

Dátová veda a technológia veľdát v službách aktúarov

Peter Schmidt¹

Abstrakt

Rýchly vývoj a veľké množstvo kvalitných údajov majú významný vplyv na život takmer všetkých ekonomických subjektov vrátane spoločností a podnikov a finančných inštitúcií. Nové technológie a dátová veda môžu byť katalyzátorom ďalšieho rozvoja disciplín, ako je napríklad aktúarstvo. Nové riešenia pre aktúarov môžu ponúkať alternatívne, štatisticky založené metódy predpovedania a odhadovania, dokonca aj pre tradičné účtovníctvo a heuristické postupy plánovania a analýzy nákladov. V tomto článku popisujeme túto metódu ukladania a spracovania dát, jej možnosti a jej prepojenie s riadiacim systémom.

Kľúčové slová

Dátová veda, veľdáta, aktúarstvo, životný cyklus

Abstract

Rapid development and a large amount of quality data have a significant impact on the lives of almost all economic entities, including companies and enterprises and financial institutions. New technologies and data science can be a catalyst for further development of disciplines such as actuarial. New solutions for actuaries may offer alternative, statistically based methods of forecasting and estimating, even for traditional accounting and heuristic planning and cost analysis procedures. In this article we describe this method of data storage and processing, its possibilities and its connection with the control system.

Key words

Data science, big data, actuarial, life cycle

JEL classification

JEL C55

1 Úvod

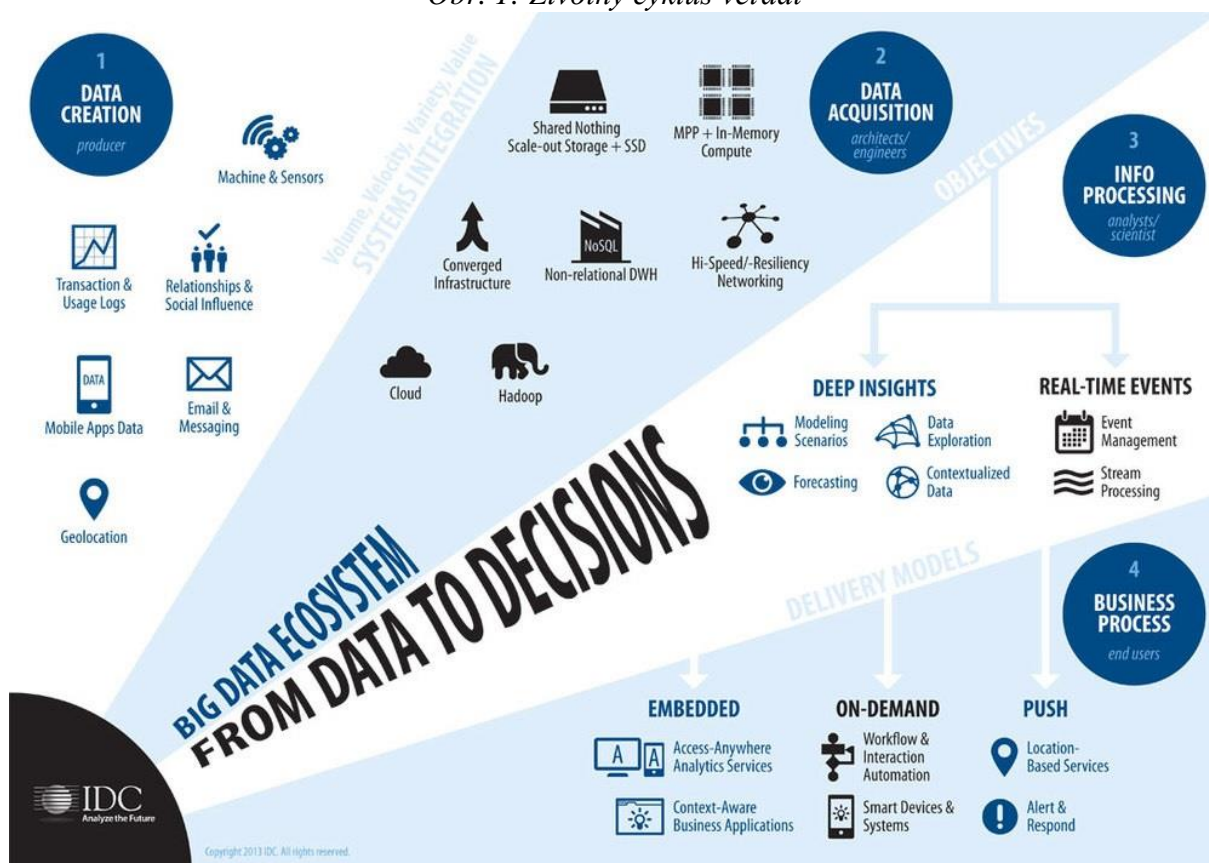
Do druhého desaťročia 21. storočia sa tempo vývoja informačných technológií (IT) zrýchlilo nebývalým tempom. Lacný telekomunikačný a iný výpočtový hardvér, ako aj softvér s otvoreným zdrojovým kódom sa stali dostupnými nielen pre odborníkov v oblasti IT, ale aj pre odborníkov v iných odvetviach a pre každodenných používateľov. Vďaka šíreniu otvorených a cenovo dostupných technológií je dnes technológia takmer v každej oblasti ľudského života. Na novú úroveň vývoja poukazuje skutočnosť, že z týchto každodenne používaných zariadení sa stávajú inteligentné takzvané „smart“ zariadenia (tento termín sa vzťahuje predovšetkým na schopnosť zariadení navzájom komunikovať), čo otvára novú dimenziu zberu údajov. Na jednej strane je množstvo údajov a rýchlosť ich zberu a na strane druhej, ide o nové, zatiaľ nezaznamenané oblasti každodenného fungovania nášho života. Odhadovalo sa, že do roku 2020 ľudstvo zhromaždí viac ako 16 zettabytov užitočných informácií, v skutočnosti ich bolo okolo 26 ZB (Turner, Ganz, Reinsel, & Minton, 2020). Efektívne ukladanie a spracovanie tohto množstva dát je predmetom technológie veľdát (Big Data), efektívnej extrakcii, využívaníu a analýze informácií sa venuje dátová veda založená na matematike a štatistike.

¹ Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra aplikovanej informatiky, Dolnozemska cesta 1, 852 35 Bratislava, peter.schmidt@euba.sk.

2 Veľdáta, dátová veda a aktuárstvo

V mnohých odvetviach predstavuje technológia veľdát pre organizácie príležitosť aj výzvy. Pojem veľdáta „sa vzťahuje na veľké množstvo údajov, ktoré sa neustále zhromažďujú prostredníctvom technologických zariadení (kreditné karty, karty vernostných bodov, sociálne médiá, zariadenia pripojené k internetu, bezdrôtové snímače a vysielacie atď.) (Chua, 2013). Významná časť údajov je neštruktúrovaná, t.j. nesleduje žiadny špeciálny, alebo preddefinovaný dátový model.“ Veda postavená na technológii veľdát je novou disciplínou. Nič to nedokazuje viac ako skutočnosť, že prvá zaznamenaná zmienka o technológii (zdieľané ukladanie a spracovanie veľkého množstva údajov prostredníctvom počítačových klastrov) pochádza z roku 2003, keď programátori spoločnosti Google uverejnili vedecký článok o novom systéme správy súborov (Ghemawat, Gobioff, & Leung, 2003). Vývoj nových technológií a nových postupov bol motivovaný potrebou získavania informácií z veľkého množstva údajov, či už v ekonomickej, vedeckej alebo dokonca spoločenskej sfére. Kombináciou relevantných častí matematiky, štatistiky a programovania sa vytvoril nový metodický rámec na dosiahnutie týchto cieľov. Aj keď veľa matematických a štatistických postupov a teórií používaných v dátovej vede existuje oveľa dlhšie ako samotná dátová veda, praktické aplikácie ako strojové učenie, alebo umelá inteligencia sa vyvíjajú paralelne s technológiou veľdát.

Obr. 1: Životný cyklus veľdát



Zdroj: (IDC- https://www.idc.com/ap/media-center/infographics/?order=date_1&topics=Big+Data%2C&modal=c65230243393266ff6e5)

3 Životný cyklus veľdát

Termín „veľdáta“ nemusí zákonite predstavovať len veľké množstvá údajov, ale aj rôzne technologické procesy a fázy zhromažďovania, ukladania a používania údajov, a v praxi na systém opísaný v systéme veľkých údajov. Obrázok 1 zobrazuje celý systém od generovania údajov po ich použitie pri rozhodovaní. V anglickej literatúre sa používa slovné spojenie Big Data Ecosystem, ktoré do slovenčiny z hľadiska obsahu môžeme preložiť skôr ako životný cyklus veľdát, nakoľko ekosystém sa používa ako skrátená verzia pre ekologický systém.

3.1 Životný cyklus veľdát je tvorený štyrmi hlavnými fázami:

3.1.1 Vytváranie údajov (Data Creation)

Za posledné desaťročie sa počet zariadení pripojených k internetu významne zvýšil v dôsledku rýchleho vývoja a šírenia internetových technológií. Odborníci odhadujú že v roku 2020 bolo k sieti pripojených vyše 30 miliárd IoT zariadení, toto číslo má v roku 2025 dosiahnuť až 75 miliárd zariadení, a to nielen nových spotrebičov (napríklad telekomunikačných zariadení), ale aj malých a veľkých domácich spotrebičov (napríklad chladničky, pračky, kávovary či dokonca zásuvky). Veľké zariadenia a stroje akými sú automobily, CNC stroje, roboty či iné výrobné zariadenia sa tiež stanú zdrojmi údajov, vďaka ich schopnosti obojsmernej dátovej komunikácie. Využitím tejto vlastnosti budú tieto zariadenia poskytovať podrobné informácie o mnohých doteraz nezmapovaných oblastiach ľudského života, správania sa ľudí či o firemných procesoch (McEwen & Cassimally, 2014).

3.1.2 Zber údajov (Data Acquisition)

Zber údajov je druhým stupňom životného cyklu. Tento stupeň je zodpovedný za zber, ukladanie, systematizáciu, infraštruktúru a distribúciu údajov. Súčasný vývoj v troch technologických oblastiach umožnil vytvorenie tohto stupňa:

- výrazné zvýšenie výkonu operačných pamätí a procesorov používaných v počítačových systémoch, ako aj zníženie jednotkových cien tohto hardvéru;
- dostupnosť efektívneho manažmentu spracovania údajov (klastrové, distribučné spracovanie, napríklad Hadoop, Teradata a pod.);
- radikálne zmena v prístupe v technológiách ukladania a archivovania veľkého množstva údajov. Presadenie sa cloudových riešení pri ukladaní ale aj spracovávaní údajov. Nástup technológie polovodičových pamäťových médií SSD, ktorých kapacita medziročne výrazne rastie a cena klesá. Rozmach vysokorýchlostných sietí umožňuje zaznamenávať, ukladať, spracovávať a distribuovať veľké množstvo údajov v reálnom čase. Budovanie „pomalých“ senzorových sietí, ktoré dokážu dodávať veľké množstvá údajov z prostredí o ktorých sme si to pred nedávnom ani nevedeli predstaviť.

3.1.3 Spracovanie údajov (Info Processing)

V prípade neštruktúrovaných databáz s veľkým objemom údajov, tradičné metódy analýzy údajov nemožno použiť alebo len veľmi obmedzene. Nové typy údajov, ich zrnitosť a časový horizont v kombinácii so zvýšenou kapacitou spracovania umožňujú prieskum, analýzu a modelovanie údajov v reálnom čase, ako aj tvorbu presných predpovedí. To v kombinácii s automatickým učením a umelou inteligenciou umožňuje plne automatizovanú správu udalostí (napríklad vyšetovanie a prevenciu podvodov v reálnom čase, čo je v poisťovníctve vážny argument). To si samozrejme vyžaduje kombináciu matematiky, štatistiky a programovania, pre ktoré dátová veda poskytuje metodický rámec (Tolle, Tansley, & Hey, 2011).

3.1.4 Podnikové procesy (Business processes)

Posledný stupeň životného cyklu nadväzuje na kombinované výsledky prvých troch stupňov a na pridanú hodnotu generovanú zo zozbieraných a spracovaných údajov. Údaje možno použiť rôznymi spôsobmi: na jednej strane môžu organizácii získať nové zdroje výnosov prostredníctvom cieleného predaja údajov a na druhej strane môže organizácia efektívnym využitím údajov, presnejších predpovedí robiť lepšie obchodné rozhodnutia. Ďalšou významnou oblasťou obchodného použitia je automatizovaná relevantná komunikácia so zákazníkmi, prieskumy v reálnom čase a pod. Organizácie, ktoré nájdu správny spôsob, ako využiť nové veľdáta, budú schopné zvýšiť svoju výkonnosť a produktivitu, môžu však tiež získať nové konkurenčné výhody a zdroje príjmu. Zle navrhnuté a nasadené systémy veľdát však môžu viesť k strate informácií, zraniteľnosti a zníženému výkonu. Tieto technológie budú v najbližších rokoch hrať dôležitú úlohu v podnikaní, mnoho organizácií je už teraz v procese objavovania, vývoja systémov a stratégií pre nové technológie. Veľa StartUpov má hlavnú podnikateľskú činnosť práve správu veľdát, ponúkajú služby spojených s dátovou vedou a rôzne ďalšie činnosti, pri ktorých sa generuje ekonomická hodnota z informácií. Otázka adaptácie teda nebude v budúcnosti luxusnou záležitosťou, ale môže byť dokonca základnou podmienkou prežitia organizácie na moderných trhoch (Cavanillas, Curry, & Wahlster, 2018).

4 Dátová veda

Z doterajších poznatkov je zrejmé, že kombinácia výrazne zvýšeného množstva, rozmanitosti a rýchlosti prísunu informácií vyžaduje nové postupy, ktoré podnikom, výskumníkom a analytikom umožnia využívať potenciál technológie Big Data. Veda založená na dátach je interdisciplinárna vedná oblasť, disciplína založená na vzťahu medzi (matematickými a štatistickými) metodikami, (programovaním) procesmi, algoritmi a systémami, ktoré vytvárajú vedecké údaje z veľkého množstva údajov uložených v rôznych formách (štruktúrovaných a neštruktúrovaných), z ktorých sa dá vytvoriť ekonomická, alebo spoločenská hodnota (Dhar, 2013). Dátová veda má niekoľko podskupín ako prieskumná analýza dát, informačné plánovanie, interaktívna vizualizácia údajov, deskriptívna štatistika, postupy odhadovania, infografika, atď. Vedci v oblasti dátovej vedy údajov sú odborníkmi v oblasti dolovania, čistenia, spracovania a prezentácie údajov. Aby dátový vedec, dokázal odhaliť skryté riešenia, musí disponovať určitými znalosťami. Ide o dve skupiny znalostí a to technické a obchodné. Technické znalosti predstavujú prehľad v matematike, štatistike, programovaní v prostredí SQL / R / Python / C / Java, dolovaní dát, vizualizácii údajov, Hadoop atď. Obchodné znalosti a zručnosti predstavujú analytické a komunikačné schopnosti, obchodné a priemyselné znalosti v oblasti dát a veľdát technológií (Kalmár, 2017).

Tab. 1: Možnosti a výzvy dátovej vedy a technológií veľkých dát

Oblasť	Možnosti	Výzvy
Dátová veda ako aktívum	<ul style="list-style-type: none"> • Pomáha spoločnostiam zhodnocovať ich dátové aktíva prostredníctvom vývoja nových metodík • Zvyšovanie hodnoty údajov prostredníctvom správcovstva a kontroly kvality 	<ul style="list-style-type: none"> • Rýchly pokles hodnoty dát, získavaním nových dát • Nevhodné využitie dát môže viesť k strate ich hodnoty • Neistota ohľadom budúceho vývoja v oblasti regulácie, DGPR a podobné obmedzenia môžu viesť k strate hodnoty údajov
Dátová veda a veľké dáta pri podpore rozhodovania	<ul style="list-style-type: none"> • Vypracovanie metodík na podporu rozhodovania v reálnom čase využitím veľkých dát • Spolupráca s ostatnými oddeleniami na odhaľovaní možností zhodnotenia interných a externých dát 	<ul style="list-style-type: none"> • Samoobsluha a automatizácia by mohli znížiť potrebu analytikov • Kultúrne bariéry môžu brániť globálnemu zdieľaniu údajov aj v rámci jednej organizácie
Dátová veda a veľké dáta pri riadení rizika	<ul style="list-style-type: none"> • Rozšírenie zdrojov dát použitých pri predpovedaní rizík s cieľom získať presnejšie predikcie • Identifikácia rizík v reálnom čase na účely zisťovania podvodov a forenzných prípadov • Používanie prediktívnej analýzy na testovanie rizika dlhodobějších investičných príležitostí na nových trhoch a produktoch 	<ul style="list-style-type: none"> • Používanie prediktívnej analýzy na testovanie rizika dlhodobějších investičných príležitostí na nových trhoch a produktoch • Zabezpečenie toho, aby sa pri používaní rôznych zdrojov údajov a analýzy veľkých dát na identifikáciu rizík nezamieňala korelácia s príčinnými súvislosťami. • Prediktívne analytické techniky budú znamenať zmeny v rozpočtovaní a výpočtoch návratnosti investícií.

Zdroj: Vlastné spracovanie

V tabuľke 1 sú zhrnuté príležitosti a výzvy, ktoré nová technológia predstavuje pre finančných, účtovných či poisťovních odborníkov. V súčasnosti už fungujú systémy schopné analyzovať údaje v reálnom čase (napríklad porovnanie plánu s faktami, zisťovanie príčin odchýlok od plánu v rovnakom čase) a zvyšovať kvalitu, frekvenciu a presnosť plánovania a predpovedania. Organizácie, ktoré sa chcú rýchlo prispôbiť novým výzvam a rozšíriť svoje odborné znalosti v oblasti dátovej vedy a technológií veľkých dát, budú môcť posilniť možnosti plánovania a analýzy organizácií, rozvíjať rozhodovacie systémy, identifikovať nové trhy a zlepšiť prevádzkovú efektívnosť, čo v konečnom dôsledku povedie k vyššiemu zisku (Chua, 2013).

Pre organizácie z oblasti finančníctva či poisťovníctva boli informácie cenným nástrojom aj doteraz. Banky, poisťovne pri svojej každodennej činnosti stále používajú množstvo pokročilých štatistických postupov a modelov, napríklad v súvislosti so svojimi aktivitami v oblasti riadenia vzťahov so zákazníkmi (CRM), kde sa tieto modely používajú na identifikáciu zákazníkov, ktorí sú ochotní nakupovať. Pri cielenom marketingu sa organizácia primárne zameriava na týchto zákazníkov počas reklamnej kampane (čím šetrí peniaze a čas pre zákazníka aj organizáciu). Ďalšou dôležitou oblasťou použitia je analýza rizík a prevencia

podvodov (prostredníctvom rozhodovacích stromov a neurónových sietí), ktorá pre finančné inštitúcie predstavuje stále nové výzvy.

Od počiatkov poisťovníctva poisťovatelia používali všetky možné nástroje na čo najlepšiu správu poistených kmeňov. To pôvodne znamenalo, že poisťné riziká sa analyzovali a rozlišovali do samostatných kategórií na základe ich rizikovosti. Takéto delenie umožňovalo stanoviť čo najspravodlivejšie poisťné poplatky na každé riziko. Riziká v združení sa permanentne monitorujú. Toto zohráva úlohu, keď poisťovatelia vypočítavajú svoje technické rezervy, ktoré ukazujú, koľko aktív musia mať na krytie svojej rizikovej pozície (Kivisaari, 2018). V extrémnom prípade môže byť člen združenia príliš riskantný a bude z neho vylúčený (aj keď právne predpisy majú pre toto vylúčenie často prísne obmedzenia). Miera rizika úzko súvisí s množstvom zmysluplných dát, ktoré ponúkajú vysokú mieru spoľahlivosti pri posúdení reálneho rizika. Pri niektorých príležitostiach sa hovorilo, že vďaka fenoménu veľkých dát a vývojom dátovej analýzy sa stane poistenie zastarané. Táto predpoveď sa zakladá na myšlienke, že s novými nástrojmi bude možné predpovedať budúcnosť tak presne, že nebude existovať priestor pre neistotu. Aj keď je pravda, že s veľkými dátami sa objavuje veľa nových výziev v súvislosti s princípom poistenia, je ale zrejmé, že princíp poistenia bude v dohľadnej budúcnosti živý a potrebný.

Koncept veľdát označuje myšlienku, že exponenciálne sa zvyšujúce množstvo digitálnych dát vytvárajú radikálne nový svet. Okrem objemu dát koncept zahŕňa pokroky v oblastiach, ktoré sa označujú ako 3V:

- veracity (pravdivosť - znamená, že je potrebné skontrolovať kvalitu údajov, opraviť skreslenia a vyčistiť šum),
- variability (rôznorodosť - údaje pochádzajú z mnohých zdrojov a často nie sú štruktúrované),
- velocity (rýchlosť – zvyšovanie rýchlosti výpočtovej techniky umožňuje používať čoraz sofistikovanejšie metódy analýzy dát so znižovaním nákladov).

Už teraz možno povedať, že možnosti, ktoré tento vývoj prináša, poskytujú aktúarom revolučne nové nástroje na správu poistených združení (Csicsman, 2018). Aktuári tradične používajú jednoduché veličiny (ako je vek, pohlavie, adresa, fajčiar / nefajčiar) na rozdelenie rizík do rôznych tarifných tried. Ak by si však aktúar mohol vytvoriť obraz o poistníkovi na základe znalostí z neformálnych zdrojov, napr. koľko hodín strávi prácou v sede, koľko krát navštevuje lekára za určité obdobie a čo tam kupuje, aký čas venuje telesnej aktivite a pod., dalo by sa aspoň teoreticky povedať, že nie je potrebné poznať vek alebo pohlavie tejto osoby. Je potrebné povedať, že v tejto oblasti je kritický nedostatok analytických poznatkov. Znalosť podrobností života poistenca by určite rádovo znížili riziko poisťovateľa, avšak by hrubo zasiahlo do osobnej integrity poistenca. V súčasnosti, keď vďaka GDPR je správa aj tých najzakladanejších údajov výzvou, je ťažké si predstaviť rozsiahly zber údajov o poistencovi. Za pár rokov, keď už budú systémy na neustále monitorovanie životných funkcií človeka bežne rozšírené, tak aj zber údajov bude jednoduchší. Poisťovne budú mať lepší prehľad o poistencovi a s ním súvisiacimi rizikami ako samotný poistenec, čiže súčasné status quo sa otočí. Na druhej strane je už dnes vidno porozumenie príčinnej súvislosti určitých údajov, ktoré vedú k rôznym výsledkom. Je potrebné dodať, že dátová revolúcia sa netýka len objemu dát, ale aj vysokej rýchlosti ich spracovania. Dávnejšie, aj keď boli k dispozícii údaje, nízky výpočtový výkon počítačov neumožňoval používanie presných výpočtových metód, nakoľko by výpočet trval neprijateľne dlho. Aj v skorších dobách boli údaje veľmi hodnotným zdrojom, na to aby sa nimi plynulo, preto sa dôraz kládol na vývoj takých postupov, ktoré umožňovali spracovávať stále rastúci objem dát. V ideálnom svete by sme mali k dispozícii úplné informácie a naše analytické nástroje by umožňovali dokonalé predpovede budúcnosti. Napriek tomu by sme nevedeli presne predpovedať čo a kedy sa stane. Stále by sme mohli hovoriť len o pravdepodobnosti, ale by sme dokázali rozlíšiť poistené riziká do presnejších kategórií. Dokázali by sme lepšie

ohodnotiť riziká a stanoviť spravodlivejšie poistné poplatky. Stále by sme však potrebovali spojiť riziká, aby sme mohli ťažiť zo stabilizačného účinku zákona veľkých čísiel. Jedným z hlavných problémov pri využívaní týchto možností je nedostatočná znalosť analytiky údajov, čo má priamy dopad na obmedzené chápanie rôznych príčinných súvislostí. Z obrovského množstva dát je možné nájsť veľa zaujímavých korelácií medzi rôznymi veličinami. Aj keď nám korelácia dvoch veličín ukáže signifikantnú závislosť ešte nie je isté, že tieto dve veličiny spolu súvisia. Je potrebné poznať kauzalitu, t.j. príčinu a dôsledok. Je mimoriadne nebezpečné, ak sa nesprávne interpretuje korelácia ako príčinná súvislosť. Bez pokroku v analytike údajov nemôžeme nové možnosti dobre využiť. Poistenie je vždy založené na modelovaní reality, pričom ako nástroj sa často používa algoritmus. Tieto modely žiaľ nie sú dokonalé. S rastúcim objemom dát rastie aj náročnosť modelu a bude čoraz ťažšie vytvoriť modely, ktoré by verne odrážali realitu. Hrozba spočíva v tom, že modely sa môžu prispôbiť iba dostupným údajom bez prediktívnej sily na situácie, ktoré nie sú v týchto údajoch zahrnuté. Tradičné chápanie v oblasti poistenia je, že informačná asymetria prináša poistníkom výhody, nakoľko poistník vie spravidla viac o svojej situácii v porovnaní s tým, čo vie poisťovateľ. Využitím pokročilých analytických techník by poisťovateľ mohol byť v zásadne v lepšej pozícii voči poistníkovi. Je reálna hrozba, že poisťovatelia zneužijú túto situáciu na úkor poistníka. V krajnom prípade budú poisťovatelia schopní segmentovať riziká do oveľa viac kategórií, ako je možné dnes, pričom niektoré riziká sa stanú nepoistiteľnými. V súčasnosti sa analytika veľdát stále intenzívne rozvíja, čo má veľký vplyv takmer na všetky oblasti života. Cieľom poisťovacích spoločností je odstrániť informačnú asymetriu.

5 Záver

Nová technológia a metodika otvára finančným a investičným inštitúciám možnosť vyvinúť sofistikované prediktívne modely, ktoré umožňujú oceňovanie dlhodobých investícií, trhov, finančných nástrojov (meny, deriváty a opcie), ktoré doposiaľ predstavovali významné riziko kvôli miere neistoty. Nová technológia veľdát umožňuje analytikom zahrnúť do prediktívnych modelov nebývalé množstvo faktorov, údajov, čím sa znižuje ich zjednodušujúci účinok a zvyšuje sa tak presnosť. Nová technológia má významný vplyv na všetky funkcie plánovania, hodnotenia, riadenia a správy informácií v oblasti poisťovníctva. Presnosť a frekvenciu plánovania možno zlepšiť pomocou nových postupov (prediktívne modelovanie nákladov alebo predaja) a nového objemu dát. Pomocou nových metodík je možné hodnotenie efektívnosti v reálnom čase. Nová technológia môže navyše viesť k rýchlejšiemu riadeniu a správe informácií (podnikové informácie, informácie o trhoch, zákazníkoch atď.), ktoré sú dostupné kdekoľvek a kedykoľvek a tým zjednodušiť rozhodovacie procesy. Pomocou strojového učenia sa v budúcnosti automatizuje veľa plánovacích a hodnotiacich procesov, čo ovplyvní efektívnosť prevádzkových procesov. V nadväznosti na súčasné tempo vývoja môžu v nasledujúcom desaťročí vzniknúť špecifické oblasti aktuárstva súvisiace s dátovou vedou a veľdátami.

Príspevok bol spracovaný v rámci riešenia grantovej úlohy VEGA 1/0221/17 *Investičné modelovanie v prostredí katastrofického poistného rizika.*

Literatúra

- [1] Cavanillas, J. M., Curry, E., & Wahlster, W. (2018). *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe*. Cham: Springer International Publishing.
- [2] Chua, F. (2013). Big data: Its power and perils. Retrieved February 21, 2021, from <http://www.accaglobal.com/futures>.

-
- [3] Csicsman, J. (2018). Van ÚJ a big Data ALATT? Napjaink legnépszerűbb It-s HÍVÓSZAVA Csak EGY lehetőség ARRA, hogy az INFORMATIKUSOK ELADJÁK Magukat, vagy VALÓDI ÚJDONSÁG? *Biztosítás és Kockázat*, 5(2), 64-80. doi:10.18530/bk.2018.2.64.
- [4] Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73. doi:10.1145/2500499.
- [5] Ghemawat, S., Gobioff, H., & Leung, S. (2003). The Google file system. *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles - SOSP '03*. doi:10.1145/945445.945450.
- [6] IDC Asia/Pacific infographics. (n.d.). Retrieved February 21, 2021, from https://www.idc.com/ap/media-center/infographics/?order=date_1&topics=Big+Data%2C&modal=c65230243393266ff6e5.
- [7] Kalmár, P. (2017). Adattudomány és „Big Data” technológia a controlling szolgálatában. *Biztosítás és Kockázat*, 5(2), 2-5. doi:https://doi.org/10.24387/CI.2017.2.1.
- [8] Kivisaari, E. (2018). Big data is coming - are you ready? *Biztosítás és Kockázat*, 5(2), 60-63. doi:10.18530/bk.2018.2.60.
- [9] McEwen, A., & Cassimally, H. (2014). *Designing the Internet of things*. Chichester, West Sussex: John Wiley and Sons.
- [10] Tolle, K. M., Tansley, D. S., & Hey, A. J. (2011). The fourth PARADIGM: Data-Intensive scientific Discovery [Point Of View]. *Proceedings of the IEEE*, 99(8), 1334-1337. doi:10.1109/jproc.2011.2155130.
- [11] Turner, V., Gantz, J. F., Reinsel, D., & Minton, S. (2020, July 17). The digital universe of opportunities: Rich data and the increasing value of the internet of things. Retrieved February 21, 2021, from <https://www.coursehero.com/file/p44nd58l/Turner-V-Gantz-J-F-Reinsel-D-Minton-S-2014-The-digital-universe-of/>.