

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

2/2020

ročník/volume 30

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

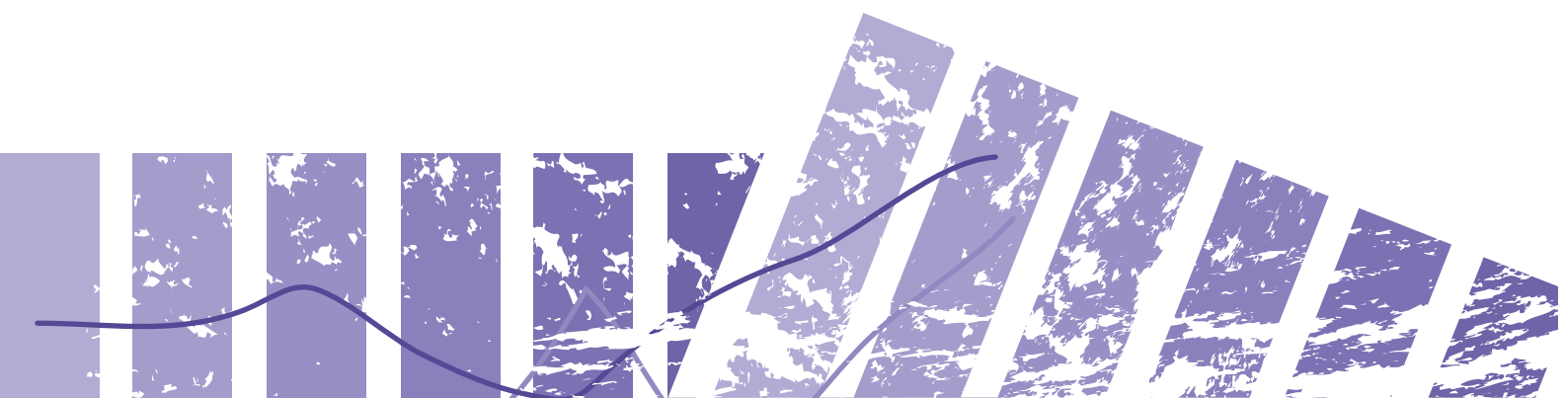
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 4

Typ článku/Type of article: informatívny článok/informative article

Strany/Pages: 57 – 72

Dátum vydania/Publication date: 15. apríl 2020/April 15, 2020



Informatívny článok/Informative article

Michal PÁLEŠ

**Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky
Ekonomickej univerzity v Bratislave**

MOŽNOSTI RIEŠENIA PROBLEMATIKY BIG DATA V JAZYKU R

ONE OF THE BIG DATA VIEWS IN R LANGUAGE

ABSTRAKT

Možno tvrdiť, že údaje sú v súčasnosti jednoznačne jedným z najhodnotnejších aktív v akejkol'vek sfére podnikania. V poisťovniach dennodenne pribúda množstvo údajov o klientoch, zmluvách, poisťných udalostiach, plneniach, majetku, záväzkoch a o iných skutočnostiach, ktoré sa opätovne využívajú v interakcii, v správach a reportoch pre manažment poisťovne a regulátorne orgány. Cieľom tohto príspevku je uviesť čitateľa do problematiky Big Data v aktuárstve a využitia v tomto kontexte jazyk R a Apache Hadoop.

ABSTRACT

It can be argued that data are clearly one of the most valuable assets in any sphere of business. A considerable amount of insurance data about clients, contracts, insurance events, claims, assets, liabilities and other events that are re-used in interactions, news and reports for the insurance company management and the regulatory authorities are increasing daily. The aim of this paper is to introduce the reader to the issue of Big Data in actuarial science and the use of R language and Apache Hadoop in this context.

KLÚČOVÉ SLOVÁ

big data, Apache Hadoop, jazyk R, aktuárstvo

KEY WORDS

big data, Apache Hadoop, R language, actuarial science

1. ÚVOD

V súčasnosti zažíva používanie Big Data značný rozmach v rôznych odvetviach. Žijeme v dobe, keď je táto technológia masívne používaná spoločnosťami na každodennej báze a často si ich použitie ani neuvedomujeme. Použitie Big Data nie je len doménou technologických firiem, ale vo veľkom ich využívajú aj banky, štátne inštitúcie či poisťovne. V čase, keď sú tieto dáta považované za štvrtý výrobný faktor, ich spoločnosti používajú viac ako kedykoľvek predtým. Získavajú tým nielen prehľad o internom prostredí firmy, ale najmä o svojich zákazníkoch, ich správani a aktuálnych trendoch. Správne použitie týchto dát znamená konkurenčnú výhodu a môže mať vplyv na budúcnosť firmy ako takej. Aby sa tieto veľké a neustále meniace objemy dát dokázali efektívne analyzovať, sú potrebné nové prístupy a platformy. Veda, ktorá sa zaoberá získavaním pridanej hodnoty z týchto dát, sa nazýva Data Science. Ide o jedno z najperspektívnejších odvetví v súčasnosti, ktoré sa neustále rozvíja veľkou rýchlosťou. Zaoberá sa analýzou, čistením, spracovaním dát a získavaním nových poznatkov, ktoré spoločnostiam umožňujú prijímať rýchlejšie a lepšie rozhodnutia. Poisťný trh a spoločnosti na ňom pôsobiace si tieto

skutočnosti veľmi dobre uvedomujú, a preto hľadajú stále nové možnosti ako skvalitniť existujúce a už implementované produkty. Pri aktuálnom trende možno niektorých aktúrov v poisťovniach považovať aj za dátových vedcov. Ich úlohou je neustále sledovať vývoj prebiehajúceho stavu a v prípade zmien v dostatočnom predstihu informovať a prispôbiť sa danej situácii. S novými možnosťami, ktoré Big Data prinášajú sa otvárajú nové možnosti aj aktúrom. Nové zdroje údajov, rastúci objem dát či meniace sa správanie zákazníkov – toto všetko sú výzvy, ktorým aktúar čelí. Dôležitým faktorom pri konkurenčnom súboji je používanie Big Data a spôsob ich spracovania. Existuje viacero možností, ako poskytnúť aktúrom prístup do týchto „veľkých dát“ a tým umožniť ich použitie v aktuárskej praxi. Jednu z takýchto možností predstavuje aj integrácia jazyka R a populárnej Big Data platformy Apache Hadoop, ktorá je jedným z najpopulárnejších open source nástrojov používaných na spracovanie Big Data.

2. BIG DATA A AKTUÁRSTVO

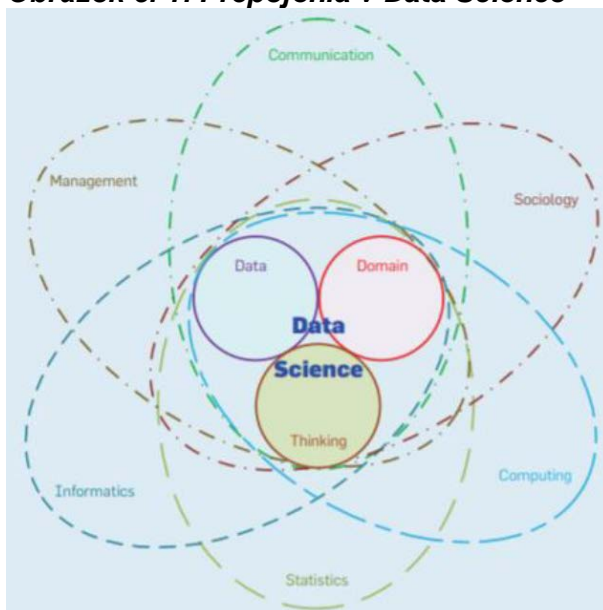
Slovné spojenie Data Science sa v posledných rokoch stáva čoraz populárnejším. Tento pojem už nie je predmetom diskusií výhradne odbornej verejnosti v oblasti informačných technológií, získal si popularitu aj v oddeleniach marketingu, predaja a je úspešným nástrojom pri prijímaní rozhodnutí vrcholového manažmentu. V doslovnom preklade ide o takzvanú dátovú vedu – účelom tejto počítačovej vedy je extrahovať zmysluplné poznatky z dát a následne ich ďalej efektívne využívať tak, aby nám priniesli určitú hodnotu. Avšak vnímať Data Science výlučne ako počítačovú vedu by nebolo celistvé. Odborníci na túto problematiku sa zhodujú, že ide o multidisciplinárny vedný odbor, ktorý využíva relevantné poznatky z oblastí informatiky, štatistiky, komunikácie, sociológie a manažmentu na analýzu dát. Tieto poznatky vznikajú analýzou dát, často hovoríme o Big Data, pričom dáta pochádzajú z rôznych informačných zdrojov a platforiem. Ide o dáta štruktúrované aj neštruktúrované, formalizované aj neformalizované a pochádzajúce z rôznych zdrojov – z firemných databáz, zo senzorov, z internetu či prieskumov. Na základe uvedomenia si týchto skutočností si môžeme utvoriť predstavu o komplexnosti a náročnosti tejto vednej disciplíny. Rozmach Data Science priamo súvisí s rozmachom Big Data, keďže spoločnosti si začali uvedomovať, že síce disponujú obrovským množstvom dát, no bez ich efektívnej analýzy a využitia nemajú žiadnu pridanú hodnotu pre spoločnosť. Uvedomenie si tejto skutočnosti viedlo spoločnosti k rozhodnutiu investovať značné finančné prostriedky práve na získanie tejto pridanej hodnoty. O dátach sa v súčasnej dobe hovorí ako o štvrtom výrobnom faktore popri tradičných troch – práci, pôde a kapitáli. Spoločnosti praktizujú tzv. Data Driven Decision Making – čo predstavuje rozhodovanie sa na základe údajov, teda rozhodnutia ktoré máme podložené dátami sú výhodnejšie, ako spoliehanie sa na intuíciu. Napríklad, marketingový manažér sa nebude spoliehať na svoje skúsenosti a odhad pri výbere typu reklamy, ale využije poznatky zistené na základe analýzy dát ako zákazníci reagujú na daný typ reklamy. Tieto metódy umožňujú spoločnostiam lepšie porozumieť tomu, čo zákazník preferuje a očakáva. Spoločnosti sa ďalej zameriavajú na to, podľa čoho sa rozhoduje, a prispôbujú sa tomu, čo im umožňuje získať konkurenčnú výhodu. Pre Data Science pozri [1], [12], [13].

Najznámejšiu a dodnes zaužívanú metódu charakterizujúcu Big Data zaviedol v roku 2001 Douglas Laney, viceprezident analytického tímu firmy Gartner. Ide o známu teóriu „troch V“. Neskôr sa k tomuto modelu pridali ešte ďalšie. Sú to začiatkové písmená anglických pojmov:

- **Data volume** (*objem údajov*),
- **Data velocity** (*rýchlosť údajov*),
- **Data variety** (*rôznorodosť údajov*),
- **Data veracity** (*pravosť údajov*),
- **Data value** (*hodnota údajov*),
- **Data volatility** (*prchivosť údajov*),
- **Data vulnerability** (*zraniteľnosť údajov*).

Špecialista, ktorý sa zaoberá Data Science sa nazýva data scientist – dátový vedec. Tento termín sa objavuje približne od roku 2008, odkedy sa na potrebu analýzy a spracovania dát začal klásť čoraz väčší význam. Ako môžeme vidieť na obrázku č. 1, výkon tohto povolania si vyžaduje znalosť mnohých vedeckých disciplín. Zároveň je však v súčasnosti považované za jedno z najžiadanejších povolání, v USA bola táto profesia označená za najatraktívnejšiu štyri roky po sebe [5]. Data scientist by mal okrem znalostí technológií disponovať nasledujúcimi schopnosťami: poznať špecifiká daného priemyslu, objasniť získané poznatky, rozumieť aktuálnej firemnej stratégii a podieľať sa na jej tvorbe a svojou prirodzenou zvedavosťou hľadať odpovede na aktuálne otázky. Z technologického hľadiska by mal dátový vedec poznať nasledujúce oblasti: rôzne programovacie jazyky (R, Python...), technologické platformy (Apache Hadoop, non-SQL databázy,...) a sledovať aktuálne technologické trendy (Cloud Computing,...).

Obrázok č. 1: Prepojenia v Data Science



Zdroj: [2]

Z opisu Data Science a profilu samotného dátového vedca vieme, že medzi jeho hlavné schopnosti a znalosti určite patrí znalosť kvantitatívnych metód, tvorba predikčných modelov, programovanie a hľadanie súvislostí v dátach a následné odporúčania a prezentácie jeho znalostí. V oblasti poisťovníctva je s oblasťou dátovej vedy spojená aj práca aktuára (pozri aj [8]). Najmä s rozmachom Big Data tieto dve profesie budú čoraz viac prepojené – potenciál Big Data v oblasti poisťovníctva a aktuárstva rozhodne existuje. Napriek tomu, stále tu môžeme pozorovať rozdiely – hlavné rozdiely medzi týmito povolaniami sú uvedené na obrázku č. 2.

Obrázok č. 2: Actuarial Science vs. Data Science

	Actuarial Science	Data Science
Basics	Mathematical Basics	
Data	Small Data	Small and Big Data
	Structured & static Data	Unstructured & dynamic Data
	Internal Data	External Data
Mathematics & Statistics	Probability Theory	Computational Statistics
	Life and Non-Life Insurance Mathematics	Algorithm
	Quantitative Risk Management	Information Theory
Computer science		Machine Learning & Visualisations
		Numeric Optimization
		Data Management
Programming languages	SAS, S Plus, R	Python, R
	SQL	SQL
	Excel / VBA	Julia, Spark, Scala
Domains/ business knowledge	Reserving, Pricing	
	ERM, ALM, Solvency	
	Accounting, Economics, Law	

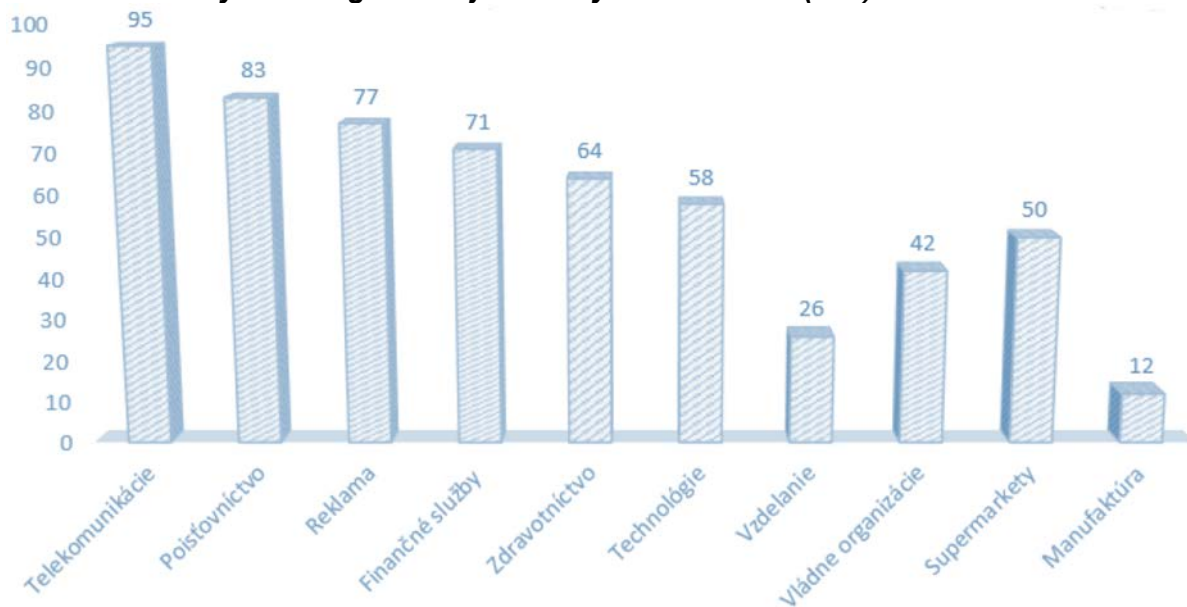
Zdroj: Swiss Association of Actuaries

Dátový vedec je viac programátorsky zameraný a medzi jeho znalosti patria aj rôzne no-SQL databázy, Hadoop či práca s neštruktúrovanými dátami. Aktuár pri svojej práci analyzuje všetky dostupné a relevantné dáta – historické, ale aj real-time (v reálnom čase). Následne pomocou rozličných techník získava požadované výsledky. Po získaní relevantných výsledkov ich aktuár prezentuje a podieľa sa na tvorbe stratégie spoločnosti. Pri svojej práci aktuári používajú rôzne matematické a štatistické programy alebo programovacie jazyky. Vďaka rýchlosti rozvoja technológií v poslednom desaťročí prešlo aj toto odvetvie významnými zmenami a nové zdroje dát si vyžadujú efektívnejšie metódy spracovania, čo predstavuje výzvu aj pre aktuárov. Z tohto dôvodu sa v tejto profesii kladie čoraz väčší dôraz na to, aby aktuár ovládal potrebné informačné technológie ako napríklad princípy databáz, programovacie jazyky či efektívne algoritmy na analýzu dát. V súčasnosti je hlavným cieľom aktuárov vytvárať presnejšie a spoľahlivejšie predikcie a modely. Dôraz sa kladie najmä na nájdanie rovnováhy medzi finančným úspechom a akceptovateľným rizikom. Dosiachnutie tejto rovnováhy predstavuje pre poisťovne ideálny stav z hľadiska ich dlhodobej stability (pozri aj [3]).

Aktuári oproti dátovým vedcom majú tú výhodu, že pracujú v oblasti, ktorá im je dôverne známa. Takisto, práca aktuárov je regulovaná legislatívou, nariadeniami a na svoju činnosť musia mať zodpovedajúce vzdelanie. Aktuár môže byť vo svojej profesii aj dobrým dátovým vedcom, no opačne to až tak neplatí. Vďaka aktuálnemu trendu rastu Big Data a rastúcej potrebe spoločností hľadať súvislosti v dátach a odhaľovať nové skutočnosti, tieto dve profesie konvergujú. Môžeme predpokladať, že aktuári v priebehu rokov budú vykonávať najmä úlohu dátových vedcov.

Poisťovníctvo ako odvetvie zohráva dôležitú úlohu v živote ľudí a firiem, pretože sa priamo zaoberá bezpečnosťou jednotlivcov a ich majetkov, pričom princíp fungovania tohto odvetvia je založený na existencii rizika. Ďalším jeho špecifikom je fakt, že poisťovníctvo nevyrába ani nespravuje žiaden fyzický produkt. V takomto prípade sú dáta pre poisťovne jedným z ich najhodnotnejších aktív. Vo všeobecnosti môžeme konštatovať, že celý princíp fungovania poisťovníctva je založený na predikcii rôznych rizík. Vysoko kompetitívny charakter tohto odvetvia a neustále zmeny správania zákazníkov tvoria predpoklad neustálych investícií do nových, efektívnych a presných spôsobov, ako predpovedať toto správanie a minimalizovať ich riziko. Všetky tieto skutočnosti naznačujú a vytvárajú predpoklady na použitie Big Data v poisťovníctve. Ako máme možnosť vidieť na obrázku č. 3, toto odvetvie si už v súčasnosti vo veľkom osvojilo používanie Big Data na svoju činnosť.

Obrázok č. 3: Využitie Big Data v jednotlivých oblastiach (v %)



Zdroj: [7], podľa Dresner Advisory Services

Jednou z najpoužívanějších techník pri práci aktuára je prediktívne modelovanie a analyzovanie. Ide o metódu, ktorá umožňuje určiť, čo sa udeje v budúcnosti na základe znalosti historických údajov. Následne sa vytvárajú modely založené na rôznych vzťahoch medzi jednotlivými premennými, ktoré boli zistené v týchto údajoch. Prediktívne modelovanie je, ako sme už uviedli, kľúčovým nástrojom Data Science a Big Data, pričom poisťovníctvo predstavuje odvetvie, ktoré tieto techniky dokázali veľmi rýchlo adaptovať. Táto metóda sa používa v aktuárskych analýzach najmä v nasledujúcich oblastiach:

- **posúdenie a hodnotenie rizík** – risk assesment; jedným z najdôležitejších oblastí pre poisťovne je stanovenie výšky poistného. Používa sa v životnom

a neživotnom poistení, napríklad pri havarijnom, zdravotnom poistení a poistení domácností. Poistovne dokážu využívať telematiku, IoT (Internet of Things) zariadenia či iné smart zariadenia na sledovanie zákazníkov a tým predpovedať dané riziko. Pomocou prediktívneho modelovania dokážu poistovne identifikovať pravdepodobnosť havárie, odcudzenia automobilu alebo inú poistnú udalosť pomocou kombinácie behaviorálnych dát a iných externých faktorov (ako napríklad počasie, stav vozovky či charakteristiky obytnej štvrte). V životnom poistení zase fitness náramky či smart hodinky umožňujú sledovať správanie a zvyky zákazníkov, čo umožňuje analýzu rizika na základe ich návykov;

- **odhalenie poistných podvodov** – fraud detection; dátová analýza spolu s kombináciou prediktívneho modelovania zlepšujú a umožňujú odhaliť kriminálnu aktivitu. Sledovaním premenných v každom poistnom plnení a ich následným porovnaním s poistnými podvodmi poistovne dokážu vytvoriť profil podozrivých nárokov na poistné plnenia. Pri odhalení zhody sa takýto prípad automaticky označí na ďalšie, podrobnejšie preskúmanie;
- **informácie o zákazníkoch** – customer insights; získavanie komplexného pochopenia správania, zvykov a potrieb zákazníkov z rôznych zdrojov je veľmi dôležité, aby poistovne mohli predvídať budúce správanie, správne produkty a identifikovať správne segmentácie zákazníkov. Informácie získané z call centier, e-mailov, sociálnych sietí, užívateľských fór či správanie zákazníkov na stránkach spoločností umožňuje poistovateľom vytvorenie unikátneho profilu jednotlivého zákazníka. Získanie prehľadu o zákazníkoch pomocou Big Data poskytuje predpovede nielen o tom, kedy zákazník pravdepodobne odíde ku konkurencii, ale formuje politiku zákazníka. To môže napomôcť poistovateľom rozvíjať dôverné vzťahy a integrovať zákazníkov správnym spôsobom a s presnými informáciami;
- **marketing** – poistovne môžu sledovať správanie spotrebiteľov v rôznych formách a vytvárať ciele reklamy, zbierať informácie pomocou cookies a iných mechanizmov či budovať modely dopytu, aby sa mohli zamerať na tých, u ktorých je pravdepodobnejšie, že si kúpia daný produkt. Tieto aktivity môžu pomôcť k zníženiu nákladov na reklamu, čo vedie aj k zníženiu celkových nákladov, a marketingové náklady prerozdeliť na iné účely;
- **vývoj produktov** – prediktívna analýza umožňuje nájsť nové oblasti pôsobenia a navrhnúť nové produkty pripravené na mieru klientov či zlepšiť už existujúce. Poistovne môžu potom ponúknuť lepšie produkty založené na analýze predchádzajúcich údajov o poistení, zdravotných záznamoch, spôsobe riadenia vozidla a životného štýlu.
- **automatizácia procesov** – poistovatelia používajú automatizáciu najmä pri jednoduchých a opakujúcich sa úkonoch, ktoré nevyžadovali ľudskú interakciu ako napríklad kontrola povinných polí či formátu dát. S nárastom Big Data je možné automatizovať aj komplexnejšie a zložitejšie úkony, ako upisovanie produktov, posúdenie majetku, overovanie nároku na plnenie, či odhalenie podvodov. Posunom k inteligentnejšej automatizácii spoločnosti môžu ušetriť obrovské množstvo času a finančných prostriedkov pomocou strojového učenia (Machine Learning);
- **data driven decision making** – prediktívne modelovanie dokáže napodobniť ľudské rozhodovanie na vytváranie nových pravidiel, ktoré sú lepšie ako tie používané predtým. Tým pádom sú získané výsledky mapované rýchlejšie a s väčšou presnosťou. Toto modelovanie prináša nesporné výhody, ale aj

určité obmedzenia – modelovanie ľudského správania môže spôsobiť, že v algoritmoch ostanú zakomponované ľudské predsudky.

Spomínané techniky je možno využiť v životnom aj neživotnom poistení. Rozvoj Big Data umožňuje nielen použitie týchto techník, ale prináša poisťovniam aj viac údajov, čo napomáha skvalitňovať ich produkty a mať lepší prehľad o zákazníkoch. V oblasti životného poistenia patria medzi tradičné zdroje dát najmä interné záznamy poisťovní, kde patrí napríklad vek poistenca, pohlavie, údaje o zdravotnom stave, či sociálnom statuse a iné. Nové, rozvíjajúce sa dáta pre životné poistenie zahŕňajú verejné záznamy, kriminálnu históriu, demografické údaje, genetické informácie, elektronické lekárske záznamy, históriu predpisov liekov, životný štýl a dáta z nosičov, ako sú napríklad smart hodinky a fitness náramky. Niektoré z týchto údajov sa používajú pred uzatvorením poistnej zmluvy, iné počas a po uzavretí ako určitá forma kontroly. Dáta zo sociálnych sietí ako Facebook, Instagram či Snapchat môžu odhaliť napríklad fajčenie alebo užívanie alkoholu, či iné dôležité informácie o zdravotnom stave poistenca.

Priekopníkom v tejto oblasti je poisťovňa John Hancock, ktorá ako prvá zaviedla zľavu na poistnom pre používateľov fitness náramkov Fitbit. Ide o náramok, ktorý je prepojený s internetom a odosiela údaje o fyzickej aktivite zákazníka. Náramok svojim klientom zadarmo zabezpečí poisťovňa. Spoločnosť uvádza, že klienti zapojení do tohto programu môžu získať zľavu na životnom poistení až do výšky 15 percent. Dáta sú pravidelne odosielané cez internet a používatelia tohto programu dostávajú body za svoju fyzickú aktivitu. Následne môžu túto aktivitu sledovať a získané body premeniť na rôzne zľavy a iné odmeny. Podľa prieskumu spoločnosti Accenture, tretina poisťovní v Severnej Amerike má vo svojom portfóliu podobný produkt, založený na sledovaní fyzickej aktivity spotrebiteľa [7].

Nové zdroje dát sa objavujú aj v neživotnom prostredí – pri poistení majetku alebo úrazovom poistení. K tradičným zdrojom dát v tejto oblasti patrili napríklad údaje o vodičovi, počet rokov skúseností, typ vozidla, počet najazdených kilometrov ročne (havarijné poistenie), typ konštrukcie budovy, geografická lokalita, druh zabezpečenia budov proti ohňu a vlámaniu (poistenie majetku), či údaje o oblasti podnikania. Nové zdroje dát v tejto oblasti sú:

- dáta o počasí, kriminalite, hustote obyvateľstva, hustote premávky, informácie zo sčítania obyvateľstva pre všetky druhy poistenia,
- telematické zariadenia v automobiloch, ktoré uľahčujú získať detailné informácie o správaní vodiča – tieto údaje sa používajú pri poistení automobilov pre firmy a jednotlivcov,
- mobilné dáta, dáta zo zariadení IoT a iných technológií, napríklad z inteligentných domov, najmä pri poistení majetku a podnikania,
- dáta zo smart zariadení, mobilov a fitness náramkov, elektronické zdravotné záznamy, záznamy o činnosti na internete, napríklad história hľadaných výrazov pri úrazovom poistení. Do tejto kategórie patria aj údaje o genofonde zákazníka a genetických predispozíciách na základe analýzy DNA, čo umožňuje genetické profilovanie.

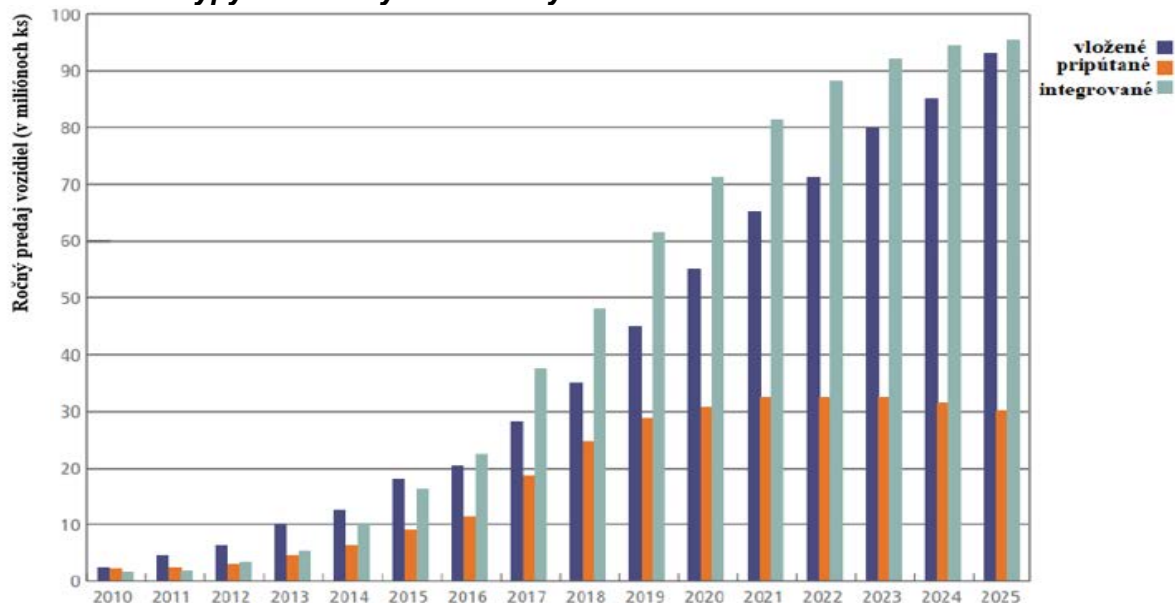
Ďalším rozmáhajúcim sa trendom v oblasti neživotného poistenia motorových vozidiel je tzv. Usage Based Insurance (UBI), známe aj ako Pay As you Drive (PAYD) alebo Pay How you Drive (PHYD), čiže poistenie založené na používaní alebo „platiť ako jazdíš“. Základnou myšlienkou UBI je, že správanie vodiča je

monitorované priamo počas toho, kedy dotyčná osoba jazdí, čo umožňuje poisťovateľom získať komplexnú predstavu o spôsobe jazdy a tým poskytovať individuálny prístup ku klientom. UBI je založené na telematike – ide o vedu, ktorá sa zaoberá diaľkovým prenosom a spracovaním dát. Najčastejšie používanou v praxi je dopravná telematika, pri ktorej ide o prenos údajov o vozidle do vzdialeného počítača, pričom sa využíva technológia GPS. Medzi najčastejšie zbierané údaje patrí počet kilometrov, poloha, rýchlosť, indikácia rýchleho zrýchlenia a spomalenia, použitie smeroviek, spôsob prechodu zákrut, použitie airbagov a iné. Úroveň zozbieraných údajov vo všeobecnosti závisí od typu použitej technológie a ochoty poistencov odosielať osobné údaje. Poisťovňa následne tieto údaje analyzuje a určí výšku poistného. Napríklad vodič často jazdiaci dlhé trasy, vo vysokých rýchlostiach bude mať vyššie poistné ako vodič jazdiaci krátke trasy pri nízkych rýchlostiach. Spôsob zberu týchto údajov sa neustále vyvíja – v začiatkoch (približne od roku 2005) sa do automobilov montovali prídavné zariadenia, ekvivalenty čiernych skriniek v leteckom priemysle, následne sa používali zariadenia na princípe USB, ktoré sa vložili do auta a v súčasnosti prevládajú mobilné aplikácie. Tento princíp v dnešnej dobe už používajú mnohé poisťovne, ako napríklad Allstate, Metromile, Root Insurance (USA), AIOI (Japonsko), Generali (EÚ), Wubi (Česká republika) či Allianz (Nemecko). Použitie tohto typu poistenia prináša poistencom viacero benefitov, najmä tým, čo ročne najazdia menej kilometrov – namiesto fixnej platby platia len za prejdené kilometre. Americká spoločnosť Metromile na svojich stránkach uvádza úsporu až 611 dolárov pri prejení 5 000 míľ (8 046 km) ročne a vo všeobecnosti dokážu vodiči s nízkym počtom kilometrov ušetriť až 50 percent za rok. [4]

Implementácia a využitie telematiky prináša do tohto segmentu poistenia veľa výhod. Poisťovne získavajú komplexný obraz o spôsobe jazdy, správaní sa zákazníkov a reálnom využití vozidiel a tieto údaje môžu využiť aj jednotlivé podniky a to prehľadom o využití firemných vozidiel a správania sa zamestnancov k nim. Zákazníkov zase láka vidina úspory na poistnom – okrem príležitostných vodičov získavajú benefit aj mladí vodiči, pretože pri klasickom spôsobe poistenia často platia vyššie poistné z dôvodu ich nízkeho veku a skúseností. Asociácia GSMA združujúca mobilných operátorov rozlišuje tri spôsoby implementácie v tejto oblasti:

- **vložené (zabudované)** – ide o inštaláciu prídavného zariadenia do motorového vozidla. Tieto zariadenia poskytujú poisťovne a zber údajov sa uskutočňuje pomocou týchto mechanizmov. V súčasnosti sú to malé skrinky, ktoré sú namontované fixne do vozidla alebo zapojené do palubnej dosky;
- **pripútané** – vodič vozidla používa mobilný telefón na odosielanie informácií pomocou mobilných dát, wifi či Bluetooth. Nevýhodou tohto typu je že je závislé od mobilného zariadenia, čo môže spôsobiť vyššie poplatky a takisto stratu dát v prípade výpadku mobilnej siete či zariadenia;
- **integrované** – aplikácie sú priamo integrované do automobilov. Ide o inteligentné automobily, ktoré obsahujú internetovú komunikáciu, pričom odpadá potreba používať mobilný telefón používateľa (napríklad systémy ako Apple CarPlay). GSMA predpokladá, že práve integrované riešenia budú v budúcnosti dominovať (pozri aj obrázok č. 4).

Obrázok č. 4: Typy telematiky v motorových vozidlách



Zdroj: [6]

Použitie telematiky so sebou takisto prináša aj určité nevýhody, plynúce najmä z nutnosti zdieľania dát s poisťovateľom. V prípade inštalácie zariadenia alebo použitia mobilnej aplikácie na zber týchto údajov, poisťovňa dostane všetky relevantné informácie, čo veľa zákazníkov považuje za narušenie ich súkromia. Keďže sa odosielajú údaje o polohe, spôsobe jazdy a rýchlosti, analýza týchto dát nemusí priniesť výhody, práve naopak – pre určitých zákazníkov môže viesť k zvýšeniu poistného (napríklad v prípade agresívneho štýlu jazdy) alebo intenzívneho používania motorového vozidla. Vo všeobecnosti sa PHYD poistenie viac oplatí vodičom s nízkym počtom najjazdených kilometrov ročne.

Používanie nových dátových zdrojov v neživotnom poistení nielenže zlepšuje segmentáciu zákazníkov, ale otvára aj nové možnosti poisťovniam, keď vďaka možnosti spracovania dát možno v reálnom čase dosiahnuť presnejšie výsledky. Až do rozvoja technológie Big Data sa aktuári pri výpočtoch spoliehali najmä na historické dáta. Klimatické zmeny sa v súčasnosti dejú rýchlejšie ako kedykoľvek predtým a globálna klíma má odlišný charakter ako pred rokmi. Tieto zmeny predstavujú nové riziko pre poisťovne, a preto je dôležité ich identifikovať a odhadnúť aktuálne trendy. Pomocou Big Data môžu poisťovne tieto skutočnosti odhaliť vďaka použitiu veľkého množstva senzorov, ktoré zachytávajú informácie o počasí, barometrickom tlaku, teplote, zmeny v prúdeniach vzduchu a podobne.

Dôležitým aspektom pri použití dát z nových zdrojov je ich kvalita. O kvalite týchto údajov pojednáva aj smernica EÚ s názvom Solventnosť II. Pre viac informácií o tejto problematike pozri [10]. Použitie Big Data v aktuárstve a poisťovníctve nesporne prináša veľa výhod a spoločnosti získavajú lepší prehľad o produktoch či zákazníkoch – avšak treba brať do úvahy aj etický aspekt použitia týchto údajov. Aktuári majú povinnosť vyplývajúcu z ich profesionalizmu dodržiavať povest' poistno-matickej profesie a nesú zodpovednosť voči verejnosti v novovznikajúcej oblasti Big Data. Dôležitou súčasťou je dodržiavanie regulácií a zákona. V mnohých situáciách majú aktuári jedinečný prehľad o výsledkoch a dôsledkoch použitíach týchto dát a musia byť ochotní s schopní tieto poznatky kľúčovým zainteresovaným

stranám (napríklad orgánom dohľadu, audítorom, verejnosti, či exekutíve) vysvetliť. Vysvetlenie svojich záverov je jedným z kľúčových atribútov poisťných matematikov. Odborný posudok aktúrov je rozhodujúci pri využívaní Big Data v poisťovniach. Posudok sa týka nielen pridanej hodnoty, ktorú Big Data prinášajú, ale aj etickosti ich využitia, pričom treba dodržiavať všetky regulácie tohto odvetvia a príslušné zákony jednotlivých krajín – napríklad v krajinách Európskej únie ochranu osobných údajov všeobecné nariadenie GDPR (General Data Protection Regulation). Toto všeobecne záväzné nariadenie vstúpilo do platnosti v roku 2018, pričom povinnosť dodržiavať ho majú všetky spoločnosti a organizácie spracúvajúce osobné údaje a pôsobiace v krajinách EÚ28, poisťovne nevynímajúc. Spoločnosti, ktoré nesplnia podmienky tohto nariadenia alebo nezabezpečia bezpečnosť osobných údajov, môžu dostať pokuty až do výšky niekoľko miliónov eur. Nové dáta ako údaje z fitness, zo sociálnych sietí, z automobilov a iných zdrojov sa dajú ľahko zneužiť. Odhalenie zneužitia údajov môže viesť k podniknutiu právnych krokov proti poisťovniam, pokutám, zákazu činnosti a poškodeniu dobrého mena spoločnosti. Je preto na individuálnom posúdení aktára, vedenia spoločnosti a iných zúčastnených strán, ako a do akej miery budú tieto nové poznatky používať.

3. JAZYK R A APACHE HADOOP

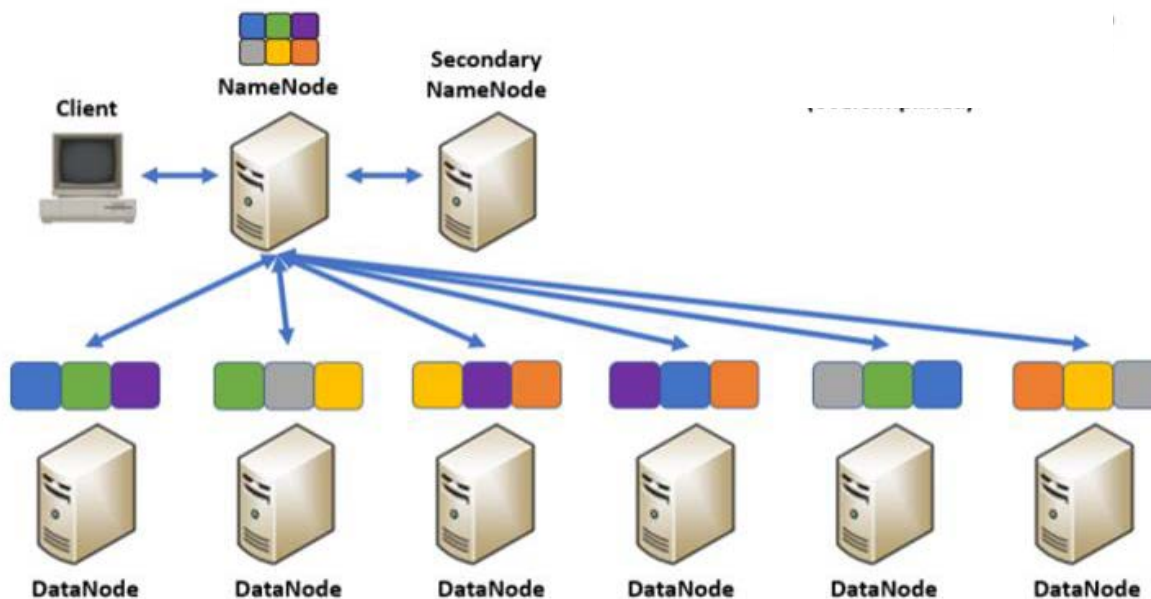
Jazyk R umožňuje používanie širokej škály štatistických a grafických techník, ako napríklad regresná analýza, analýza časových radov, štatistické testy, zhľukovanie a i. Všetky tieto skutočnosti z neho robia ideálnu voľbu pre Data Science, analýzu Big Data a Machine Learning. Nakoľko aktár vo svojej praxi používa spomenuté techniky, nie je prekvapujúce, že R je populárnym nástrojom v tejto oblasti. Pre komplexné informácie o jazyku R a jeho využití v aktuárstve pozri [9].

Hadoop predstavuje open source projekt od spoločnosti Apache. Ide o softvérový framework, ktorý dokáže spracovať rozsiahle súbory v rámci počítačového klastra pomocou distribuovaného spracovania dát. Jednou z kľúčových charakteristík Hadoopu je jeho škálovateľnosť – dokáže pôsobiť na jednom serveri, ale aj na tisícoch počítačov. V súčasnosti ide o dominantnú platformu na spracovanie Big Data. Obsahuje viacero komponentov, ktoré zabezpečujú fungovanie klastra. Medzi základné komponenty patria:

- **HDFS** (Hadoop Distributed File System) – distribuovaný súborový systém používaný na spracovanie Big data naprieč klastrom. Je určený pre komoditný hardvér a je vysoko odolným voči jeho zlyhaniu. Dáta sú v ňom uložené v blokoch, ktoré sú následne replikované. Je napísaný v jazyku Java.
- **MapReduce** – programový model určený na spracovanie Big Data. Je navrhnutý tak, aby poskytoval spracované výsledky v prijateľnom čase. Pôvodne bol vyvinutý spoločnosťou Google.
- **YARN** (Yet Another Resource Negotiator) – doslova ďalší vyjednávač zdrojov. Jeho úlohou je efektívna alokácia zdrojov klastra na vykonávanie jednotlivých úloh. Apache Hadoop je založený na master/slave architektúre. Klastor pozostáva z viacerých serverov/počítačov, ktoré nazývame nody, respektíve uzly. Zvyčajne obsahuje aspoň jednu NameNode – hlavný server, ktorý riadi a rozdeľuje úlohy ostatným, podriadeným serverom, ktoré nazývame DataNodes. Týchto serverov môže byť v klastri ľubovoľný počet, pričom sú všetky riadené NameNode. Dáta uložené v HDFS naprieč klastrom sú replikované – to koľko krát majú byť replikované, určuje replikačný faktor, ktorý zabezpečí, že v prípade zlyhania jednej z DataNodes nedochádza k strate

údajov. O tom, kde sú jednotlivé bloky uložené má prehľad NameNode. Ak zlyhá jedna z DataNode, nastane automatický replikácia dát medzi ostatné, stále dostupné uzly a tým pádom sa zabezpečí dodržanie replikačného faktora. Informácie o tom, kde sú dáta v HDFS uložené má k dispozícii NameNode v svojej pamäti RAM. Aby sa predišlo jej zlyhaniu, klaster často obsahuje aj sekundárnu NameNode. Tá prevezme úlohu riadenia klastra v prípade výpadku primárnej NameNode a zabezpečí chod klastra. Pozri aj obrázok č. 5.

Obrázok č. 5: Zjednodušená architektúra Hadoop klastra



Zdroj: [11]

V závislosti od replikačného faktora dokáže klaster fungovať aj pri výpadku viacerých Nodes naraz. Po výmene chybných DataNodes a ich nahradení novými sú na tieto servery dáta znova automaticky replikované. To, že niektorá z Nodes zlyhá je očakávané, nie je dôležité či nastane, ale kedy. Tento princíp fungovania klastra nám nielen eliminuje hrozbu straty dát uložených v HDFS a potrebu záloh, ale aj ponúka značnú výpočtovú silu na spracovanie týchto dát. Existuje viacero distribúcií Hadoopu, najväčšími hráčmi na trhu sú spoločnosti Cloudera, Hortonworks a MapR. Všetky uvedené ponúkajú svoje vlastné distribúcie a pokročilé nástroje na manažovanie Hadoop klastrov. Najväčším podielom na súčasnom trhu disponuje Cloudera, ktorá niekedy býva označovaná za prvú komerčnú Hadoop spoločnosť. Ponúka vlastný produkt pod názvom Cloudera's Distribution Including Apache Hadoop (CDH), ktorý okrem už spomínaných komponentov používa množstvo iných nástrojov, ktoré pomáhajú k efektívnemu spracovaniu veľkých dát ako Apache Hive, HBase, Impala, a iné. Dôležitým prvkom CDH je Cloudera Manager – centralizovaný správca celého klastra, v ktorom je možné spravovať, monitorovať a nastavovať jednotlivé služby Apache Hadoop.

Tabuľka č. 1: Knižnice (packages) kolekcie RHadoop

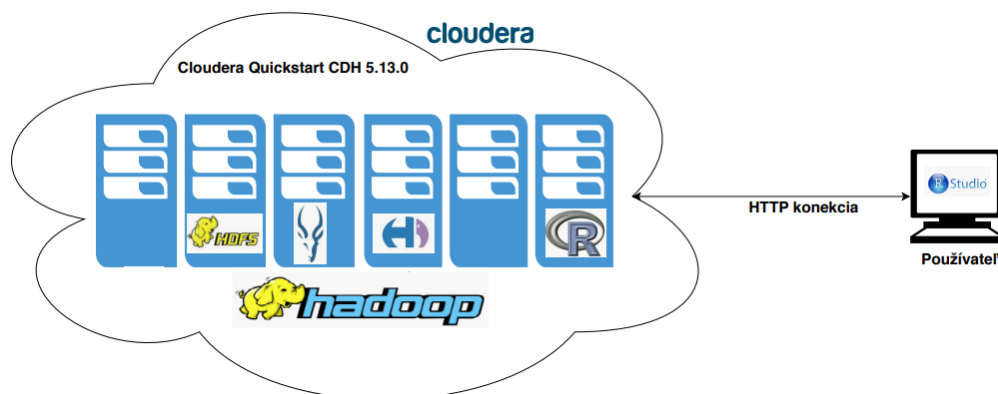
Knižnica	Opis
<i>Rhdfs</i>	Poskytuje základnú konektivitu medzi R a HDFS. Používatelia môžu čítať, zapisovať a modifikovať súbory z HDFS priamo z R. Tento balíček stačí nainštalovať na nóde, na ktorej beží samotný R server.
<i>Rhbase</i>	Poskytuje základnú konektivitu medzi R a databázou HBASE s použitím Apache Thrift servera. Používatelia môžu čítať, zapisovať a modifikovať tabuľky v databáze priamo z R. Tento balíček stačí nainštalovať na nóde, na ktorej beží samotný R server.
<i>Plymr</i>	Umožňuje používateľovi vykonávať bežné dátové operácie na dátach uložených priamo v Hadoope. Vychádza z populárneho balíčka plyr. Tento balíček je nutné nainštalovať na každú nódu v klastrí.
<i>rmr2</i>	Umožňuje vykonávanie analýz s použitím Hadoop MapReduce funkcionality. Tento balíček je nutné nainštalovať na každú nódu v klastrí.
<i>Ravro</i>	Poskytuje možnosť čítania a zápisu súborov typu avro pomocou R z HDFS. Tento balíček stačí nainštalovať na nóde, na ktorej beží samotný R server.

Zdroj: [7]

Ako platformu na spracovanie Big Data v [7] bol zvolený produkt Cloudera Quickstart. Ide o funkčnú simuláciu skutočného Hadoop klastra na jednom počítači. Obsahuje všetky potrebné služby a služby funkcionality skutočného, distribuovaného Hadoop klastra. Ako pracovné prostredie jazyka R bol použitý R Studio server, ktorý je nainštalovaný priamo na serveri, kde beží Cloudera Quickstart. Pre použitie tejto metódy sme sa rozhodli z dôvodu uľahčenia komunikácie Hadoopu a R. Prepojenie lokálneho R Studia so vzdialeným Hadoop klastrom nemá praktický význam a vyžadovalo by si zložité nastavenia komunikácie služieb – použitím R Studio server túto potrebu eliminujeme. Výhodou je aj jeho dostupnosť zo vzdialených počítačov cez URL prakticky z ľubovoľného počítača bez potreby inštalácie jazyka R na lokálnom počítači používateľa. Tento server sa vizuálne nijako nelíši od prostredia R Studio.

Na realizáciu konektie medzi R a Hadoopom sa používa kolekcia knižníc zvaná RHadoop. Skladá sa z viacerých balíčkov, ktoré umožňujú manipuláciu a analýzu dát uložených v Hadoop, ktoré uvádza tabuľka č. 1. Návrh architektúry riešenia zachytáva obrázok č. 6.

Obrázok č. 6: Architektúra možného prepojenia Hadoop a R



Zdroj: [7]

V súvislosti s Big Data a aktuárstvom, pochopiteľne, existuje viacero možností spracovania dát uložených v prostredí Hadoop. Okrem použitia externých nástrojov ako R a ich spojením s technológiou Big Data. Hadoop ponúka svoje vlastné, integrované riešenie na prístup a prácu s jeho ekosystémom – Hadoop User Experience (ďalej len HUE). Ide o open-source webový UI pre Hadoop, ktorý podporuje prechádzanie, vizualizáciu či dopytovanie dát v klastri. Tento komponent je súčasťou všetkých hlavných distribúcií Hadoopu, prípadne je možné ho dodatočne na klaster nainštalovať. Autentifikácia používateľa prebieha priamo na úrovni HUE, za pridelovanie prístupov je zodpovedný systémový administrátor. V prípade, že používateľ nemá vytvorený priečinok na HDFS, po registrácii v HUE sa priečinok automaticky vytvorí. Ak používateľ existuje, HUE sa zosynchronizuje s už existujúcim priečinkom a bude ho registrovať ako domovský. Po prihlásení do HUE sa spustí krátky tutoriál o jeho základných funkciách.

4. ZÁVER

Oba predstavené prístupy umožňujú spracovanie Big Data podľa potrieb používateľa. Každý z prístupov má svoje výhody aj nevýhody, pričom aj R aj HUE obsahujú veľké množstvo funkcií, ktoré aktuár pri svojej práci dokáže využiť. Hodnotenie oboch prístupov je uvedené v tabuľke č. 2 (1 – najlepšie, 5 – najhoršie). Po vyhodnotení všetkých kritérií máme možnosť vidieť, že obe technológie majú približne rovnaké hodnotenie, každá z nich vyniká v niečom inom. Môžeme konštatovať, že obe prinášajú široké možnosti spracovania Big Data a ďalej je už na stratégií jednotlivých spoločností, resp. rozhodnutí aktuárov, ktoré budú používať. Je logické, že aktuári so silnou znalosťou jazyka R budú preferovať spojenie jazyka R a Hadoopu na svoje analýzy. Pre prácu s HDFS a inými službami je však potrebná aspoň základná znalosť princípov technológií Big Data, preto aktuár pre lepšie pochopenie prepojenia technológií by mal absolvovať rôzne školenia ohľadom Hadoopu. Znalosť systému Linux, resp. aspoň jeho základných príkazov bude takisto veľkou výhodou. Výhodou R je, že aktuár môže použiť všetky funkcie a balíčky, ktoré R ponúka a následne ich aplikovať na Big Data.

Tabuľka č. 2: Komparácia prostredí R a HUE

	R	HUE
Užívateľské prostredie	2	1
Rýchlosť spracovania	2	2
Náročnosť nastavenia a prepojenia s Hadoop	3	1
Komplexnosť riešenia	2	3
Integrácia s inými technológiami	1	2
Variabilita možností spracovania	1	3
Riešenia pre Big Data	3	1
Možnosti rozšírenia funkcionality	1	3

Zdroj: [7]

HUE môžeme považovať za jednoduchšie prostredie z hľadiska nastavení potrebných na zabezpečenie konektivity. Ide o nástroj, ktorý prichádza s väčšinou Hadoop distribúcií ako integrovaný. HUE bolo vytvorené na priame spracovanie Big Data používateľmi, čo znamená, že ponúka rozsiahlejšie a robustnejšie nástroje ako R. Z hľadiska výkonnosti má HUE takisto výhodu oproti R – využívaním spojenej výpočtovej sily počítačového klastra môžeme dosiahnuť vynikajúce výsledky v rýchlosti spracovania dát aj v reálnom čase. Prostredie HUE je veľmi intuitívne a za krátky čas si ho osvoja aj noví používatelia. Ťažiskom práce s touto službou je znalosť SQL. Aktuár si následne môže vybrať, ktorú z Hadoop technológií použije pri svojich požiadavkách. Užívateľsky príjemný front-end služby prináša aj rozsiahle možnosti prechádzania súborov na HDFS a vizualizácie dát. Súbory môžeme jednoducho, na pár kliknutí nahráť, resp. stiahnuť aj z lokálneho počítača. Práca s dátami je takisto jednoduchá a intuitívna – vrátane vytvárania nových tabuliek z databázy. HUE je nepochybne nástrojom, ktorý hrá hlavnú úlohu v spracovaní Big Data v spoločnostiach, aktuárstvo nie je výnimkou. Aktuári teda nájdu v HUE široké spektrum možností, ktoré môžu použiť pri svojej práci.

Ktorú technológiu bude poisťovňa resp. oddelenie aktuárstva používať, závisí od ich preferencie či internej firemnej stratégie. Obe ponúkajú komplexné možnosti práce s Big Data a závisí od rozhodnutia kompetentných osôb, ktorý zo spôsobov použijú v každodennej aktuárskej praxi. Tento príspevok bol spracovaný najmä na základe snahy prezentovať závery získané pri tvorbe práce [7], ktorá bola ocenená Slovenskou spoločnosťou aktuárov. Práca obsahuje zdrojové kódy v R a obsahuje manuál na prepojenie R a Apache Hadoop. Samozrejme, dôležité je aj finančné hľadisko zavádzania daných platforiem v organizáciách, ktorému sa tu špeciálne nevenujeme.

Príspevok bol vytvorený v rámci projektu VEGA č. 1/0120/18 – Moderné nástroje riadenia rizika v interných modeloch poisťovní v kontexte direktívy Solvency II a VEGA č. 1/0647/19 – Moderné nástroje na riadenie a modelovanie rizík v neživotnom poistení.

LITERATÚRA

- [1] BERKLEY. What is Data science [online]. UC Berkley School of Information, 2018. [online]. [cit. 26. 1. 2020]. Dostupné na: <<https://datascience.berkeley.edu>>
- [2] CAO, L.: Data Science: Challenges and Directions. In Communications of the ACM 2017, č. 8, s. 59 – 68. ISSN 000-10782. [online]. [cit. 26. 1. 2020]. Dostupné na: <<https://doi.org>>
- [3] CIPRA, T.: Riziko ve financích a pojišťovnictví: Basel III a Solvency II. Praha: Ekopress, 2015. ISBN 978-80-87865-24-8.
- [4] FINANCIAL WEB. Pros and Cons of Pay as You Drive Insurance. 2016. [online]. [cit. 26. 1. 2020]. Dostupné na: <<https://www.finweb.com/>>
- [5] GLASDOOR. 50 Best Jobs in America. 2018. [online]. [cit. 26. 1. 2020]. Dostupné na: <<https://www.glassdoor.com>>
- [6] GSMA. 2025 Every Car Connected. 2012. [online]. [cit. 26.1.2020]. Dostupné na: <<https://www.gsma.com>>
- [7] MASÁR, J.: Problematika Big Data v súčasnej aktuárskej praxi. Diplomová práca: Ekonomická univerzita v Bratislave, 2019. 69 s.
- [8] PÁLEŠ, M.: Aktuárstvo v režime Solventnosť II (S riešenými príkladmi v jazyku R). Bratislava: Vydavateľstvo EKONÓM, 2016. 189 s. ISBN 978-80-225-4288-3.
- [9] PÁLEŠ, M.: Jazyk R pre aktuárov. Bratislava: Letra Edu, 2019. 349 s. ISBN 978-80-89962-26-6.

- [10] PÁLEŠ, M. Kvalita údajov a jej význam pre aktúarov. In: Slovenská štatistika a demografia, 2019, č. 1, s. 38 – 53.
- [11] SCHNEIDER, J.: Hadoop Essentials. Bratislava: Datavard, s.r.o., 2018, (interný dokument).
- [12] PIERSON, L.: Data Science for Dummies. Hoboken: John Wiley & Sons, Inc., 2017. 365 s. ISBN 978-1-119-32763-9.
- [13] PROVOST, F. – FAWCETT, T.: Data Science And Its Relationship to Big Data and Data-Driven Decision Making. 2013. [online]. [cit. 26. 1. 2020]. Dostupné na: <<https://www.liebertpub.com>>.

RESUMÉ

Informácie hrajú v dnešnom svete veľmi dôležitú rolu. Žijeme v dobe keď sa informácie stali najdôležitejšou konkurenčnou výhodou. V minulosti ľudia neboli obklopení toľkými informáciami ako dnes, a preto dnešná spoločnosť čelí problémom s kvalitou dát, nie s jej nedostatkom. Informácie ovplyvňujú každého z nás v našom rozhodovaní, a čím sú kvalitnejšie, tým nám umožňujú lepšie sa rozhodovať. Kvalitné dáta, z ktorých získame potrebné relevantné informácie, nám umožňujú znižovať straty, zvyšovať ziskovosť, analyzovať rôzne oblasti, jednoducho povedané, vedia nám zefektívniť fungovanie spoločnosti. Big Data priamo súvisia s vedným odborom Data Science. Existujú isté spoločné prieniky medzi povoláním dátových vedcov a aktúarov a v súčasnosti aj niektorí aktuári plnia v poisťovniach úlohu dátových vedcov. Na základe tohto predpokladu možno konštatovať potrebu použitia Big Data v tejto profesii. V súčasnosti už možno uviesť rôzne aplikačné príklady a spôsoby používania Big Data v poisťovníctve. Práca aktuára vyžaduje znalosť mnohých softvérových nástrojov, ktoré im pomáhajú dosahovať želané výsledky. Populárnym sa najmä v posledných rokoch stal jazyk R, ktorý predstavuje výkonnú platformu na spracovanie, analýzu a vizualizáciu dát. Popri jazyku R jeden z najpopulárnejších open source nástrojov používaných na spracovanie Big Data je Apache Hadoop. Možno realizovať porovnanie jazyka R a Hadoop User Experience – ich možností, ktoré pri spracovaní veľkých dát ponúkajú, architektúru a princípy ich fungovania, ako aj dôvody ich popularity. Na základe týchto skúseností možno zvoliť tú platformu, ktorá je ideálnym nástrojom pre prácu aktúarov.

RESUME

Information plays a very important role in today's world. We are living in an era when information has become the most important competitive advantage. In the past, people were not surrounded by as much information as today, and today's society is facing data quality problems, not their deficiency. The information affects all of us in our decision making, and the better it is, the better decisions it enable us to make. High-quality data, from which we obtain the necessary relevant information, provide us with reducing losses, increasing profitability, and analyzing various areas and to put it simply, they can improve the functioning of the society. Big Data is directly related to Data Science. There are some common intersections between the profession of data scientists and actuarians, and at present some actuarians also act as data scientists in insurance companies. The necessity of using Big Data in this profession can be concluded on the basis of this assumption. Various application examples and ways of using Big Data in the insurance industry can already be mentioned. Actuarial work requires knowledge of various software tools helping them to achieve the desired results. Especially in recent years, the R language has become popular as a powerful platform for data processing, analysis and visualization. In addition to the R

language, one of the most popular open source tools used in the processing of Big Data is the Apache Hadoop. Consequently, a comparison of the R language and the Hadoop User Experience can be carried out- their possibilities offered in the processing of big data, architecture and principles of their functioning, as well as reasons for their popularity. These experiences enable the selection of the platform that is the ideal work tool of actuarians.

PROFESIJNÝ ŽIVOTOPIS

Ing. Michal Páleš, PhD., pôsobí ako vedúci Katedry matematiky a aktuárstva Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave. V rámci pedagogickej činnosti vyučuje predmety matematika, teória pravdepodobnosti, softvérové aplikácie pre aktúarov, teória rizika v poistení, úvod do aktuárstva a vybrané kapitoly z matematiky. Vo svojej vedeckej práci sa orientuje na aktuársku vedu, využitie kvantitatívnych metód v ekonómii a softvérovú podporu riadenia rizík (najmä jazyk R). Je autorom viacerých medzinárodne ocenených vysokoškolských učebníc, vedeckých monografií a článkov z oblasti aktuárstva.

KONTAKT

pales.euba@gmail.com