

INTERNET RZECZY I EKSPLOMACJA DANYCH, CZYLI JAK ODNALEŹĆ IGŁĘ W STOGU SIANA

Piotr Ablewski

Institut Fizyki, Wydział Fizyki, Astronomii i Informatyki Stosowanej,
Uniwersytet Mikołaja Kopernika w Toruniu
ul. Grudziądzka 5/7, 87 100 Toruń
piotr@ablewski.pl

Abstract. Data mining and Internet of Things are one of the hottest topics in modern informatics. Combination of them makes a great chance to introduce computational thinking and algorithmics to pupils and students. Simple set up with Raspberry Pi microcomputer connected to the Internet makes an opportunity to teach about data mining and IoT from the basics. In the digital environment strongly simulating young people, these skills need to be developed with the special emphasis on practical skills.

1. Wstęp

„Inteligentne urządzenia” zaczynają nas otaczać ze wszystkich możliwych stron. Są one zarówno gadżetami, które pozwalają upraszczać codzienne czynności, jak i narzędziami wspomagającymi ludzkie ułomności lub kontrolującymi terapię medyczną. Urządzenia wykorzystywane w gospodarstwie domowym (sprzęt AGD i RTV), kontrolery ogrzewania, klimatyzacji, systemy wbudowane (choćby te implementowane na potrzeby zwiększenia bezpieczeństwa pojazdów) są bardzo często połączone z siecią Internet. Mogą się one ze sobą komunikować, wymieniając dane i kooperując przy wykonaniu większych zadań. Urządzenia te mogą również przesyłać dane do miejsc, gdzie będą one gromadzone. Część danych poddawanych jest analizie w czasie rzeczywistym, co pozwala na sprawne działanie urządzeń wchodzących w skład Internetu Rzeczy (IoT), bądź Internetu Wszechrzeczy (IoE). Gromadzone dane stają się częścią baz danych, w tym również struktur dużych zbiorów danych (ang. big data). Analiza tych zbiorów nie jest zadaniem trywialnym. Spowodowane jest to złożonością i różnorodnością danych w nich zgromadzonych (spełniają one zazwyczaj warunek 3V, a więc volume, velocity and variety, czyli duża ilość, prędkość przetwarzania i różnorodność danych). Celem analizy jest wydobycie z danych konkretnej

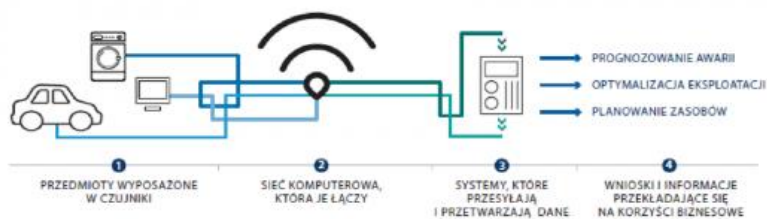
wiedzy. Proces ten nazywany jest eksploracją lub kopaniem danych (ang. data mining) [1,2,3,4].

Dane gromadzone przez IoT pochodzą ze środowiska, którego elementem jest człowiek. Są więc one źródłem informacji na temat ludzkiego życia. Oczywiście można to wykorzystać w szczytnym celu – do poprawy jakości lub ułatwienia życia, bardziej optymalnego zaplanowania codziennych zadań czy automatyzacji pewnych procesów. Część z danych może nieść ze sobą również wartość antropologiczną i stać się obiektem dalszych badań. Można wyobrazić sobie również, że dane posłużą do inwigilacji człowieka i próby niepostrzeżonego przejęcia kontroli nad jego tożsamością lub zachowaniem.

Pomimo zasygnalizowanych zagrożeń, pozytywne strony rozwoju IoT biorą górę, Internet Rzeczy rośnie w siłę i coraz więcej urządzeń codziennego użytku otrzymuje połączenie z siecią. Niniejsza publikacja ma na celu pokazanie, jak wykorzystując popularne oprogramowanie i proste, jednopłytkowe urządzenia, wprowadzać młodych ludzi do świata IoT oraz zachęcać ich do poszukiwania wiedzy w dużych zbiorach danych, zaczynając od prostych modeli opisujących otaczający nas świat.

2. Wstęp do IoT – Arduino oraz Raspberry Pi

Internet rzeczy zaczyna się od urządzenia, które podłączone jest do Internetu i potrafi wymieniać dane z innymi urządzeniami (patrz Rysunek 1). W przypadku gotowych rozwiązań są to zazwyczaj urządzenia wyposażone w jednopłytkowy mikrokomputer lub mikrokontroler, który potrafi odczytywać dane z sensorów i przekazywać je światu w zrozumiałej formie.



Rysunek 1 Schemat ideowy Internetu rzeczy (źródło: P. Choroś, Wykorzystanie analityki biznesowej w Internecie Rzeczy, SAS Institute, s. 5.)

W sieci znaleźć można coraz więcej gotowych zestawów uruchomieniowych, które najlepiej określić byłoby jako urządzenia z kategorii noszonych (ang. wearables). Znaleźć można również setki roboto pochodnych urządzeń, które przystosowywane są chociażby do dokładnego sprzątnięcia ludzkich domostw, bez konieczności angażowa-

nia w to człowieka. Jedyną akcją, którą wykonać musi człowiek to zaplanowanie ich pracy. Producenci sprzętu zazwyczaj udostępniają w tym celu dedykowaną aplikację mobilną, którą wykorzystać można zarówno do zaplanowania pracy urządzenia jak i do jej kontroli. Urządzenia te wyposażone są w szereg czujników, które pozwalają zarówno ocenić zakres prac, jakie powinny zostać wykonane, oszacować bądź przewidzieć problemy, które mogą pojawić się w trakcie pracy lub zorientować rzeczony urządzenie w przestrzeni.

Małe urządzenia przeznaczone do noszenia zazwyczaj również udostępniają pewne informacje pochodzące z czujników w nich zamontowanych. Potrafią one dokonywać wstępnej i szybkiej analizy tych danych, ostrzegając człowieka o możliwych zagrożeniach.

Tak więc urządzenie IoT można scharakteryzować jako urządzenie elektroniczne wyposażone w czujniki oraz pozwalające na komunikację z otoczeniem poprzez powszechnie dostępny interfejs.

Aby rozpocząć swoją przygodę z IoT, skorzystać można z gotowego, ogólnodostępnego rozwiązania przystosowanego do konkretnych działań, albo postarać się odtworzyć urządzenie z wykorzystaniem prostego zestawu uruchomieniowego. Wybierając narzędzie, które posłuży jako pomoc dydaktyczna przy IoT warto również zastanowić się, czy istnieje układ o szerszym zastosowaniu niż jedynie IoT i wspomóż pracę dydaktyczną związaną również z innymi tematami.

Przy poszukiwaniu układów opartych na mikrokontrolerach bezkonkurencyjnym wyborem wydaje się być platforma Arduino [5]. Arduino to cały ekosystem – zarówno hardware, jak i software. Sercem platformy jest 8 bitowy mikrokontroler Atmel AVR, którego porty wyprowadzone są na płytkę drukowaną w postaci bardzo czytelnie opisanych pinów. Dodatkowo mnogość rozwiązań dostarczanych przez twórców Arduino, pozwoli na wybór płytki optymalnej dla danego projektu – zarówno pod względem konkretnego mikrokontrolera, jak i pod względem dostępnych wyprowadzeń i rozmiarów całej konstrukcji. Arduino to również zestawy komponentów, czujników i gotowych rozwiązań w postaci tzw. shield'ów, czyli gotowych płytek drukowanych implementujących komponenty niezbędne do pełnienia przez układy odpowiednich funkcji (np. wspomniane wcześniej podłączenie do sieci Internet). Platforma programistyczna zintegrowana z projektem to Arduino IDE – przyjazne narzędzie bazujące na projekcie Processing, z którym zintegrowany jest język programowania bliźniaczo podobny do C. Dla bardzo początkujących użytkowników, bądź młodszych uczniów, którzy dopiero zaczynają swoją przygodę z tworzeniem kodu, przeznaczone są projekty bliźniaczo podobne do środowiska Scratch, gdzie gotowy program można złożyć z bloków. Bardziej zaawansowanym użytkownikom przypaść do gustu może kod pisany w języku C i kompilowany poprzez avr gcc. Wybór edytora kodu jest w tym przypadku dowolny.

Arduino nie oferuje jednak warstwy pośredniczącej w postaci systemu operacyjnego. Rozwiązanie te ma swoje pozytywne strony – jest szybkie, proste i mało awaryjne. Jednak nie daje one tak rozbudowanych możliwości dzielenia się danymi jak ma to miejsce w przypadku współpracy z systemem operacyjnym.

Najpopularniejszą platformą, która w prosty sposób integruje schemat wejścia i wyjścia znany z mikrokontrolerów z systemem operacyjnym jest Raspberry Pi [6]. Raspberry Pi to jednopłytkowy komputer oparty o architekturę ARM (czyli architekturę znaną ze smartfonów i tabletów, a coraz szerzej planowaną jako następcę architektur x86 i x86 64 dla przenośnych komputerów). Oczywiście istnieje wiele klonów niniejszego rozwiązania, jednak ich niedogodności i niedopracowanie wciąż zdają się górować nad zaletami. Raspberry Pi oferuje współpracę z systemami operacyjnymi opartymi na jądrze GNU/Linux, Microsoft Windows 10 IoT Core oraz RISC OS. Dostępne rozwiązania pozwalają na uniwersalne wykorzystanie tego narzędzia w procesie dydaktycznym, nie tylko skupiając się na rozwiązaniach związanych z IoT. Warto również mieć na uwadze świetną dokumentację techniczną, rozbudowane narzędzia służące do konfiguracji urządzenia i rzeszę ludzi wspierających projekt, którzy w ramach idei Wolnego Oprogramowania chętnie dzielą się z użytkownikami swoją wiedzą i doświadczeniem. Fundacja związana z projektem, bardzo mocno wspiera działania dydaktyczne prowadząc bloga [7], skupiając twórców i edukatorów na oficjalnym forum [8] i stronie wsparcia [9] projektu oraz wydaje dostępny nieodpłatnie w wersji elektronicznej magazyn The MagPi Magazine [10], w którym znaleźć można opisy ciekawych projektów, nadających się do realizacji w trakcie lekcji informatyki lub zajęć rozwijających zainteresowania uczniów. Nie łąda argumentem przemawiającym za tym rozwiązaniem jest niewątpliwie cena – wielostanowiskową pracownię opartą na Raspberry Pi można zbudować za cenę porównywalną z jednym lub dwoma stanowiskami komputerowymi opartymi na komputerach klasy PC.

3. IoT i Raspberry Pi

Na potrzeby praktycznej realizacji idei przedstawionej w niniejszej pracy, w charakterze urządzenia będącego jednym z elementów infrastruktury IoT, warto wykorzystać mikrokomputer Raspberry Pi. System operacyjny Raspbian – dystrybucja systemu GNU/Linux oparta na Debianie (aktualnie w wersji Stretch), która powstała specjalnie z myślą o mikrokomputerze RaspberryPi, posiada zestaw oprogramowania znacznie ułatwiającego pracę z czujnikami oraz płytkami testowymi. Przeglądając gotowe projekty wykorzystujące Raspberry Pi jako główne narzędzie, łatwo zauważyć mnogość gotowych rozwiązań oraz ułatwień dla twórców, chcących wykorzystywać tę platformę na szerszą skalę. Wielu producentów dostarcza gotowe zestawy akcesoriów przekształcające Raspberry Pi w gotowe narzędzie, które może zasilić szeregi urządzeń

IoT – począwszy od prostych systemów przydomowych stacji pogodowych, skończywszy na systemach zarządzających ogrzewaniem, czy implementujących rozwiązania znane z systemów alarmowych i przeciwwłamaniowych. Dodatkowo, dzięki interfejsowi GPIO (ang. General Purpose Input/Output), Raspberry Pi łatwo rozszerzyć o dodatkowe elementy będące realizacją własnych pomysłów.

Jeśli zależy nam na możliwie jak najprostszym rozwiązaniu, wykorzystać można czujniki oparte na szynie 1 wire [11], która umożliwia podłączenie wielu urządzeń do wspólnego portu komunikacyjnego, przez co stopień skomplikowania budowanego układu znacznie maleje. Układy obsługujące szynę 1 wire to między innymi detektory światła, czujniki temperatury, wilgotności – układy reagujące na zmiany prostych parametrów fizycznych. Takie czujniki można podłączyć pod Raspberry Pi bez konieczności lutowania układów, czy inwestowania w zestawy uruchomieniowe zawierające płytki drukowane, dopasowane do portów posiadanego przez nas urządzenia. W tym przypadku wystarczające jest wykorzystanie płytki prototypowej umożliwiającej montaż elementów elektronicznych „na wcisk”.

Rozwiązanie takie ma również aspekt dydaktyczny – może zainteresować uczniów i zachęcić ich do budowania swoich własnych układów, poznawania arkanów elektroniki i podstaw fizyki. Projektowanie układów traktować można jako mało inwazyjny sposób na zrozumienie podstaw działania obwodów elektronicznych, sposobów komunikacji czy działania systemu operacyjnego. Każde podłączane do Raspberry Pi urządzenie należy poprawnie skonfigurować, upewniając się czy porty działają w odpowiednim trybie oraz ustalając prędkości komunikacji po stronie urządzenia i komputera. Rozwiązanie takie niejako zmusza ucznia do poznania struktury wykorzystanego systemu operacyjnego i metod jego konfiguracji.

Raspberry Pi nie dostarcza gotowego narzędzia do komunikacji z podłączanymi urządzeniami. W przypadku wykorzystania systemu operacyjnego z rodziny GNU/Linux dane pozyskiwać można z odpowiednich plików stowarzyszonych z konkretnymi urządzeniami – jest to nie lada zaleta tego rozwiązania. Analiza tych danych to nic innego praca na strumieniach plikowych lub po prostu łańcuchach znaków. Wykorzystać można w tym celu dowolny język programowania – analiza odbywać może się poprzez skrypty powłoki jak i wyspecjalizowane narzędzia. Oczywiście producenci gotowych układów bardzo często udostępniają do nich biblioteki ułatwiające pracę i dostarczające prosty interfejs komunikacji z urządzeniem. Doświadczenie pokazuje, że większość z tych bibliotek dostępnych jest z poziomu języka C/C++ lub Python, które zdają się królować na platformie Raspberry Pi. Oczywiście w Internecie znaleźć można również otwarte implementacje bibliotek umożliwiających komunikację z zewnętrznymi urządzeniami (np. tymi, które wykorzystują wspomnianą wcześniej szynę 1 wire). Samodzielna implementacja może w tym przypadku sprawić niedoświadczonym programistom pewne problemy, gdyż wymaga wiedzy na temat funkcjo-

nowania wykorzystywanego protokołu komunikacyjnego i sprowadza się zazwyczaj do interpretacji ciągów bitowych na dane zrozumiałe przez pozostałe komponenty oprogramowania. Przykładowe dane tego typu znaleźć można w dokumentacji chociażby 1 wire [11].

Raspberry Pi posiada interfejs Ethernet, a w najnowszej swojej wersji również zintegrowany chip umożliwiający komunikację z wykorzystaniem technologii WiFi, tak więc idealnie spełnia założenia przyjęte wcześniej dla urządzenia IoT – potrafi komunikować się ze światem zewnętrznym. Wykorzystanie systemu operacyjnego daje możliwość szybkiej implementacji serwera WWW, czy to w postaci gotowego rozwiązania takiego jak Apache, nginx Node.js itp. czy w postaci własnej implementacji opartej na komunikacji poprzez protokół HTTP. Dzięki temu dane generowane przez zbudowane urządzenie mogą być pobierane przez inne układy lub kierowane do bazy danych, aby poddać je późniejszej obróbce.

Wykorzystanie Raspberry Pi w charakterze urządzenia IoT niesie jak widać wiele korzyści. Nie dość, że pozwala na implementację szerokiej gamy rozwiązań i zasymulowanie praktycznie dowolnego urządzenia IoT, to jeszcze dostarcza walorów dydaktycznych – zarówno tych związanych stricte z informatyką, jak również z elektroniką i fizyką.

4. Eksploracja danych

Termin eksploracja danych dotyczy poszukiwania informacji w dużych zbiorach danych, zazwyczaj w rozbudowanych bazach. Wywodzi się on z angielskiego określenia Data mining, które w bezpośrednim tłumaczeniu oznacza kopanie danych. W literaturze spotkać można również określenia takie jak pozyskiwanie, ekstrakcja czy zdobywanie danych. W gruncie rzeczy oznaczają one zazwyczaj tę samą czynność – poszukiwanie wiedzy w dużych zbiorach danych. Człowiekowi zależy, aby proces ten był jak najbardziej zautomatyzowany – po pierwsze ze względu na znacznie większą szybkość przetwarzania danych przez komputery, jak również ograniczoną ludzką uwagę, która jest w stanie przeoczyć pewne prawidłowości występujące w danych.

Techniki eksploracji danych oparte są na metodach statystycznych, uczeniu maszynowym, sieciach neuronowych, logice rozmytej, metodach ewolucyjnych i wizualizacji danych. Celem eksploracji jest natomiast znalezienie prawidłowości, klasyfikacja oraz jakościowa lub ilościowa analiza danych. Efektem działania algorytmów wykorzystywanych przy eksploracji danych jest zazwyczaj reguła logiczna, którą spełniają analizowane dane, drzewo decyzyjne lub sieć neuronowa potrafiąca podejmować decyzje na podstawie wiedzy zgromadzonej w przedstawionych jej danych.

Dane poddawane analizie są zazwyczaj zgromadzone w dużych bazach danych, zwanych strukturami big data. Charakteryzują się one dużą ilością danych, ich różno-

rodnością i koniecznością szybkiego przetwarzania (czyli wspomniana we wstępie zasada modelu 3V). Bazy te, ze względu na ogromną ilość danych, zwane są również hurtowniami danych.

Rynek dostarcza różnorodne oprogramowanie służące do eksploracji danych. Są to zazwyczaj narzędzia zintegrowane z najczęściej wykorzystywanymi systemami bazodanowymi, pochodzące od firm takich jak Microsoft, Oracle, IBM, HP, które pozwalają zamienić niezrozumiałe dane na formę akceptowalną przez człowieka. Kolejnym krokiem jest analiza danych, przy której wykorzystuje się pakiety do statystycznej obróbki danych, narzędzia pozwalające na klasyfikację danych oraz implementacje metod uczenia maszynowego i sieci neuronowe. Dla niewprawnego użytkownika wykorzystanie tych narzędzi ogranicza się do kliknięcia kilku opcji lub wywołania narzędzia z określonym zestawem parametrów wejściowych. W tym kroku można oczywiście pokusić się o jak najbardziej optymalne implementacje algorytmów wykorzystanych do analizy danych. Testując nowe algorytmy lub wybierając ten optymalny warto pokusić się sprawdzenie poprawności jego działania i porównać go z zaimplementowanymi już algorytmami.

Narzędzia do eksploracji danych potrafią dużo kosztować. Jednak w sieci znaleźć można wiele darmowych implementacji algorytmów do analizy danych. Jedną z darmowych propozycji jest program Weka 3 [12] – zestaw algorytmów uczenia maszynowego, nakierowanych na analizę danych. Zawiera on narzędzia do filtrowania, analizy i prezentacji danych, przez co jest kompletnym narzędziem przeznaczonym na potrzeby eksploracji danych. Internetowe środowisko twórców wspierających projekt jest bardzo liczne, przez co łatwo znaleźć materiały dydaktyczne prezentujące metody zaimplementowane w tym pakiecie oprogramowania. Łatwo znaleźć również gotowe scenariusze lekcji i kursy on line związane z analizą danych przy pomocy pakietu Weka.

Ważnym elementem eksploracji danych jest ich prezentacja, która potrafi znacznie ułatwić pracę. Czytelna prezentacja surowych danych ułatwia wybór odpowiedniej metody dalszej obróbki. Prezentacja danych poddanych analizie to ostatni etap prac związanych z eksploracją danych. Gotowe oprogramowanie pozwala na nie tylko na czytelne wyświetlanie danych i zaprezentowanie ich w graficznej formie przyjemnej dla ludzkiego oka, ale również na uchwycenie ich najważniejszych cech. Zazwyczaj wykresy opatrzone są również miarami statystycznymi, dzięki czemu łatwo zinterpretować otrzymane wyniki. Oprogramowanie takie jak Weka posiada wbudowany moduł do prezentacji danych. Oczywiście ta część analizy jest kluczowym krokiem całej procedury – nawet najlepsza analiza danych bez ich poprawnej prezentacji nie przynosi oczekiwanych rezultatów. W sieci znaleźć można wiele gotowych skryptów rozwiązujących problem czytelnej prezentacji danych. Wśród nich znajdują się zarówno

proste rozwiązania oparte na programie gnuplot, jak również te bardziej rozwinięte, bazujące głównie na bibliotekach Qt i Matplotlib.

Coraz popularniejsze stają się narzędzia webowe stworzone z wykorzystaniem języka HTML5 wspieranego skryptami JavaScript. Oferują one zazwyczaj wykresy oparte na grafice wektorowej w formacie SVG, który staje się standardem prezentacji danych.

5. Narzędzia oparte o chmurę komputerową

Narzędzia do analizy dużych struktur danych wymagają zazwyczaj dużych zasobów mocy obliczeniowej (jest to przecież jedno z założeń modelu 3V). Idealnym rozwiązaniem wydaje się więc zastosowanie w analizie narzędzi opartych na chmurze komputerowej.

Oprócz rozwiązań autorskich uruchamianych na skalowalnych w prosty sposób maszynach wirtualnych działających w środowisku chmury komputerowej, najwięksi gracze na rynku oferują wyspecjalizowane rozwiązania, zoptymalizowane na potrzeby metod wykorzystywanych przy eksploracji danych. Microsoft oferuje usługę Azure Machine Learning [13], która nie wymaga nawet specjalistycznej wiedzy. Udostępnione zostało środowisko webowe, które pozwala ułożyć program z gotowych bloków, pozwalając na tworzenie aplikacji metodą „przeciągnij i upuść”. Bardziej zaawansowani użytkownicy mogą tworzyć aplikacje z wykorzystaniem wysokopoziomowych języków programowania (np. C#, R czy Python).

Dzięki temu rozwiązaniu problem eksploracji danych nie jest problemem na poziomie implementacji i możliwe jest szybkie tworzenie gotowych rozwiązań. Komercyjne używanie platformy Microsoft Azure jest płatne, ale Microsoft udostępnia darmowe konta oferujące funkcjonalność wystarczającą do nauki dla uczniów i studentów.

Ciekawe rozwiązanie oferuje również firma Google. Ich rozwiązaniem to TensorFlow [14] – otwartoźródłowa biblioteka wykorzystywana w uczeniu maszynowym i sieciach neuronowych. Twórcy zapewnili kompatybilność z najbardziej popularnymi językami programowania poprzez udostępnione API. Biblioteka ta wykorzystuje zarówno CPU jak i GPU, może więc zostać zintegrowana z dowolnym rozwiązaniem chmurowym opartym na wirtualizacji maszyn.

6. Pierwsze kroki w eksploracji danych

Oczywiście korzystanie z gotowych rozwiązań ma sens, gdy głównym celem jest rozwiązanie postawionego problemu. Naukę metod eksploracji danych warto rozpocząć od znacznie prostszych przypadków. Skorzystać można z gotowych zbiorów danych, które poddawać można analizie. Warto przejrzeć tutaj dane oferowane jako standardowe zbiory testowe w programie Weka, w szczególności chyba najpopular-

niejsze w środowisku osób zajmujących się tematyką eksploracji danych, testowe zbiory Iris, dotyczące różnych gatunków irysów.

Na początek dobrym rozwiązaniem są również dane pozyskiwane z własnych elementów infrastruktury IoT, na przykład dane z czujników temperatury lub wilgotności podłączonych do Raspberry Pi. Można wtedy rozpocząć naukę prostych modeli regresji, klasyfikacji danych, czy filtrowania danych z tzw. outlier'ów, czyli danych nie pasujących do całego zbioru. Gotowe scenariusze takiego podejścia, wraz z opisem potrzebnego hardware'u i oprogramowaniem znaleźć można w [15].

Dalsze etapy nauki przenieść można do gotowych rozwiązań, np. programu Weka, w którym łatwo pokazać różnicę między różnymi podejściami do tego samego problemu i dodatkowo sprawdzić, czy zastosowane podejście jest odpowiednie. Kolejne kroki powinny być oparte o narzędzia związane z chmurą obliczeniową, gdyż jest to naturalne środowisko pracy specjalistów od eksploracji danych.

Rozpoczynając naukę od prostych modeli, które są dobrze znane z lekcji fizyki (np. zmiany temperatury w prostych układach fizycznych) nabrać można pewnej intuicji, która jest niezbędna dla osoby chętnej do zagłębienia się w świat wielkich danych. Znając model łatwo sprawdzić, czy zastosowane metody statystyczne lub algorytmy uczenia maszynowego przynoszą pożądany efekt i czy możliwe jest na podstawie zbudowanych automatycznie modeli odczytanie cennych informacji związanych z badanym zagadnieniem.

Kolejne kroki to praca z mniej deterministycznymi danymi, o strukturze, która nie do końca jest przewidywalna i oczywista. Dlatego właśnie połączenie nauki metod eksploracji z danymi pochodzącymi ze świata IoT jest tak ciekawym zagadnieniem. IoT jest istną kopalnią danych, a znalezienie w nich cennych informacji jest jak szukanie igły w stogu siana.

Nie tylko IoT generuje ciekawe z punktu widzenia eksploracji dane. Świetnymi źródłami są również zaawansowane eksperymenty fizyczne, trwające latami i generujące terabajty surowych danych, które należy poddać analizie, aby wydobyć z nich dane ciekawe z punktu widzenia nauki. Królują tutaj oczywiście największe projekty – Wielki Zderzacz Hadronów badający cząstki Modelu Standardowego i olbrzymie interferometry VIRGO i LIGO poszukujące śladów fal grawitacyjnych.

7. Podsumowanie

Eksploracja danych jest jednym z najgorętszych tematów powiązanych z informatyką w ostatnich latach. Coraz więcej ofert pracy związanych jest z tym zagadnieniem, a specjaliści od przeszukiwania danych – datolodzy (ang. data scientist) – są rozchwytywani na rynku pracy.

Naukę eksploracji danych warto połączyć z inną gałęzią informatyki, która również jest obecnie bardzo mocno rozwijana – Internetem rzeczy. Dane wygenerowane przez ten ekosystem są pełne niewyjaśnionych zależności i skrywają wiele ciekawych tajemnic.

Nieoceniona jest również wartość dydaktyczna takiej metody wprowadzania do świata eksploracji danych i IoT – uczniowie mają okazję poznać i zrozumieć podstawy nie tylko tych dwóch tematów, ale również fizyki, elektroniki i algorytmiki. Metoda ta rozwija myślenie komputacyjne i jest świetnym przykładem interdyscyplinarnego podejścia do tematu. Tematu trudnego i niezwykle skomplikowanego, który daje się przedstawić w formie prostej i przyjemnej w nauce, co powoduje, że zastosowane podejście nie powinno odstraszać młodych adeptów, a skłaniać ich do myślenia i poszukiwania niekonwencjonalnych rozwiązań.

Literatura

1. R. Agrawal, T. Imielinski, A. Swami, Mining Association Rules Between Sets of Items in Large Databases, Proc. ACM SIGMOD Conference, pp. 207 216, Washington DC, USA, May 1993
2. R. Agrawal, M. Mehta, J. Shafer, R. Srikant, A. Arning, T. Bollinger, The Quest Data Mining System, Proc. of the 2nd Int'l Conference on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, August 1996
3. R. Agrawal, R. Srikant, *Fast Algorithms for Mining Association Rules*, Proc. 20th Int'l Conf. Very Large Data Bases, pp. 478 499, Santiago, Chile, 1994
4. M. Zakrzewicz, Data Mining i odkrywanie wiedzy w bazach danych, Materiały konf. Polish Oracle Users Group PLOUG'97, str. 57 67, Zakopane, 1997
5. Witryna internetowa projektu Arduino, <https://www.arduino.cc>, ostatni dostęp 20.06.2019 roku.
6. Witryna internetowa projektu Raspberry Pi, <https://www.raspberrypi.org>, ostatni dostęp 30.05.2019 roku.
7. Blog twórców projektu Raspberry Pi, <https://www.raspberrypi.org/blog/>, ostatni dostęp 30.05.2019 roku.
8. Portal skupiający społeczność projektu Raspberry Pi, www.raspberrypi.org/community/, ostatni dostęp 30.05.2019 roku.
9. Forum społeczności projektu Raspberry Pi, <https://www.raspberrypi.org/forums/>, ostatni dostęp 30.05.2019 roku.

10. Witryna magazynu The MagPi Magazine, <https://www.raspberrypi.org/magpi/>, ostatni dostęp 30.05.2019 roku.
11. Dokumentacja szyny 1 wire na stronach firmy Maxim Intergrated, https://www.maximintegrated.com/en/app_notes/index.mvp/id/1796, ostatni dostęp 30.05.2019 roku.
12. Witryna internetowa projektu Weka 3, <https://www.cs.waikato.ac.nz/ml/weka/>, ostatni dostęp 30.05.2019 roku
13. Witryna internetowa projektu Azure Machine Learning, <https://azure.microsoft.com/pl pl/services/machine learning service/>, ostatni dostęp 30.05.2019 roku.
14. Witryna internetowa projektu TensorFlow, <https://www.tensorflow.org>, ostatni dostęp 30.05.2019 roku.
15. Witryna internetowa autora, <https://ablewski.pl/raspi weather>, ostatni dostęp 30.05.2019 roku.