

BIG DATA ANALYTICS FOR INSURANCE¹

Jovica Stanković, PhD

Faculty of Economics, University of Niš, Niš, Serbia,
jovica.stankovic@eknfak.ni.ac.rs

Jelena Z. Stanković, PhD

Faculty of Economics, University of Niš, Niš, Serbia,
jelenas@eknfak.ni.ac.rs

ABSTRACT

Availability of large amount of data from different data sources and implementation of Big Data Analytics (BDA) in insurance industry transformed the insurance product lifecycle. The rapid change is evident throughout the whole insurance product lifecycle: from the product design, underwriting and pricing, their marketing and distribution, through to claims processing and the ongoing management of customer relationships. The adoption of new technologies, especially for data collection and advanced analysis, represents a significant competitive advantage in the insurance market. Therefore, in this paper the impact of the BDA usage on insurance sector is analysed and potential benefits and risks associated with the use of BDA are presented.

KEY WORDS: Big Data Analytics, Digitalization, Insurance.

INTRODUCTION

The process of digitalization enabled usage of new data sources and data sets, with very fast creation of data of enormous size. The data are available from multiple sources. Traditional data set were sourced directly from insureds themselves, but nowadays they are also collected from other sources such as mobile devices, sensors and the internet, both within and outside an insurance context. New data sources became available in the era

¹

of digitalization. These data sets include the Internet of Things (IoT) data (physical activity and health condition obtained from wearable devices, driving behavior (telematics), smart home data etc.), online data (web searches, social media posts, online shopping), insurers' own data (call center logs, digital claim reports, data from mobile apps) and other digital personal and non-personal data. The main challenge is to manage the complexity of data sources, types and speed at which they grow.

Data Science defined as a “systematic study and analysis of various data sources, which understands the meaning of data and uses them as a means for effective decision making and problem solving” (Akerkar&Sajja, 2016) enables the organization to operate in more efficient and productive manner; to identify new opportunities and challenges and to create a strong image of its brand. The purpose of Data Science is to facilitate the implementation of various processes related to data, such as data collection, data cleaning, presentation and evaluation of data, analysis and usage of data to obtain information related to operations. Considering the fact that insurance industry was one of the pioneers in extensive use of information technology, Data Science tools can be ideal for supporting numerous, successive and segmented tasks of insurance lifecycle. The adoption of Data Science has significantly altered the relationship between insurance companies and insureds and contributed to commercialization of insurance services. Therefore, the aim of this paper is to explain the Big Data and Big Data Analytics, with special attention paid to the usage of Big Data in insurance industry.

THE DEFINITION AND CHARACTERISTICS OF BIG DATA

Big Data is the term originated in 2008 and represents “data of the quantities, which exceed the possibilities of commonly used software for the storage, processing and data management” (Dumbill, 2013). Accordingly, the term “Big Data” is used for data sets so large or complex that traditional data processing applications are not adequate for working with them. The term “Big Data”, however is not limited just to the data and to the specific size of the data set, but often refers to the use of predictive analytics, algorithms and artificial intelligence, analysis of user behavior or certain other advanced methods of data analysis to extract hidden value from the data. Therefore, it can be concluded that Big Data, as the ability to process massive amounts of data collected over the Internet by analyzing web logs, data generated during

the transaction or obtained from various sensors, represent the key component of Data Science.

Literature review and industry overview are offering numerous definitions for the term Big Data, but common part of most important and widely accepted is that Big Data are usually described by 4V (John&Misra, 2017): variety, velocity, volume and veracity or 5V (Boobier, 2016): volume, velocity, variety, veracity and value.

Volume – the huge amount, bigger than the amount of data processed by traditional applications, of structured, semi-structured and unstructured data that is available. For many companies the analysis and use of social media data are becoming necessary since billions of users of various social network applications daily produce billions of data. In addition, various applications generate very large amounts of data (various sensors, smart homes, behavior and location monitoring, etc.).

Velocity – There are two kinds of velocity connected to the Big Data definition. The first kind refers to the speed at which data are generated, while the second aspect concerns the possibility for real time processing of large data sets. In an insurance context, perhaps the most obvious example is the whole issue of vehicle telematics information or wearable health monitoring sensors data.

Variety – Variety refers to the diversity of data types and the diversity of data sources. While in the previous period the application was primarily concerned with structured data stored mainly in various types of information systems' transactional databases, today there is increasing need to use sources of unstructured data, such as video content, images, files, data from websites across the Internet, etc. In an insurance context this might include data which are based on weather, location, sensors, and also traditional structured data. For instance, in the case of catastrophe modelling insurers combine policy data, policyholder input (from social media), weather, voice analysis from contact centers, and other key data sources which may affect the model.

Veracity–Veracity refers to the reliability of the data, since all data sources are not equally accurate and reliable. This is particularly important issue of the data originating from different web sites or social networks. In an insurance context this may relate to the amount of “spam” and off-topic posts on a social media, or even fake news and data.

Value –The value of the data can be quantified in different ways. One way is to measure value of the data to the user, while another is to measure the value is the cost of acquiring key data to give that information, for example the creditworthiness of a customer. In an insurance context, one example might be the added value of telematics driving behavior data to improve motor insurance risk profiles and policy pricing.

TYPES OF BIG DATA ANALYTICS

Analytics, or the analysis of data, is generally recognized as the key by which data insights are obtained. Analytics unlocks the ‘value’ of the data.

There are several levels of data analytics:

Descriptive analytics - Analytics which serves simply to report on what has happened or what is happening. It uses data aggregation and data mining to provide the needed insights and enable better understanding of the relationship between customers and products. In insurance, this might relate to the reporting, for example of claims for a given date or period.

Predictive analytics - seeks to predict probabilities of future events. Predictive analytics represents “an application of advanced statistical, information software, or operation research methods to identify predictive variables and build predictive models to identify trends and relationships not readily observed in a descriptive analysis” (Schniederjans et al., 2014). Predictive analytics try to understand the future using different statistical models and forecast techniques, and as a result to answer the question – “What could happen?”. Predictive analytics can be used for making the projection of insurance sales and premium revenue, or for making predictions about which customers are likely to be fraudulent, based on data about fraud behaviours in the past and typical types of claims most likely to give rise to fraud. Analysts typically use regression analysis, predictive models, forecasts, statistical analysis, scoring and other methods.

Prescriptive analytics - Analytics which anticipates future events and outcomes, and makes suggestion - “prescribes” what should be done about it. Prescriptive analytics represents “an application of decision science, management science, and applied mathematical techniques to make best use of allocable resources”(Schniederjans et al., 2014). Prescriptive analytics

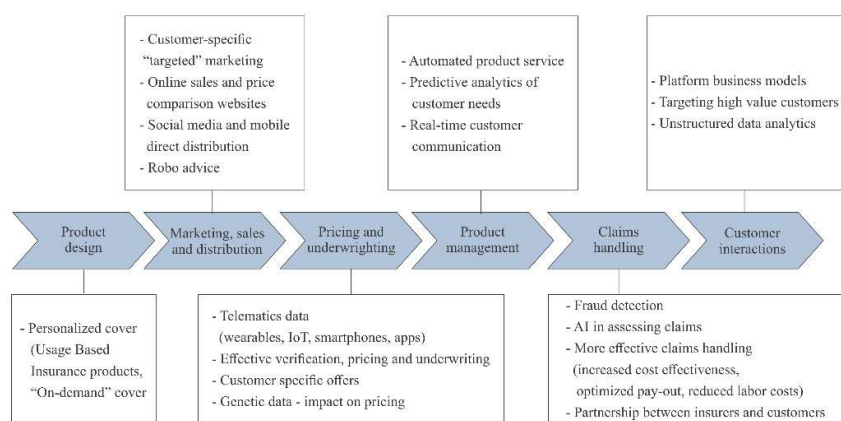
mainly use optimization and simulation algorithms to advise on possible outcomes and try to answer the question “What should happen?”. Prescriptive analytics can be used in various scenarios where there are too many variables, options, constraints and data sets. The prescriptive analytics defines the best possible solution to a specific problem, and recommends one or more courses of action which can be measured and refined based on results. In insurance, P&C insurance providers can use catastrophe modelling price plans for pricing, policy conditions and optimizing portfolios to minimize risk.

Cognitive analytics - Next generation analytics is likely to be “cognitive” in nature, based on machine learning and collection of techniques and algorithms for identifying patterns in large, complex, or high-velocity data sets. It includes sophisticated statistical models, predictive analytics, machine learning, neural networks, text analytics, image analytics, and other advanced data mining techniques (Hurwitz et al., 2015). Considering insurance industry cognitive system can improve personalization of life insurance. Premium value could be determined based on insureds’ personal habits, tracked by wearable devices such as a smart watch, compared with hundreds of thousands of other insured sin order to determine the risk level, which could cut costs as insurance companies better understand actual risks. All these types of analytics are combined with increasing quality of visualization (dashboards, advanced graphics, graphical mapping etc.), which enables users to better understand the data. The same may be said of geospatial analytics – the analytics of place – which incorporates geocoding into the analytical data to give a sense of location in any decision. It is important to emphasize that analytics is not the goal, but what could be achieved with the analytics. Obtained data insights should be combined with best practices, decision making, and strategic and operational judgments.

THE USE OF BIG DATA ANALYTICS IN THE INSURANCE PRODUCT LIFECYCLE

There are a lot of potential usages of big data analytics in various components of the insurance product lifecycle. However, the speed and scope of changes in insurance business caused by the BDA implementation took their own course. The specifics of insurance companies’ operations, especially in terms of data protection and strict regulatory requirements, have increased the risks and negative effects of the insurance products’ innovations (Klapkiv, &Klapkiv, 2017). Nevertheless, recent analysis of the

insurance product lifecycles on more developed markets demonstrates numerous cases of big data deployment. Although the basic structure of the product lifecycle remains the same, BDA provides the possibility to transform the lifecycle and/or change each element of this structure (picture 1).



Picture 1. The use of BDA in insurance product life cycle

Many insurers start the transformation process by using BDA in certain activities of the products' lifecycle, for example, in marketing and sales (Stanković et al., 2022). Behavioral and lifestyle indicators based on insureds' everyday activities can be used to identify specific needs and to anticipate risks of certain insureds' segments. These information enables insurers to reveal the gaps between the current insurance coverage and insureds' requirements. For this purpose, search engine optimization is being increasingly employed, as well as social media updates and location tracking data (OECD, 2020). Targeted marketing campaigns are usually delivered in real time using smart devices, specific websites or social media platforms.

The low frequency and unsatisfactory quality of interactions between insurance market participants (Hocking et al., 2014) made automated selling services, such as robo-advice and price comparison websites, exceedingly important for insureds. However, in conditions of financial instability, direct interaction between the insured and the insurer can enable the building of long-term relationships and trust.

When it comes to product design, the application of BDA enables insurers to create customer specific products based on the different types of data. In

insurance, IoT currently finds its application in the field of motor and health insurance. In motor vehicle insurance, the emphasis is on smart vehicles, which means vehicles equipped with on-board-diagnostics devices that enable monitoring of parameters related to driver behavior or car components in real time (Stankovic, 2014). With the increase in the number of vehicles equipped with telematics monitoring devices, a huge amount of data is provided, which provides many opportunities for designing usage-based insurance products and “on-demand” covers. The basic feature of these products is that the insurance price, i.e., the premium, is determined on the basis of new sets of data, such as: the length of the road traveled, the time of day when driving, the insured's driving behavior and others. The larger data set generated by the use of in-vehicle telematics monitoring devices ensures a precise quantification of risk and an understanding of policyholder behavior. The ever-widening availability and application of devices such as smart watches, bracelets and other wearable devices, which can monitor pulse, blood pressure, blood oxygen level, activity (number of steps, hours of sleep), geolocation, intake of food and fluids in the body, calorie consumption and blood sugar level, made it possible to personalize health insurance products. Such insurance products have a double role: they can be used in the function of risk prevention by rewarding the insured, who have healthy lifestyle and driving habits, by calculating a lower insurance premium, as well as risk reduction in case of health problems and accidents (activating the alarm and calling the emergency service).

BDA usage in insurance has a potential to change the nature of risk due to reduced uncertainty. Using volume, variety and velocity of big data predictive analytics based on granular individual-based approach to pricing provides the possibility to insure previously assumed uninsurable risks at reasonable price (McFall et al., 2020). Moreover, advanced statistical modeling applied to traditional and new data enhanced dynamic insurance premium optimization. The additional benefit of BDA usage for the insureds is the possibility to understand the size of the risk exposure and the appropriate amount of insurance coverage.

BDA plays a major role in the claims management in insurance industry. Matching the claims against the profiles and past claim patterns enables insurers to reveal fraudulent or deviant claims, as well as behavior of the claimant and people and organizations involved in the claim (Ravi, & Kamaruddin, 2017). More efficient claims management results in reduced costs for insurers and insureds. However, BDA implementation in insurance is connected with both, benefits and risks for insurers and insureds.

RISKS AND BENEFITS OF BDA IMPLEMENTATION IN INSURANCE

The BDA implementation in insurance and development of digital technologies create unique opportunities for improving both, economic performances of insurers and insureds' well-being. However, its impact on insurance, as well as other industries, development in long term cannot yet be reliably assessed. Current issues regarding BDA implementation create certain risks (table 1). The most important one is information security. Storage, protection and third-party usage of insureds' data is of vital importance for digital transformation of each activity of the insurance product lifecycle. This is a key challenge to supervisors, who need to define a balanced approach to promote innovations while maintaining the expected level of insureds' protection. Therefore, supervisors may require that insurance companies and intermediaries provide policies and procedures for the protection and use of information on insureds. The fact that insurers may outsource various services, such as cloud services or other platforms, the design of algorithms and provision or sourcing of large volumes of data, in order to develop and organize digitalized processes, opens the issue of third-party risk management. Some jurisdictions limit the number of service providers available to provide these services, while others require extension of insurers' policies and procedures on the use of BDA to third party providers.

The most important risks for insureds are the risks of unfair treatment, discrimination, or concerns over access to, or exclusion from, insurance services. Namely, increasing use of BDA by insurers has the potential to strengthen the insurance market through more competitive pricing, wider availability of coverage and innovative product design.

Table 1. Benefits and risks of BDA implementation in insurance product lifecycle

Benefits / risks	Phase
• <i>Benefits for insured:</i> increased insurability of new risks; design of personalized insurance products • <i>Benefits for insurer:</i> wider customer access; reduced underwriting costs	Product design
<i>Risks:</i> insurance product design may not be transparently and clearly explained to insureds and supervisors to understand its structure and purpose	
• <i>Benefits for insured:</i> communication throughout the life of the policy; improved overall insured experience • <i>Benefits for insurer:</i> personalized marketing and appropriate level of disclosure; reduced marketing costs	Marketing, sales and distribution
<i>Risks:</i> highlighting or limiting particular information to produce certain actions from the insured; risk of passive purchases and of the purchase of a cover that is not needed	
• <i>Benefits for insured:</i> insuring new risk at affordable prices; adjusted prices at a more granular insured-specific level • <i>Benefits for insurer:</i> insurers can identify and assess risks more precisely	Pricing and underwriting
<i>Risks:</i> reduced affordability and inclusiveness of insurance products; potentially reduced revenue flows for insurers	
• <i>Benefits for insured:</i> reduced premiums through reduction of fraud incidence; faster claims payments and enhanced customer experience through automated claims handling • <i>Benefits for insurer:</i> reduced claims payments; reduced operating costs	Claims handling
<i>Risks:</i> violation of contextual integrity; risk of unwarranted secondary use of data; risk of increase in unsubstantiated rejections and unlawful discrimination in the claims decision making process due to over reliance on generic claims settlement models	

Source: According to IAIS (2018, 2020)

On the other hand, innovations based on BDA could reduce the number of insurance providers, because only insurers with strong insureds bases and

investment capabilities can monetize access to insureds data. This could potentially create privileged positions for one or a few companies resulting in limited options for insureds. Therefore, in respect of product design, marketing and sales, due attention needs to be given to achieving fair customer outcomes in terms of products that meet consumers' needs, the design and use of algorithms and the use of customer data. Digital innovations can change and potentially improve the customer experience and reduce insurers' operating costs, but appropriate and responsible use of new technologies should be provided.

CONCLUSION

The coming years are moving towards the explosion of the Internet of Things (IoT) and artificial intelligence that will enable insurance companies to gain competitive advantage by improving the quality of the relationship with the insureds, increasing the margins and offering tailor made services. Taking into account the impact of Big Data Analytics on the transformation of the insurance product lifecycle, it can be concluded that the changes relate to: (1) the way BDA changes the interaction between insurers and insureds, which can now also take place through social networks and virtual advisors; (2) how BDA can be used to automate, standardize, and improve the efficiency of business processes through, for example, online sales and digitization of claims records; and (3) the possibilities that BDA provides for the modification of existing and the development of new insurance products.

The basis for BDA implementation in insurance product lifecycle in terms of data and a certain level of technical equipment exists in most insurance companies. However, certain activities, such as management of information systems, risk management, sales, are becoming more important, and their realization requires the engagement of external experts or companies. The adoption of new technologies, especially for data collection and advanced analysis, has the potential to strengthen the insurance market through more competitive pricing, wider availability of coverage and innovative product design. However, the advancements in BDA and its implementation in insurance have to promote ultimately consistent delivery of fair outcomes to insureds. The fairness of outcomes must be achieved irrespective of the adoption of new technologies or other innovations by insurers, including the use of BDA. Therefore, supervisors have adopted various recommendations and guidelines setting out numerous policy areas that policymakers should

consider in respect of big data usage in insurance emphasizing issues of fairness, ethics, accountability and transparency.

REFERENCE:

1. Akerkar, R. and Sajja, P.S. (2016). *Intelligent techniques for data science*. Cham: Springer International Publishing.
2. Boobier, T. (2016). *Analytics for insurance: The real business of Big Data*. John Wiley & Sons.
3. Dumbill, E. (2013). Making Sense of Big Data. *Big data*, 1(1), p.1-2.
4. Hocking, J., Wood, A., Dally, N., Pan, K., Lin, B., Ban, H., ...& Lee, S. (2014). *Insurance and Technology: Evolution and Revolution in a Digital World*. Morgan Stanley Research Global, The Boston Consulting Group.
5. Hurwitz, J. S., Kaufman, M., & Bowles, A. (2015). *Cognitive computing and big data analytics*. John Wiley & Sons, p.87
6. IAIS (2018). *Issues Paper on Increasing Digitalisation in Insurance and its Potential Impact on Consumer Outcomes*. International Association of Insurance Supervisors.
7. IAIS (2020). *Issues Paper on the Use of Big Data Analytics in Insurance*. International Association of Insurance Supervisors.
8. John, T., & Misra, P. (2017). *Data lake for enterprises*. Packt Publishing Ltd.
9. Klapkiv, L., & Klapkiv, J. (2017). Technological innovations in the insurance industry. *Journal of Insurance, Financial Markets and Consumer Protection*, 26 (4/2017), 67-78. 1
10. McFall, L., Meyers, G., & Hoyweghen, I. V. (2020). The personalisation of insurance: Data, behaviour and innovation. *Big Data & Society*, 7(2), 1-11.
11. OECD (2020). *The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector*
12. Ravi, V., & Kamaruddin, S. (2017). Big data analytics enabled smart financial services: opportunities and challenges. In *International Conference on Big Data Analytics* (pp. 15-39). Springer, Cham.
13. Schniederjans, M.J., Schniederjans, D.G. and Starkey, C.M., (2014). *Business analytics principles, concepts, and applications: what, why, and how*. Pearson Education., p.95
14. Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1), 3-9.
15. Stanković, J., Stanković, J. Z., Tomić, Z. (2022). Factors affecting digital transformation of insurance in the Republic of Serbia. *Economic Themes*, 60(1), 133-149.