution and returning to the beginning of the process until satisfactory solution is reached.

4. Algorithms¹⁴

Machine learning algorithms differ from standard algorithms in information technology, because in machine learning, the data create model rather than starting the algorithm with data input. The main advantage of this approach is that it eliminates biases and misconceptions from the model. Choosing the right algorithm is not an easy task and it takes an extensive experience to make a choice that is good enough to produce a model suited to the specific problem. Most commonly, algorithms are written in one of the following programming languages: Java, Python, and R.

The most interesting machine learning algorithms are:

- Linear regression - this algorithm is mostly used in machine learning. These algorithms are used for statistical analyses by managing to establish links between data. Algorithms using regression can quantify the strength of the correlation between variables in a given dataset, and based on historical data predict the values of variables in the future.
- Clustering - objects with similar properties are grouped into clusters so that all the objects in one cluster are more similar to each other than the objects from other clusters. Algorithm firstly recognizes the object parameters and based on the parameters subsequently classifies the objects.
- Decision Tree – A flowchart-like tree structure is used to display decision results. Each end node of a tree represents a possible outcome of the algorithm, while branches come out of the decision nodes depending on the value of the parameter in the node.

In addition to the above, used are instance-based algorithms, regularization, rule-based machine learning, neural networks, etc.

5. Preconditions for Intensive Development of Machine Learning Application

Although the main principles of machine learning were developed more than fifty years ago, the application in practice has been particularly intensified in recent years, because it was not until then that the necessary preconditions were met.

Modern computer processors have now become able to deal with a large amount of data and still run fast enough. Data warehousing costs have dramatically dropped. Relatively simple possibilities were created enabling the connection of a large number of computers into clusters that can analyse large amounts of data at a satisfactory speed. Due to general digitalisation, many different data (meteorological, medical, social networks, etc.) have become commercially available through cloud services and APIs. Through open source communities (e.g.

¹⁴ Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*, IBM Limited Edition, John Wiley & Son, New York, USA

Google's TensorFlow), machine learning algorithms have become free and easily accessible to those interested.

Owing to the fulfilment of these prerequisites, the insurance industry was not left behind in the intensive application of machine learning.

IV. Application of Machine Learning in Insurance

Insurance industries worldwide have already implemented numerous projects that were based on machine learning.

The analysis of Big Data Analytics conducted by the European Insurance and Occupational Pensions Authority (EIOPA)¹⁵ has shown that in motor and health insurance, about one third of European insurance companies use machine learning in their business. A total of 222 insurance undertakings and intermediaries from 28 jurisdictions have participated in this thematic review. Machine learning methods and software are already actively used by 31% of companies, and another 24% are at a proof of concept stage. These advanced methods enable accurate assessments of different trends with or without human intervention, increasing the efficiency of decision-making and thus reducing operational costs.

The analysis conducted by EIOPA has shown that traditional data sources (demographic data, exposure data and claims data) are increasingly combined with machine learning technology, rather than with new sources (like telematics data, genetic data, credit card data). The most important result of the use of traditional data in a new manner is that it enables the development of products better tailored to individual insureds and better risk assessment. In addition, insurers increasingly use the data outsourced from third-party data vendors to calculate credit scores, driving scores, claims scores, etc.

1. Application of Machine Learning across Insurance Industry

The EIOPA study has shown that machine learning may be used in the following areas of insurance¹⁶:

- Development of new products – introduction of usage-based products (e.g. driving of a vehicle), products tailored to the needs of individual insured person, new types of insurance (e.g. cyber insurance), etc.
- Pricing and underwriting – enhanced risk assessment, introduction of new, additional elements to policy pricing, predictive models for claims, price optimisation, policy cancellation modelling, etc.
- Sales – automated advice, eliminating intermediaries from policy selling, sophisticated Customer Relationship Management System, increased number of interactions with insureds, etc.
- Post-sale services and assistance – smart applications of mobile devices, customer

¹⁵ EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu

¹⁶ EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu

service available 24/7, chatbots, automatic warnings of natural hazards such as floods, storms, etc.

- Claims management – advanced analytics in the prevention of insurance frauds, verification of photographs as a support to loss assessment on vehicles, automated invoice verification e.g. from repair shops and their payment, etc.

2. Challenges in Machine Learning Implementation in Insurance

The introduction of machine learning to insurance industry may do more harm than good if the following challenges are not addressed:

- Discrimination – given that the learning usually relies on historic data and that machine learning algorithm cannot be prohibited from finding evident connections which, according to the regulations of the European Union, are not allowed for use when defining tariffs (e.g. the information that women live longer than men), it is necessary to eventually verify whether software recommendations are in compliance with regulations.
- Non-transparency of the model – as already explained, the operation of the machine learning software is a black box to customers and may lead to problems in Solvency II regime which requires that the models are transparent and verifiable by auditors.
- Unfair commercial practices – price optimisation may lead the software to include in its calculation not only the risk factors but also the sensitivity factors of particular groups to change in prices or reluctance of particular groups to search the market for a better offer. In addition, the software could find what percentage of the full loss amount the majority of claimants are willing to accept and drop the lawsuit regarding the rest of the amount. Such a practice would not be ethical and could not be considered a fair business relationship with insured persons.
- Abuse of genetics data – since the genetics data may become available through external companies, software may use them for pricing in health or life insurance, which is not in conformity with the EU insurance practice.
- Non-compliance with the GDPR – the companies using the machine learning black box can hardly claim to be compliant with the EU General Data Protection Regulation (GDPR) i.e. to properly handle personal data, which is possibly subject to very high fines in the European Union.
- Compromised data – modern business is increasingly exposed to cyber risks and may lead to a situation where the people operating the machine learning software are not aware that it uses compromised data and draws inadequate conclusions.
- Inadequacy of traditional insurance policies¹⁷ – if a software with machine learning capabilities causes damage, the question may be posed as to who is accountable? Insurance issues are tricky as the potential stakeholders in any AI application range from the algorithm designer, coder and integrator and the owner of the data sets, to the manufacturer of the product using them. Hidden risks of artificial intelligence will force insurers to advance their policies.

¹⁷ Papović, B. (2019) Review of Article Artificial Intelligence – Hidden Risks, source *Insurance Post*, November 2018, pp. 37–38, *Insurance Trends* 1/2019, pp. 107-108

V. Examples of Applying Machine Learning in Insurance Practice

One of machine learning examples realized in practice and also used by insurance companies is the classification of e-mails into spam or regular mail.

In addition, there are several insurance-specific examples where this technology was applied. According to the 2016 official data¹⁸, the top four insurance companies in the USA use machine learning in different business processes¹⁹, whereas in Serbia, the Association of Serbian Insurers uses machine learning algorithms to combat frauds.

1. State Farm²⁰

The biggest insurance company in 2016, State Farm, used machine learning to classify the safe driving of each driver who had their insurance policy and then offered the insurance product for each driver class.

Three years ago, State Farm launched a competition for the most effective solution for driver classification based on the photographs that showed behaviour of traffic participants. Drivers were classified into ten categories: safe driving, texting, operating the radio, talking on the phone, etc. The first place application was the one that utilized machine learning and two neural network models.

2. Liberty Mutual²¹

One of the biggest American insurance companies, Liberty Mutual, founded Solaria Labs, a specialised company dealing with innovations. One of the projects developed by Solaria Labs in 2017 is the development of a portal with an open application programming interface or API (Application Programming Interface). The purpose of the portal is to integrate Solaria Labs' own IT projects with public data, with the aim of developing the traffic safety application based on machine learning.

In addition, they have developed a mobile app which assists the drivers who had a car crash to quickly assess vehicle damage using a cellphone camera. The training set of machine learning application consisted of thousands of traffic accident images paired with the data on the repair costs of the photographed vehicle.

¹⁸ National Association of Insurance Commissioners website www.naic.org

¹⁹ Sennaar, K. (2019). How America's Top 4 Insurance Companies are Using Machine Learning, Emerj Artificial Intelligence Research, <http://emerj.com>, Boston, USA

²⁰ www.statefarm.com

²¹ www.libertymutual.com

3. Allstate²²

Insurance company Allstate has developed a virtual assistant, Chat-bot ABIE (Allstate Business Insurance Expert), based on machine learning methods, to help its agents in the sales to corporate clients. Previously, their agents had only been trained in selling homeowners and health insurance, so when the company decided to entrust them with the sale of more complex property insurance products, they required help. ABIE provided agents with a step-by-step guide in English for applying tariffs, pricing, and preparing a variety of corporate insurance products. Since ABIE provided quality assistance, agents benefited from its services on a monthly basis, in about 25,000 cases.

4. Progressive²³

Major American insurance company, Progressive, leverages machine learning algorithms for predictive analytics to study the data on client drivers with the aim to achieve a better understanding of market trends and possibility to further develop motor insurance products. Their telematics app, Snapshot, collected 20 billion kilometres of driving data in 2016.

5. Association of Serbian Insurers²⁴

Almost all major Serbian insurance companies have the organizational units established for fraud prevention, which are more or less successful in combating frauds. In order to successfully combat fraudsters in the field of motor insurance, in 2015, the Association of Serbian Insurers provided insurance companies with a powerful tool to combat frauds. FROPS (Fraud Risk Operational Performance Solution) of the UK-based company Salviol²⁵, software for fraud prevention, is an analytical tool used in insurance and other financial sectors. The main objective of FROPS is to collect, compare, research and analyse large amounts of data. It looks for anomalies, discrepancies or inconsistencies in the data in order to detect fraudulent actions which lead to loss of revenues. FROPS analyses the widest range of information to provide the most efficient analytical environment with maximum precision of key fraud indicators. It indexes the data from which it calculates the risk of fraud and operates independently from other programming environments. This solution uses predictive analytics and machine learning to identify new types of insurance fraud.

The key indicators of fraud are the damage parameters that indicate a high likelihood of fraudulent actions. They represent a critical point for fraud detection and prevention systems. Proper combination of these indicators narrows down

²² www.allstate.com

²³ www.progressive.com

²⁴ www.uos.rs

²⁵ www.salviol.com

the set of potentially fraudulent claims. In further control and analysis, the Anti-Fraud Department focuses on this set, which increases its effectiveness. FROPS recognizes the unstructured data, text files, scanned documents, etc. and enables their search, analysis and categorization. The data thus prepared give the analyst the insight necessary for the investigation process.

In addition, FROPS utilizes quantitative technique to analyse social networks combining the organisational theory with mathematical models. Based on these algorithms, the analyst can better understand the dynamics of groups, networks and organizations. It can identify persons or organization that hides information, monitor and analyse how networks evolve over time, and very quickly identify important characteristics of the analysed group.

VI. Conclusion

Numerous machine-learning projects to combat frauds, define tariffs, settle claims, provide support to customers and agents, have already been implemented in insurance industry worldwide, particularly in health and motor insurance.

This paper aims to shed light on the application of machine learning in insurance and draw attention to the challenges of machine learning, and thus encourage Serbian insurance companies to use machine learning techniques with the aim to advance and modernize their internal business processes.

In view of numerous advantages insurance industry can expect from the introduction of modern technologies described in this paper and supported by the examples from foreign and domestic practice, in the coming years, it is expected that Serbian insurance companies will benefit from a more extensive use of machine learning.

Literature

- EIOPA (2019). Big Data Analytics in Motor in Health Insurance: A Thematic Review, www.eiopa.europa.eu
- Henckaerts, R., Cote, M.-P., Antonio, K. and Verbelen, R. (2019). *Boosting Insights in Insurance Tariff Plans with Tree-based Machine Learning*, Cornell University arXiv:1904.10890, New York, USA
- Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*, IBM Limited Edition, John Wiley & Son, New York, USA
- Kose, I., Gokturk, M. and Kilic, K. (2015). An Interactive Machine-learning-based Electronic Fraud and Abuse Detection System in Healthcare Insurance, *Applied Soft Computing* 36, pp. 283-299
- McGuire, G., Taylor, G. and Miller, H. (2018). Self-assembling Insurance Claim Models Using Regularized Regression and Machine Learning, <http://dx.doi.org/10.2139/ssrn.3241906>, Sydney, Australia
- Milovanović, A. (2015). Seminar Paper „Mašinsko učenje“, Faculty of Mathematics of the University of Belgrade
- Mitchell, T., (1997). *Machine Learning*, McGraw Hill and MIT Press, USA

- National Association of Insurance Commissioners website www.naic.org
- Nikolić, M. and Zečević, A. (2019). *Mašinsko učenje*, Faculty of Mathematics of the University of Belgrade
- Novaković, J. (2013). Rešavanje klasifikacionih problema mašinskog učenja, *Reinženjering poslovnih procesa*, Vol. 4, the Faculty of Technical Sciences in Čačak
- Papović, B. (2019). Review of Article Artificial Intelligence – Hidden Risks, source *Insurance Post*, November 2018, pp. 37–38, *Insurance Trends* 1/2019, pp. 107-108
- Pavlović, B. (2018). Blockchain Technology in Insurance and Reinsurance, the 16th International Symposium *New Challenges in Insurance Market*, Aranđelovac
- Pavlović, B. (2019). Robot Usage in Insurance, the 17th International Symposium *Insurance on the Eve of the Fourth Industrial Revolution*, Zlatibor
- Ruder, S. (2018). A Review of the Neural History of Natural Language Processing, www.ruder.io
- Samuel, L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of research and development*, 3(3), pp. 210-229.
- Sennaar, K. (2019). How America's Top 4 Insurance Companies are Using Machine Learning, Emerj Artificial Intelligence Research, <http://emerj.com>, Boston, USA
- Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014). *DeepFace: Closing the Gap to Human-Level in Face Verification*, Facebook AI Research, Menlo Park, USA
- Wuthrich, M. (2016). Machine Learning in Individual Claims reserving, *Swiss Finance Institute Research Paper Series No. 16-67*, Geneva, Swiss
- www.allstate.com
- www.jigsaw.google.com
- www.libertymutual.com
- www.progressive.com
- www.salviol.com
- www.statefarm.com
- www.uos.rs

Translated by: **Zorica Simović**