

Wstęp

Dziedzina analityki danych obejmuje procesy analizy dowolnych rodzajów danych w celu wydobycia z nich informacji, realizowane z wykorzystaniem specjalizowanych systemów przetwarzania i oprogramowania [*M. Niezgódka*].

Analytyka danych jest podstawą zmniejszania niepewności, redukcji entropii, ale jej oparcie o Big data i wspomaganie przetwarzania skomplikowanymi narzędziami informatycznymi, w tym rozwiązaniami sztucznej inteligencji niesie też nowe zagrożenia.

Skupiając się na rozwoju narzędzi analitycznych i możliwościach ich użycia nie należy zapominać, że podstawą dla prawidłowej oceny świata poddawanego analizie jest poprawne określenie, jakie dane i dla jakiego celu chcemy przetwarzać [*W. Wiśniewski*]. Objaśniając termin Big Data jako „moc danych” podkreślamy, że danych jest mnóstwo, ale też, że ich zasoby kreują dodatkową moc dla wszystkich, którzy je przetwarzają [*B. Szafranski*].

Skalę zjawiska pokazują wyniki wyszukiwania Google, według których w Internecie słowa dane i analityka występują miliardy razy [„data” ponad 13 miliardów, „dane” – blisko pół miliarda, „analytics” blisko 2 miliardy, „analityka” – 7 milionów] a określenia złożone „data analytics” – 80 mln i „analityka danych” – 19 tys. razy. Do tego obszaru odnosi się też określenie „data science” (77 mln razy)]. Określenie analityka danych jest w dotychczasowej polskiej literaturze spolszczeniem angielskich „data analytics” jak i „data science”.

Dla usprawniania procesów decyzyjnych stosuje się analitykę: opisową, diagnostyczną, predykcyjną i preskrypcyjną. Ta ostatnia, zwana też rekomendacyjną, zorientowana jest na określenie działań, które należy podjąć aby wyeliminować przyszły problem lub wykorzystać obecny trend i wymaga danych z różnych źródeł oraz zaawansowanych narzędzi [*P. Wiszniewski*].

Postępuje danetyzacja, to znaczy kwantyfikowanie (liczenie, mierzenie, ważenie) i przekładanie wszelkich aspektów rzeczywistości na dane, by móc poddać je agregowaniu i algorytmizacji. U postaw dążenia do redukcji obszarów które nie są jej poddane leży przekonanie, że dane są kluczem do zrozumienia rzeczywistości [M. Ganczar].

Konkurowanie w świecie opartym na danych wymaga profesjonalistów. Postulowane są różne nowe profile kompetencyjne analityków danych czy badaczy danych [J. Dygaszewicz]. W zakresie znaczenia określeń data analyst i data scientist, ścierają się dwa poglądy. Według pierwszego z nich data scientist, przywoływany bez spolszczenia lub tłumaczony jako „mistrz danych” ma możliwe do zastosowania w każdej dziedzinie szerokie umiejętności, obejmujące m.in. wspieraną przez zaawansowane narzędzia eksplorację szybkozmiennych, nieustrukturyzowanych danych, pochodzących z sieciowych źródeł o nieznannej jakości, co odróżnia go od tradycyjnego analityka danych. Według zwolenników drugiego poglądu, poszczególne cechy „data scientist” odróżniają dobrego analityka danych od złego a jako całość przypisane absolwentowi kierunku data science są nierealistycznym ideałem.

Różnorodność obszarów analityki danych, której przykłady omówiono w kilkunastu rozdziałach tej książki a także potrzeba respektowania w pracach analitycznych wymogów technicznych, prawnych i etycznych [J. Cytorowski], zapewniania cyberbezpieczeństwa i ochrony danych osobowych [A. Gryszczyńska], pokazuje, że dla sprostania nowym wyzwaniom potrzebne są multidyscyplinarne zespoły. Konieczność ich tworzenia stwierdza się też w przedstawionej w sierpniu 2019 r. „Polityce Rozwoju Sztucznej Inteligencji w Polsce na lata 2019–2027” [s. 27].

Rysuje się z jednej strony potrzeba stworzenia możliwości pogłębienia znajomości nowych technik specjalistom zarządzającym danymi, znającym dobrze dziedzinę, której dane dotyczą aby współdziałać z architektami AI i wdrożeniowcami IT, m.in. w optymalizacji procesu uczenia się AI. Z drugiej strony, uwzględniając dotychczasowe trudności informatyzacji sektora publicznego, potrzebne wydaje się, aby osoby mające już teoretyczną i praktyczną znajomość jego procesów informacyjno-decyzyjnych i predyspozycje osobowościowe do realizacji zadań analitycznych dokształcały się w zakresie statystyki i informatyki tak, aby koordynować inicjatywy, rozdzielać zadania a niektóre realizować samodzielnie. Uniwersytet Kardynała Stefana Wyszyńskiego wspólnie z Głównym Urzędem Statystycznym przygotował studia podyplomowe, które mają umożliwić zaspokajanie takich potrzeb, sygnalizowanych zwłaszcza w sektorze publicznym.

Mogą być w zmniejszaniu niepewności przydatne także rozważania zawarte w 33 rozdziałach tej książki, odnoszących się do rozległych obszarów analityki danych, choć należy zastrzec, że podejmowane są w nich z konieczności tylko zagadnienia wybrane. Pozwalają one na ogólne rozpoznanie szans lepszego wykorzystania potencjału współczesnej analityki danych w realizowaniu zadań różnych podmiotów, co poszerza krąg adresatów tej publikacji o kadry kierownicze i inne osoby zainteresowane refleksją nad nowymi tendencjami, w tym np. symptomami zmian roli Chmur.

Seria Internet, której dziewiąty tom oddajemy Czytelnikom, jest wynikiem interdyscyplinarnej współpracy. Obierane dla dorocznych konferencji i monografii Uniwersytetu Kardynała Stefana Wyszyńskiego i Naukowego Centrum Prawno-Informatycznego tematy, okazują się – także tym razem – zgodne z ogłaszanymi później globalnymi Top trendami na dany rok. Zespołowa analiza pomaga rozwijać i wdrażać innowacyjne rozwiązania ale i zachować ostrożność wobec takich nowości, które nie są godne zaufania, przynoszą efekty odmienne od zapowiadanych a ich atrakcyjność szybko przemija.

Inicjatywą Rektora i Senatu UKSW wspierającą dalszy rozwój, konkretyzowaną z upoważnienia Ks. Rektora *Stanisława Dziekońskiego* przez prof. *Irenę Lipowicz*, jest organizacja multidyscyplinarnego centrum ochrony danych. Radzie Naukowej przewodniczy Zastępca Europejskiego Inspektora Ochrony Danych dr *Wojciech R. Wiśniewski* a jej członkowie też są współautorami tej książki. Zaproszenie do współpracy w ramach Centrum kierujemy do szerokiego grona ekspertów, w tym również do naszych Czytelników.

Grażyna Szpor