Zeszyty Prawnicze
20.1 / 2020, s. 297-305
DOI 10.21697/zp.2020.20.1.15

Sprawozdanie z konferencji naukowej „Bezpieczeństwo w Internecie. Analityka danych", Warszawa, 6-7 czerwca 2019 r.

Analityka danych jest stosowana w coraz większej liczbie procesów zarówno w sektorze prywatnym, jak i publicznym. Dla dalszego, bardziej efektywnego i dynamicznego rozwoju jej wykorzystania potrzebne są zmiany legislacyjne i organizacyjne. To główna konkluzja 11. ogólnopolskiej konferencji naukowej z cyklu Bezpieczeństwo w Internecie.

Odbyła się ona 6 i 7 czerwca 2019 r. w Auli Schumana Auditorium Maximum Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie (UKSW). Wśród jej organizatorów, oprócz uczelni, było Naukowe Centrum Prawno-Informatyczne, Urząd Ochrony Danych Osobowych (UODO) oraz Urząd Komunikacji Elektronicznej (UKE).

Przedmiotem rozważań zaproszonych ekspertów stały się zagadnienia związane z analityką danych. Prelegenci byli zgodni co do tego, że analityka danych już teraz dzięki zastosowaniu nowoczesnych systemów teleinformatycznych odgrywa ogromną rolę i to znaczenie będzie coraz większe dla różnego rodzaju procesów.

„Analityka jest czymś, co jest powszechnie dostępne i wykorzystywane. To jest powód do zadowolenia, ponieważ analiza danych jest dobra z punktu widzenia społecznego. Powinniśmy jednak pamiętać, że wielkie zasoby danych to również wielka odpowiedzialność" – zauważył dr Wojciech R. Wiewiórowski, wówczas zastępca Europejskiego Inspektora Ochrony Danych.

Prelegent zwrócił uwagę, że w ramach tej odpowiedzialności należy w sposób przemyślany konstruować systemy informacyjne, w tym bazujące na sztucznej inteligencji, które wykorzystują mechanizmy analizy danych. Sztuczna inteligencja bazuje bowiem na zasobach danych, których dostarcza jej człowiek.

„Jeżeli nie przekaże się systemowi wszystkich informacji, które są dostępne, będzie on działał w zawężonym zakresie z przekonaniem, że to, co otrzymał, stanowi reprezentację całego świata" – wskazał dr Wojciech R. Wiewiórowski.

Właśnie przykład tego, do czego prowadzi brak staranności, podał prof. Jerzy Cytowski z UKSW. Opowiadał o systemie teleinformatycznym COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), który był wykorzystywany przez wymiar sprawiedliwości w Stanach Zjednoczonych. Na bazie wielu różnych czynników przewidywał ryzyko popełnienia przez podsądnego kolejnego przestępstwa (recydywy). Miało to wspierać sędziów przy podejmowaniu decyzji co do wysokości wyroku lub zastosowania możliwości przedterminowego zwolnienia z zakładu karnego.

„Okazało się, że osoby ciemnoskóre miały mniejsze szanse na wcześniejsze warunkowe zwolnienie, ponieważ w ich przypadku system przeszacowywał ryzyko, omyłkowo typując ich jako osoby, które wrócą na przestępczą ścieżkę. System bazujący na analityce danych nauczył się podejmować decyzje z uwzględnieniem uprzedzeń rasowych. Tego ryzyka nie wyeliminowano na etapie projektowania" – zwrócił uwagę prof. Jerzy Cytowski.

Dlatego też, jego zdaniem, systemy informacyjne powinny być przed ich wdrożeniem do różnego rodzaju procesów decyzyjnych testowane analogicznie jak nowe leki.

Minimalizowanie ryzyka związanego z wykorzystaniem analityki danych wymaga, oprócz dobrych jakościowo systemów, również wyedukowanych kadr, co zauważył Włodzimierz Marciński, prezes Polskiego Towarzystwa Informatycznego.

„Moim zdaniem analityka danych nie występuje samodzielnie. Zawsze jest silnie skorelowana z jakąś kategorią, np. analityka danych ekonomicznych, analityka danych administracji czy też analityka medyczna. W tym kierunku idzie edukacja. Obecnie w Polsce nie ma bowiem odrębnego kierunku o nazwie analityka danych. Natomiast tę tematykę porusza się na 149 kierunkach na 57 uczelniach wyższych. Najczęściej analitykę umieszcza się na kierunku informatyka i ekonometria, dalej

jest matematyka, analityka medyczna, bibliotekoznawstwo czy też analityka gospodarcza" – wymienił prelegent.

Z tym poglądem zgodził się dr inż. Janusz Dygaszewicz, dyrektor Departamentu Systemów Teleinformatycznych, Geostatystyki i Spisów w Głównym Urzędzie Statystycznym (GUS), który wskazał, że budowanie odpowiednich kompetencji analityka danych odbywa się poprzez kształcenie, w czasie którego wypracowuje się u przyszłego specjalisty umiejętności analityczno-statystyczne, zdolność zdroworozsądkowego myślenia, ale również tzw. *hacking skills*, które mogą okazać się przydatne przy zabezpieczaniu analizowanych zasobów.

W czasie konferencji przedstawiciele różnych branż dzielili się doświadczeniami ze stosowania analityki danych oraz wskazywali problemy, które występują na tym tle. Część z nich przybiera charakter barier do efektywniejszego wykorzystania narzędzi analitycznych.

W administracji państwowej największe w kraju zasoby danych analizowane są z powodzeniem na potrzeby statystyki publicznej. W tym przypadku największy nacisk kładzie się na jakość opracowywanych danych, co okazuje się szczególnie istotne w epoce *big data*. Duże ilości danych nie zawsze bowiem niosą ze sobą wartość dodaną.

„Jesteśmy elementem systemu informacyjnego państwa. Informacje, które dostarczamy, mają zapewnić bezpieczeństwo informacyjne państwa, gospodarki i społeczeństwa" – zauważyła Anna Długosz, zastępczyni dyrektora Departamentu Systemów Teleinformatycznych, Geostatystyki i Spisów w GUS.

Weryfikacja jakości danych odbywa się niezależnie od źródła ich pozyskiwania, a więc również w przypadku informacji pochodzących ze zbiorów administracyjnych. Anna Długosz wskazała, że GUS opracował metodę eliminowania błędów z danych pochodzących z rejestrów państwowych i ewidencji samorządowych, uzupełniania braków oraz standaryzowania danych w celu uzyskania zbiorów kompletnych pod względem podmiotowym i przedmiotowym, dzięki czemu odpowiadają standardom klasyfikacyjnym. Dopiero takie zbiory są wprowadzane do systemów GUS.

Na dużych zasobach danych operuje również Zakład Ubezpieczeń Społecznych (ZUS), na co zwrócili uwagę dr hab. Katarzyna Roszewska

z UKSW i dr Robert Marczak z ZUS. Jednocześnie wskazali, że brakuje standardów i mechanizmów kontroli w zakresie tego, jak funkcjonują wykorzystywane przez urząd algorytmy oraz jak wpływają na prawa ubezpieczonych. Dochodzi więc do sytuacji, w której ustawodawca nie potrafi poradzić sobie z tym, że algorytmy, choć pomocne w codziennej pracy urzędu, zaczynają mieć wpływ na prawa socjalne klientów ZUS.

„Taki stan rzeczy negatywnie wpływa na rozwój analityki w sektorze ubezpieczeń społecznych. Można by ją bowiem wykorzystywać szerzej, wyjść poza analitykę skupiającą się na czynnościach *stricte* obliczeniowych i przetwarzać dalej idące treści. Algorytmy – jak w przypadku ubezpieczeń gospodarczych – mogłyby pomóc chociażby ustalać, czy niepoprawne działanie ubezpieczonego, np. nieuiszczenie składki, ma charakter zawiniony czy niezawiniony" – tłumaczyli prelegenci.

Przykładem wykorzystania analityki danych w administracji publicznej jest również System Informacyjny o Infrastrukturze Szerokopasmowej (SIIS) administrowany przez UKE. Służy on nie tylko zbieraniu informacji od operatorów telekomunikacyjnych, lecz także opracowywaniu tych danych. Tak powstałe analizy są potem wykorzystywane przy podejmowaniu decyzji regulacyjnych czy też dokonywaniu podziału środków finansowych z Unii Europejskiej w ramach programów rozwoju sieci telekomunikacyjnych. Marcin Błasiak, przedstawiciel UKE, przyznał, że urząd boryka się z tym samym problemem, co GUS, a więc niezadowalającą jakością danych źródłowych. Chodzi głównie o dane adresowe: ulica Marii Skłodowskiej-Curie była sprawozdawana przez operatorów na 151 różnych sposobów, zaś aleja Komisji Edukacji Narodowej na 164 sposoby. W rezultacie UKE, dokonując walidacji danych, musi dokonywać ich korekty i standaryzowania. Jakość danych przekłada się bowiem bezpośrednio na jakość opracowań będących wynikiem analizy tych danych.

Zaawansowana analityka danych pozostaje w służbie administracji, gdy ta walczy z nieopodatkowanymi źródłami dochodów. Okazuje się, na co zwróciła uwagę w swoim wystąpieniu dr Klara Dygaszewicz z UKSW, że często zastosowanie mechanizmów analitycznych pozwala walczyć z nielegalnymi zjawiskami w sferach, które przez długi czas pozostawały poza kontrolą. Przykładem jest obrót kryptowalutami. Aparat skarbowy

pozyskuje dane stron transakcji od operatorów giełd walut wirtualnych, a także zapis historii transakcji. Następnie informacje te są poddawane analizie przy wykorzystywaniu technik *big data*, jak algorytmy stworzone przez analityków zatrudnionych w Ministerstwie Finansów.

„Do analizy może być wykorzystana również technologia blockchain, na której opierają się kryptowaluty. W USA pojawiły się programy, które analizują informacje zapisane w łańcuchu bloków i działając w oparciu o te informacje, identyfikują poszczególne transakcje na rynku, przypisują je określonym użytkownikom, a następnie wskazują wysokość podatku. Słabością tego systemu jest to, że w świecie kryptowalut mamy pseudonimowość" – mówiła dr Klara Dygaszewicz.

Nie tylko jednak administracja wykorzystuje mechanizmy analityczne do prowadzenia kontroli wobec innych podmiotów. Zdaniem dr. Mariusza Maciejewskiego z Akademii Leona Koźmińskiego mogą być one stosowane również, gdy odwrócimy role, a więc w nadzorowaniu działalności samych podmiotów sektora publicznego. Analityka danych lokuje się wówczas na etapie monitorowania podmiotu kontrolowanego i wychwytywania zjawisk patologicznych za pośrednictwem porównywania stanu rzeczywistego z modelem nieprawidłowości. Na przykład, gdy Najwyższa Izba Kontroli bada przestrzeganie zakazu łączenia funkcji w radach nadzorczych kilku spółek państwowych bądź komunalnych.

„Wystarczyłoby zebrać numery PESEL członków rad nadzorczych, a następnie porównać je z numerami ujętymi w Krajowym Rejestrze Sądowym" – wskazał prelegent.

Innym obszarem kontrolowania działań administracji w opisywany sposób powinny być zamówienia publiczne.

„Cechy nieprawidłowości ujmowane wówczas w modelach to krótki czas na złożenie oferty, znacznie wyższa cena w umowie niż określona w założeniach do przetargu, czy też złożona jedna oferta. W modelach można również uwzględniać zachowania konsumpcyjne osób zajmujących się zamówieniami publicznymi. Brytyjczycy już to robią. Monitorują portale społecznościowe i jeśli urzędnik chwali się chociażby drogim autem, to trzeba sprawdzić, czy jego zakup ma pokrycie w przychodach" – zaproponował dr Mariusz Maciejewski.

Przywoływane przez ekspertów przykłady stanowią dowód na to, że analityka danych może służyć optymalizowaniu procesów e-usług w sektorze publicznym. Upowszechnienie tego mechanizmu zależy od stworzenia instytucji odpowiedzialnej za analitykę publiczną. O pracach nad powołaniem zintegrowanej platformy analitycznej mówił dr Maciej Kawecki, wówczas dyrektor Departamentu Zarządzania Danymi w Ministerstwie Cyfryzacji (MC).

„Instytucja zarządzająca platformą będzie w oparciu o rozwiązania prawne, infrastrukturalne i techniczne uprawniona do pozyskiwania poddanych pseudonimizacji danych niezbędnych do przeprowadzania analizy, a następnie opracowywania raportu prezentującego wyniki badań. Przykładem wykorzystania platformy może być sytuacja, gdy minister zdrowia będzie chciał wprowadzić refundację określonego leku, ale przed podjęciem decyzji chciałby wiedzieć, z jaką skalą problemu ma do czynienia, gdzie występują ogniska chorobowe. Dziś przeprowadzenie takiej analizy jest utrudnione bądź wręcz niemożliwe, gdyż brakuje instytucji niezwiązanej szeregiem różnych tajemnic" – wyjaśniał dr Maciej Kawecki.

W kontrze do wystąpienia przedstawiciela resortu cyfryzacji Piotr Drobek, wówczas dyrektor Zespołu Analiz i Strategii w UODO, wyraził wątpliwość, czy przy projekcie zintegrowanej platformy analitycznej pomyślano o przeprowadzeniu oceny skutków dla ochrony danych.

„Założenie, że w ramach platformy będą wykorzystywane dane po pseudonimizacji lub zanonimizowane, powinno być traktowane jako założenie wstępne. W efekcie należy zbadać ryzyka, które mogą się pojawić w związku z tym, że na potrzeby prowadzonych analiz będą pozyskiwane informacje z zasobów zawierających dane osobowe" – wskazał Piotr Drobek.

Operator danych ma zacząć działać pod koniec 2021 r. Na początku z platformy ze względów bezpieczeństwa ma korzystać jedynie administracja centralna: Kancelaria Prezesa Rady Ministrów, MC, Ministerstwo Zdrowia, Narodowy Fundusz Zdrowia, ZUS. Partnerami projektu są uczelnie.

Wśród tematów dotyczących sektora prywatnego dużo uwagi poświęcono podmiotom finansowym.

Paweł Sawicki, radca prawny z Polskiej Izby Ubezpieczeń, mówił o tym, że dużą część działań instytucji zrzeszającej ubezpieczycieli stanowi analityka danych.

„Artykuł 426 ust. 2 pkt 6 ustawy z 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej (tekst jedn. Dz. U. z 2019 r. poz. 381 ze zm.) stanowi, że do zadań izby należy pozyskiwanie, gromadzenie, przetwarzanie i przekazywanie informacji o funkcjonowaniu rynków ubezpieczeniowych w Polsce i poza jej granicami oraz opracowywanie na ich podstawie i udostępnianie analiz i prognoz. Realizacja tego obowiązku nie byłaby możliwa bez zaawansowanej analityki prowadzonej przy pomocy informatycznych baz danych" – wskazał prelegent.

Analityka jest rozbudowana również w bankowości, gdzie służy chociażby do podejmowania decyzji kredytowych. I choć jej stosowanie okazuje się niezwykle przydatne, to już wyjaśnianie mechanizmów działania modeli analitycznych wywołuje praktyczne problemy. Temu zagadnieniu swoje wystąpienie poświęcił dr Arwid Mednis z Uniwersytetu Warszawskiego. Opowiadał, w jaki sposób wdrażane są zapisy wprowadzonego w 2019 r. art. 70a ustawy z 29 sierpnia 1997 r. – Prawo bankowe (tekst jedn. Dz. U. z 2019 r. poz. 2357), zgodnie z którym banki i inne instytucje upoważnione do udzielania kredytów są zobowiązane na wniosek podmiotów ubiegających się o kredyt do wyjaśnienia w formie pisemnej tego, jakie czynniki miały wpływ na dokonaną ocenę zdolności kredytowej.

„Trudność ze stosowaniem tego przepisu bierze się stąd, że tabele scoringowe, które służą do oceny zdolności kredytowej w niektórych bankach, obejmują nawet ok. 140 czynników, z których każdemu przypisana jest odpowiednia skala punktowa. Jest bank, który analizuje nawet rodzaje zakupów dokonywanych kartą płatniczą. Jeżeli kupujemy podkładki filcowe pod meble, to znaczy, że dbamy o posiadane rzeczy i to poprawia naszą ocenę jako potencjalnych kredytobiorców. Ale już regularne nabywanie alkoholu w niedzielę rano działa na naszą niekorzyść. Mnogość czynników oraz występowanie takich niuansów oznacza, że banki nie są w stanie w wyczerpujący sposób wykazać najważniejszych danych, w oparciu o które podjęto decyzję. Dlatego też sektor pracuje nad

dobrymi praktykami dotyczącymi sposobu realizacji wprowadzonego ustawą obowiązku" – wskazał prelegent.

Monika Jóźwiak, radca prawny ze Związku Banków Polskich, opowiadała o problemach w efektywnej współpracy międzysektorowej przy wymianie informacji o fraudach zidentyfikowanych dzięki analizie przepływów środków pieniężnych w bankach, firmach pożyczkowych i innych instytucjach finansowych. Artykuł 106d prawa bankowego dopuszcza co prawda wymianę między instytucjami informacji objętych tajemnicą bankową w przypadku przestępstw dokonywanych na szkodę banków oraz innych instytucji finansowych, ale nie może być on stosowany w przypadku zidentyfikowania incydentów obejmujących kilka sektorów (oprócz finansowego), jak na przykład ubezpieczeniowy czy telekomunikacyjny. Brak podstaw do sprawnej wymiany danych utrudnia kompleksową analizę zdarzenia mogącego nosić znamiona czynu zabronionego.

„W takim przypadku o zwolnienie z tajemnicy bankowej czy telekomunikacyjnej trzeba występować do prokuratora. Tymczasem jej uzyskanie trwa średnio od dwóch tygodni do nawet dwóch miesięcy, co często nie tylko utrudnia, ale wręcz niweczy działania operacyjne dotyczące dynamicznych przepływów pieniężnych. Z perspektywy instytucji finansowych wymiana informacji przy pomocy policji i prokuratury okazuje się nieefektywna" – opisywała Monika Jóźwiak. Jej zdaniem potrzebne są zmiany legislacyjne umożliwiające bezpośrednią międzysektorową wymianę danych. Jednocześnie modyfikacje przepisów powinny być na tyle precyzyjne, żeby nie poszły za daleko.

Istnienie przywołanych problemów potwierdziła dr inż. Agnieszka Gryszczyńska z UKSW oraz prokuratorka w Prokuraturze Okręgowej w Warszawie.

„Tajemnice sektorowe mają chronić klientów, więc nie powinno być tak, że nie można udostępniać danych klienta, żeby go chronić w sytuacji, gdy to właśnie ujawnienie danych przyczyniłoby się do tej ochrony" – wskazała dr inż. Agnieszka Gryszyńska.

Fakt, że coraz więcej procesów w życiu człowieka zależy od wyników analizy danych, w tym dokonywanej przez systemy bazujące na sztucznej inteligencji, wymusza konieczność stosowania dotychczasowych

przepisów w nowym kontekście, gdyż brakuje specjalistycznych regulacji. Przykładem, na który uwagę zwróciła dr inż. Agnieszka Gryszyńska, jest ustawa z 6 czerwca 1997 r. – Kodeks karny (tekst jedn. Dz. U. z 2019 r. poz. 1950), gdzie nie ma przepisów, które odnosiłyby się wyłącznie do zbierania i zabezpieczania dowodów elektronicznych. „Skoro jednak można żądać wydania rzeczy, w tym nośnika danych cyfrowych, to organ procesowy może zażądać również wydania danych, a nie tylko sprzętu, na którym zostały utrwalone. Na potrzeby materiału dowodowego nie trzeba więc rekwirować całego nośnika, prokuratura czy policja mogą ograniczyć się do pobrania danych" – wyjaśniła prelegentka.

Innym przykładem jest odpowiedzialność za produkt niebezpieczny uregulowana w prawie cywilnym i kwestia tego, czy można ją stosować w odniesieniu do produktów informatycznych, w szczególności systemów samouczących się. Doktor Leszek Bosek z Uniwersytetu Warszawskiego na tak postawione pytanie udzielił odpowiedzi twierdzącej. Jego zdaniem producenci oprogramowania, które po jego oddaniu rozwija się w oparciu o mechanizmy samouczenia się, powinni zadbać o wbudowanie w rdzeń tych programów narzędzi zabezpieczających przed wykonywaniem przez nie szkodliwych czynności. Jeśli zaś obowiązek bezpieczeństwa nie będzie realizowany, wówczas należy uznać, że taki produkt już pierwotnie jest niebezpieczny.

Dwudniowa dyskusja poświęcona analityce danych była 11. konferencją naukową z cyklu Bezpieczeństwo w Internecie odbywającą się na UKSW. W poprzednich latach poruszano takie zagadnienia, jak między innymi ochrona danych osobowych, informacja przestrzenna czy też strategie cyberbezpieczeństwa.

Piotr Pieńkosz*

* Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie.