



## **Izabella Krzemińska**

Automatic Data-Based Personality Assessment as a Method of Electronic Services Auto-Personalisation

Automatyczne wyznaczanie osobowości z danych jako metoda auto-personalizacji usług elektronicznych

### **Streszczenie dysertacji**

Promotor: Prof. dr hab. Witold Abramowicz

Promotor Pomocniczy: Dr Marcin Szmydt

## Motywacja

W kontekście wszechobecnej cyfrowej transformacji i symbiozy ludzi z urządzeniami, takimi jak smartfony, ślady cyfrowej aktywności użytkowników są coraz częściej wykorzystywane do profilowania i klasyfikowania (Kosinski et al., 2013). Jednak w dzisiejszym świecie stare metody segmentacji, czy też profilowania zawodzą. Dlatego, nowe determinanty zachowań i potrzeb użytkowników są stale poszukiwane w celu lepszego, aktywnego dostosowywania usług. W 1997 roku Philip Kotler zaproponował segmentację rynków konsumenckich w oparciu o zmienne geograficzne, demograficzne, psychograficzne i behawioralne. Polegała ona na podzieleniu klientów na grupy na podstawie stylu życia i osobowości (Kotler i Turner, 1997). Takie techniki segmentacyjne, oparte na doborze odpowiednich deskryptorów, zwiększały skuteczność działań marketingowych na rynku (Wedel i Kamakura, 2002). Zmienne wykorzystywane do segmentacji należy zawsze rozpatrywać pod kątem ich mierzalności, dostępności i zdolności do różnicowania klientów. Już Becker and Connor (1981) segmentowali klientów na podstawie osobowości, poszukując lepszego rozpoznania różnic indywidualnych, niż te wyznaczone przy wykorzystaniu metod demograficznych. Klienci z tej samej grupy pod względem demograficznym mogą być różnorodni psychograficznie (Kotler i Turner, 1997). Przed erą cyfrową, psychografia była wdrażana na podstawie kwestionariuszy papierowych, które nie są pozbawione ograniczeń takich jak czasochłonność, nakład pracy czy czynniki zniekształcające percepcję osoby badanej. Dodatkowo wykorzystanie papierowych kwestionariuszy osobowych na dużą skalę do celów biznesowych jest problematyczne zarówno dla użytkownika jak i dostawcy usługi.

We współczesnych koncepcjach Przemysłu 4.0 wymagany jest wysoki poziom indywidualizacji produktów. Firmy starają się poznać unikalne potrzeby jednostek (Trzecieliński, 2020). Możliwości oferowane przez świat cyfrowy w zakresie analizy różnic indywidualnych stale rosną. Jednak zbieranie danych z historii korzystania z usługi również angażuje dużo zasobów i wymaga czasu potrzebnego na obserwację. Pierwszą motywacją do napisania tej rozprawy jest właśnie poszukiwanie nowych klasyfikatorów zachowań i potrzeb klientów, które będą mogły być obliczone automatycznie, z dostępnych danych. Z biznesowego punktu widzenia zaletą takich automatycznych klasyfikatorów byłaby ich dostępność od pierwszego momentu korzystania z usługi.

Powszechna cyfryzacja umożliwiła dokładniejsze badania ludzkich zachowań dzięki zwiększającym się możliwościom technicznym. Co najistotniejsze, pozwala badać ślady zachowań naturalnych, a nie zachowania w warunkach laboratoryjnych czy deklaracje dotyczące zachowań (kwestionariusze zachowań). Dodatkowo, w przedsiębiorstwach typu zwinnego (ang. agile), postulowanych w koncepcjach Przemysłu 4.0, personalizacja ma kluczowe znaczenie w budowaniu przewagi konkurencyjnej na rynku. Klasyfikacje oparte na osobowości użytkownika mogą być użyte do personalizacji usługi czy produktu. Profil osobowości traktowany jako wyróżnik potrzeb użytkowników jest coraz częściej używany w mediach społecznościowych (SM) i systemach rekomendujących (RS). Powszechnie znane są badania Michała Kosińskiego dotyczące osobowości liczonej ze śladów w SM (Kosinski et al., 2013) oraz modułarny system o nazwie Watson autorstwa IBM liczący profil osobowości z tekstu (Watson, 2014). Nawet międzynarodowa afera związana z Cambridge Analytica (Hinds et al., 2020) nie była w stanie zatrzymać trendu wykorzystania danych cyfrowych do profilowania. Dostosowywane do zmieniającej się rzeczywistości prawo o ochronie prywatności skutkuje coraz większymi restrykcjami dotyczącymi użycia danych cyfrowych. Obserwujemy stopniową hermetyzację serwisów i monopolizację dostępu do danych o użytkownikach przez największych graczy na rynku.

O tym jak ważne jest dostosowywanie do użytkownika zaawansowanych technologicznie usług przekonała się firma Amazon, która sprzedała swoją pierwszą wirtualną asystentkę głosową o imieniu Alexa w milionach egzemplarzy. I chociaż odnotowano duży sukces rynkowy, to produkt nie był używany zgodnie z przeznaczeniem zaplanowanym przez jego twórców (celem było zwiększenie zakupów na platformie Amazon). Alexa posiada 90 tys. udogodnień i funkcjonalności, a 90% użytkowników Alexy użyło jej tylko raz do zrealizowania zakupów (Anand, 2009 (dostęp z dnia 30, sierpnia 2019)). Personalizacja usług zazwyczaj opiera się głównie na logach aktywności użytkowników w usłudze. W sytuacji, w której używanie usługi jest sporadyczne (jak w przypadku Alexy) nie wystąpi element uczenia się i dostosowywania do użytkownika. Kluczowe w przypadku takich usług, wydaje się znalezienie sposobu na przedstawienie użytkownikowi najodpowiedniejszych funkcjonalności, w celu zminimalizowania ryzyka niezadowolenia lub odrzucenia usługi. Wymaga to posiadania wyznaczników do klasyfikowania użytkowników względem potrzeb od początku korzystania z usługi.

Po aferze Cambridge Analytica, dodatkową zachętą do realizacji badań do celów dysertacji była ochrona prywatności danych w procesie personalizacji. Coraz częściej dyskusja na temat wykorzystania danych cyfrowych dotyczy problemów etycznych. W wielu krajach znajduje to odzwierciedlenie w wprowadzeniu odpowiednich norm prawnych. Opisane są już nowe mechanizmy ekonomiczne dotyczące działalności cyfrowych gigantów, których siła opiera się na wykorzystaniu posiadanych danych związanych z sferą prywatną użytkowników narzędzi informatycznych. W 2019 r. Soshana Zuboff opisała w swojej książce nowy rodzaj systemu gospodarczego (np. GoogleAdWords), skupionego wokół utowarowienia danych osobowych i nazwała go kapitalizmem nadzoru (Zuboff, 2019). Według autorki wcześniejszy kapitalizm przemysłowy wykorzystywał naturę i środowisko, a kapitalizm nadzoru stosuje mechanizmy totalitarne w stosunku do natury ludzkiej. Niewątpliwie, gromadzenie danych może przynieść korzyści jednostkom w społeczeństwie poprzez optymalizacje, na przykład oferowane w rozwiązaniach tworzonych dla Inteligentnych Miast. Jednak wykorzystywanie tych danych w kontekście zarobkowym np. reklama, modyfikowanie zachowań (rekomendacje, podtrzymywanie uwagi) może zagrażać ludzkiej wolności, autonomii i prawu do prywatności. Korzyści dla społeczeństwa (bezpieczeństwo, automatyzacja, ułatwianie życia) nie zawsze przewyższają koszty ponoszone przez osoby indywidualnie.

Część populacji nie jest świadoma tego, że płaci za korzystanie z usług danymi, dlatego oszacowanie postaw i zjawisk z tym związanych jest trudne. Ze wstępnych badań jakościowych przeprowadzonych na potrzeby dysertacji wynika, że niektórzy respondenci racjonalizowali swoje decyzje dotyczące bezpłatnego udostępniania swoich danych firmom Google czy Facebook i nie dostrzegają w tym zagrożenia. Według badania (Feliksiak, 2016), 46% respondentów akceptuje ten rodzaj kontroli a 30% jest przeciwny. Również raport Amnesty International z 2019 roku wspomina o platformach, takich jak Google i Facebook, które zbierają dane o użytkownikach i na podstawie algorytmów aktywnie wpływają na zachowania użytkowników (Isaak i Hanna, 2018). Regulacje prawne nie nadążają za pomysłowością firm w pozyskiwaniu, wykorzystywaniu i odsprzedaży danych o użytkownikach usług elektronicznych (Finck, 2021). Większość z tych operacji na danych jest niezrozumiała dla przeciętnego użytkownika usług cyfrowych. Ponadto, dobrowolna zgoda na utrwalanie danych behawioral-

nych jako niezbędny element działania usługi elektronicznej nie daje już podstawy do wnioskowania, że ludzie świadomie dokonują wyborów dotyczących udostępniania swoich danych. Dlatego kolejną motywacją do przeprowadzenia badań było sprawdzenie możliwości personalizacji usługi przy zachowaniu pełnej poufności danych o użytkowniku. Zarówno dane, jak i profil użytkownika mogłyby być chronione przed dostawcą usługi, albo współdzielone za zgodą użytkownika.

Personalizacja w takim prywatnym trybie przyniosłaby korzyści użytkownikowi i niekoniecznie musiałaby oznaczać utratę prywatności. Korzyści dla dostawcy usługi byłyby pośrednie poprzez wzrost zadowolenia i lojalności Klienta. Przy założeniu takiej pełnej ochrony prywatności użytkownika, zarówno liczenie profilu, jak i mechanizm dostosowywania usługi do profilu, musiałby odbywać się w sposób całkowicie zautomatyzowany na urządzeniu Klienta.

Podsumowując, uzasadnione jest przeprowadzenie badań, których celem byłoby poszukiwanie nowych klasyfikatorów pomocnych w różnicowaniu ludzi ze względu na ich potrzeby oraz zbadanie możliwości ich użycia do personalizowania usługi, z zagwarantowaniem ochrony prywatności użytkownika. Biorąc pod uwagę przedstawione powyżej cele, sformułowano w dysertacji następującą **tezę**:

*Automatyczna identyfikacja profilu osobowości użytkownika na podstawie danych dostępnych podczas instalacji usługi elektronicznej może służyć jako klasyfikator ułatwiający personalizację tej usługi.*

Teza jest weryfikowana na przykładzie kategorii usług elektronicznych jaką jest aplikacja na smartfona.

## **Metodyka badawcza**

W badaniach zaprojektowanych na potrzebę dysertacji, zastosowano metodykę Design Science, zaproponowaną przez Hevner and Chateerjee (2012). Głównym paradygmatem w podejściu Hevnera jest poszerzanie aktualnego stanu wiedzy poprzez budowanie i ewaluację artefaktów. Metodyka Design Science została wybrana ze względu na poznawczy i eksploracyjny charakter badań nad wyznaczeniem osobowości z danych. Drugim powodem był sam cel badań, czyli ocena jakości zbudowanych artefaktów, a nie określanie związków przyczynowo skutkowych. Ewaluacja artefaktów

poprzez eksperymenty w procesie badawczym ma na celu wykazanie ich nowatorskiego charakteru. W dysertacji zastosowano metodykę ewaluacji artefaktów zaproponowaną przez Venable et al. (2016).

Głównym celem badań było zaproponowanie nowego sposobu personalizowania serwisów elektronicznych, takich jak aplikacje mobilne, w oparciu o profil osobowości użytkownika aplikacji. Profil użytkownika byłby tworzony automatycznie na podstawie danych, znajdujących się w smartfonie podczas instalacji aplikacji. Istotnym założeniem dla projektu jest ograniczenie przetwarzania danych tylko do urządzenia końcowego (smartfon) i umożliwienie personalizacji pomimo braku ujawniania danych użytkownika dostawcy aplikacji. Decyzja projektowa o ograniczeniu przetwarzania danych miała na celu stworzenie warunków umożliwiających zachowanie poufności i ochrony zarówno danych, jak i wyników przetwarzania tych danych.

W ramach proponowanego przez Hevnera iteracyjnego Cyklu Relewanacji, dedykowanego definiowaniu problemu badawczego, przeprowadzono systematyczny przegląd istniejących rozwiązań w zakresie technik wyliczania profilu osobowości z pozostawianych przez użytkowników śladów cyfrowych, jak i technik personalizacji usług. Z kolei Cykl Rygoru, służący między innymi określeniu poziomu innowacyjności tworzonych artefaktów, obejmował analizę istniejącej bazy wiedzy w tym zakresie. Badanie obejmowało szeroki zakres tematyczny, ponieważ rozprawa dotyczy zastosowania wiedzy psychologicznej w sferze personalizacji usług, czyli w obszarze zarządzania przedsiębiorstwem.

## **Cele badawcze**

Głównym celem rozprawy doktorskiej było opracowanie nowatorskich, skutecznych i wystarczająco trafnych metod oceny potrzeb użytkownika określonych profilem osobowości, wykrywanego automatycznie na podstawie danych dostępnych w czasie instalacji usługi. W ramach zaproponowanego rozwiązania, opracowano model oparty o techniki uczenia maszynowego, służący do określania osobowości użytkownika na podstawie ograniczonej ilości danych pochodzących z jego telefonu komórkowego. Implementacja modelu bezpośrednio do aplikacji umożliwia użycie obliczonego profilu osobowości użytkownika do personalizacji usługi już od momentu instalacji aplikacji na smartfonie. Kluczowym założeniem dla tworzonej metody była realizacja zdefini-

ionowanego celu w warunkach ochrony prywatności danych użytkownika. Mając na uwadze przedstawione motywacje i zdefiniowane cele oraz stan wiedzy na temat personalizowania usług i adaptowania ich do potrzeb użytkownika sformułowano następujące pytania badawcze:

(RQ.1): Czy możliwe jest stworzenie automatycznej metody określania osobowości użytkownika na podstawie danych z telefonu komórkowego dostępnych w czasie instalacji aplikacji mobilnej?

(RQ.2.): Czy możliwe jest stworzenie automatycznej personalizacji z zachowaniem pełnej prywatności danych użytkownika, bez konieczności udostępniania danych dostawcy usługi?

(RQ.3.): Czy możliwe jest zastosowanie metody automatycznego wykrywania osobowości do personalizacji aplikacji mobilnej?

Podczas prac badawczych zaprojektowano innowacyjne podejście do budowania profilu użytkownika aplikacji. Profil taki jest użyteczny biznesowo i umożliwia klasyfikowanie i wyodrębnianie użytkowników o szczególnych potrzebach względem usług elektronicznych. Dostęp do takiej wiedzy o użytkowniku daje możliwość spersonalizowania jego doświadczeń od samego początku korzystania z usługi. Cały cykl badań oparto na danych pozyskanych od 3,5 tysiąca różnych osób. W zaprojektowanych badaniach do wyznaczania osobowości z danych, unikalnym wkładem do aktualnego stanu wiedzy jest użycie różnorodnych rodzajów danych zastanych w smartfonie w momencie instalacji aplikacji. Unikalne jest też użycie w badaniach znormalizowanych klas użytych do personalizacji usługi. Zgodnie z najlepszą wiedzą autora, jest to pierwsza próba zrealizowania zarówno liczenia profilu osobowości z danych zastanych w telefonie, jak i lokalnego wyliczania profilu osobowości na urządzeniu końcowym oraz wykorzystania tego profilu jako klasyfikatora dla personalizacji przy zachowaniu pełnej poufności danych użytkownika.

## **Struktura pracy**

Rozprawa składa się z 6 rozdziałów, których układ jest odzwierciedleniem przeprowadzonego procesu badawczego. Motywacja, pytania badawcze, cele badania, wybrana metodyka, teza pracy wraz z dwoma zdefiniowanymi artefaktami, zostały przedstaw-

ione w rozdziale pierwszym. W drugim rozdziale zostały opisane koncepcje teoretyczne powiązane z celami pracy wraz z wynikami przeglądu literatury jak również omówione zostały pokrewne badania, których celem było wyznaczanie osobowości z danych. Rozdział drugi kończy się omówieniem biznesowych potrzeb związanych z personalizacją, podsumowaniem zdefiniowanych braków w istniejących rozwiązaniach i kończy się prezentacją zaprojektowanego rozwiązania. Kolejny rozdział, trzeci, poświęcony jest omówieniu badań wstępnych oraz specjalnie stworzonym narzędziom, dzięki którym zebrano dane niezbędne do realizacji badania głównego. Przedstawiono w nim również założenia i szczegółowo omówiono dane, wraz z etapami ich przetwarzania. W kolejnym rozdziale, czwartym, zaprezentowano proces tworzenia modelu, opisano na jakiej podstawie podejmowano kolejne decyzje o wyborze zestawu zmiennych oraz użytych technik uczenia maszynowego. Celem tych eksperymentów i wielokrotnych weryfikacji było uzyskanie modelu wyznaczania klas profilu osobowości użytkownika o najlepszych parametrach jakościowych (trafność, precyzja, wskaźnik F1). Rozdział 4-ty zawiera też opis metody implementacji rozwiązania w usłudze. Rozdział 5-ty to opis ewaluacji stworzonych rozwiązań, łączący metody ewaluacji z eksperymentów laboratoryjnych na danych z testami na użytkownikach (mi. testy szybkości z jaką model wylicza profil na smartfonach). Ostatni, szósty rozdział dysertacji, poświęcony jest ocenie wkładu przeprowadzonych badań do istniejącej bazy wiedzy i omówieniu dalszych prac, które umożliwią rozwój rozwiązania lub poszerzą jego wykorzystanie w innych obszarach.

## **Analiza literatury**

Ze względu na tematykę badań przeprowadzono przegląd literatury w trzech obszarach: psychologicznych koncepcji osobowości, istniejących przykładów wyznaczania profilu osobowości użytkownika z danych oraz personalizacji usług.

Historia teorii osobowości jest starsza niż historia psychologii jako nauki i osobowość nie ma jednej definicji podzielanej przez wszystkich psychologów. Teorie osobowości są zorientowane funkcjonalnie, są związane z samym pomiarem osobowości i skupiają się na mechanizmach adaptacyjnych, determinantach zachowań oraz zaburzeniach i powiązanych z nimi terapii. W psychologii różnic indywidualnych, osobowość definiuje się jako zespół cech, które są zorganizowane, względnie stałe w czasie oraz wpływają



na poznanie i zachowanie ludzi (Strelau i Zawadzki, 2012). W latach 50-tych, XX wieku, poszukiwanie stałych i mierzalnych wymiarów osobowości oparto na analizie leksykalnej, według której wszystkie istotne aspekty osobowości zostały zakodowane w językach. Głównym celem tych badań (McCrae i Costa, 1987) było wyodrębnienie odwzorowań elementów językowych (głównie przymiotników) i stworzenie na ich podstawie struktury odzwierciedlającej ludzką osobowość. Zaczęto je interpretować jako tendencje lub dyspozycje do określonych reakcji, uczuć i zachowań. Tak powstała Teoria Wielkiej Piątki, która klasyfikuje cechy osobowości według pięciu wymiarów: *Ekstrawersja* (E), *Stabilność Emocjonalna* (SE), *Otwartość na Doświadczenie* (O), *Sumiennność* (S), *Ugodowość* (U). Model Wielkiej Piątki wydaje się idealny do klasyfikowania ludzi ze względu na ich potrzeby. Jest modelem potwierdzonym empirycznie w licznych badaniach (De Raad, 2012), udowodniono związki pomiędzy Wielką Piątką a zachowaniem i preferencjami ludzi (Judge i Ilies, 2002) oraz ma podstawy biologiczne (wykazano związek Wielkiej Piątki z temperamentem (Angleitner, 1991).

Badania nad wyznaczeniem osobowości użytkownika z jego śladów cyfrowych są realizowane od 10 lat. Osobowość jest wyznaczana najczęściej na podstawie danych z SM, głównie za pomocą dużych zbiorów danych tekstowych, dostępnych jako otwarte dane na Twitterze lub MyPersonality (Bin Tareaf et al., 2019; Khan et al., 2020). Istnieją też próby przewidywania osobowości na podstawie zdjęcia profilowego w SM np. (Liu et al., 2016). Pojedyncze badania eksplorują inne rodzaje danych, m.in. rejestry: połączeń telefonicznych (Montjoye et al., 2013), zdarzeń z aplikacji (Xu et al., 2016), ruchów gałek ocznych (Berkovsky et al., 2019), zdarzeń z urządzeń ubieralnych (ang. wearables) (Kalimeri et al., 2013). Niektóre z badań opartych o dane z SM są niemożliwe do powtórzenia ze względu na prawne ograniczenia w dostępie do danych osobowych, co uniemożliwia ich wykorzystanie poza samą usługą (Facebook, Twitter), ale same badania są cenne z perspektywy poznawczej. Cechą charakterystyczną większości badań jest to, że opierają się na dużej ilości danych z historii korzystania z usługi. Wydaje się również, że niektórzy badacze skupiają się bardziej na doskonaleniu samych metod uczenia maszynowego, niż na użyteczności tworzonych modeli w praktyce (Bin Tareaf et al., 2019; Khan et al., 2020). Pomimo starań, nie znaleziono przypadków opisujących wykorzystanie takiego modelu do celów biznesowych innych niż personalizacja w SM czy SR (Matz et al., 2017; Ning et al., 2019).

Biorąc pod uwagę techniki modelowania, uczenie maszynowe jest obecnie dominującą techniką służącą do przewidywania osobowości na podstawie danych (Gjurković et al., 2020; Li et al., 2021). Sama Wielka Piątka jest najczęściej używanym modelem do określania osobowości w kontekście danych cyfrowych (23 z 35 przeanalizowanych artykułów). O jej popularności może świadczyć istnienie Cybernetycznej Wielkiej Piątki opracowanej przez DeYoung (2015). To co jest istotne ze względu na zaproponowane badania, to fakt, że w analizowanych badaniach w literaturze, profil osobowości jest zwykle zbiorem dyskretnych zmiennych binarnych. Choć jest to wygodne w modelowaniu (znika problem niezbalansowania klas), to jednak nie ma uzasadnienia w przypadku personalizacji. W przypadku personalizacji, cechy mają służyć do identyfikacji tych użytkowników, którzy znacząco różnią się od typowego użytkownika pod względem potrzeb i motywacji do działania. A zastosowanie klas binarnych z mniejszym prawdopodobieństwem odzwierciedla zróżnicowanie zachowań ludzi (Stajner i Yenikent, 2021).

Jedną z podstawowych cech nowoczesnego przedsiębiorstwa typu zwinnego jest elastyczne dostosowywanie produktów finalnych do aktualnych potrzeb klientów. W otoczeniu o dużej zmienności opracowano strategie przeciwdziałania problemom wynikającym ze zbyt długiego czasu pomiędzy stworzeniem koncepcji produktu a dostarczeniem go Klientowi. Literatura przedmiotu opisuje różne strategie logistyczne przedsiębiorstw, a jedną z nich jest *odraczanie* (późna personalizacja lub opóźnione różnicowanie produktu, ang. postponement). Głównym powodem stosowania strategii opóźniania jest uzyskanie przewagi konkurencyjnej poprzez dostarczenie Klientowi najbardziej aktualnego spersonalizowanego produktu (Swaminathan i Lee, 2003). Wg Brdulak (2012), jedną ze strategii zarządzania w warunkach niepewności jest tworzenie elastycznych, samoregulujących się systemów opartych o sztuczną inteligencję. Technologie umożliwiają rozwój takiego elastycznego, samoregulującego się systemu wykorzystującego dostępne dane do tworzenia wiedzy o kliencie. Taki samoregulacyjny system obejmuje również możliwość odroczenia powstania ostatecznego kształtu produktu do czasu zainstalowania usługi u Klienta.

Fan and Poole (2006) przytoczyli aż 22 różne definicje personalizacji. Definicje są formułowane dla wielu różnych dyscyplin: marketing & e-commerce, kognitywistyka, nauki społeczne, informatyka i psychologia środowiskowa. Autorzy dokonali syntezy i

zapropowali ogólną definicję personalizacji (Blom, 2000): „*personalizacja to proces, który zmienia funkcjonalność, interfejs, dostęp do informacji i zawartość lub odrębność systemu, aby zwiększyć jego znaczenie dla osoby lub kategorii jednostek.*”.

Jednak z punktu widzenia celów rozprawy bardziej istotne od definicji personalizacji są ograniczenia istniejących metod. Najpopularniejsze modele automatycznej personalizacji są tworzone na podstawie zebranych danych z usługi (śledzenie, tagowanie, zbieranie zdarzeń użytkownika w SM, platformach telewizyjnych, aplikacjach sprzedażowych, bankowe itp.). Krytycznym problemem tych metod wydaje się czas potrzebny na poznanie preferencji użytkownika, który jest odwrotnie proporcjonalny do intensywności i częstotliwości użytkowania. Dlatego personalizacja może być wdrożona od pierwszego momentu użytkowania, tylko wtedy, gdy użytkownik poświęci czas na początkowe zdefiniowanie swoich preferencji (określa się to kustomizacją (Sundar i Marathe, 2010)). Jednak dla usług złożonych z setek funkcjonalności taki sposób dostosowywania usługi do użytkownika wydaje się trudny, a nawet niemożliwy. Użytkownikowi trudno jest określić preferencje w stosunku do nieznanych jeszcze technologii, z których wcześniej nigdy nie korzystał.

Obecnie większość zaawansowanych badań nad dynamicznym i adaptacyjnym personalizowaniem usług opartym na osobowości użytkownika odbywa się w kontekście robotyki i usług inteligentnych. Cenny wgląd w to, jaki wpływ ma Wielka Piątka na interakcje z usługami elektronicznymi pochodzi z badań interakcji człowiek — sztuczna inteligencja. Na przykład wg Matthews (2008) proces adaptacji usługi jest prostszy w przypadku wysokiej E, wysokiej S i wysokiej SE. Osoby neurotyczne (niska SE) nie są odporne na stres, czemu towarzyszy wyższy poziom lęku i mniejsza zdolność adaptacji do nieznanych warunków. Z kolei wysoka O, sprawia, że poznawanie i odkrywanie nowych rzeczy dostarcza użytkownikowi satysfakcji i użytkownik nie powinien być jej pozbawiany. Personalizacja doświadczeń użytkownika powinna uwzględnić różnice w preferencjach wynikające z osobowości i w ten sposób zapewnić lepsze ich dopasowanie do użytkownika.

Przegląd literatury dotyczący metod określania osobowości na podstawie danych oraz metod automatycznej personalizacji ujawnił luki w istniejących rozwiązaniach. Podsumowując:

- 1 Żadna z metod oparta o dane z używania usługi nie umożliwia policzenia profilu w momencie instalacji usługi. Użyteczność i dokładność estymacji profilu jest powiązana nie tylko z metodą wyliczania, ale też z dostępnością danych.
- 2 Metody wyznaczania osobowości z danych nie zapewniają pełnej ochrony prywatności użytkownika. Profile liczone są na danych niezanonimizowanych, przesyłanych, utrwalanych i przetwarzanych poza urządzeniem klienta. Przekazując dane, użytkownik traci prywatność i kontrolę nad tym, do czego te dane są wykorzystane.
- 3 Według wyników przeglądu literatury, nie ma jeszcze metody automatycznej personalizacji usługi w oparciu o profil osobowości wyznaczany z danych. Praktykowane jest wykorzystywanie danych z usługi w celu poprawy rekomendacji, uporządkowania treści lub dopasowania reklam. Nie ma jednak automatycznych metod personalizacji interfejsu i doświadczeń użytkownika od momentu instalacji usługi. Żadna ze znalezionych metoda nie łączy w sobie korzyści (dla użytkownika) z automatycznego profilowania i automatycznej personalizacji.
- 4 W analizowanej literaturze niewiele uwagi poświęcono omówieniu etyki i zasad wykorzystywania cyfrowych danych użytkowników usług. Prywatność to temat, który wymaga regulacji i badań również z prawnego punktu widzenia.
- 5 Zaobserwowano niewłaściwe i niespójne z teorią psychologiczną podejście do tworzenia klas opartych na kwestionariuszach osobowości. Większość badań opiera się na etykietach binarnych, co ułatwia budowanie modelu, ale w praktyce dyskwalifikuje użycie profilu w personalizacji usługi (przy klasach binarnych w obu grupach większość stanowią typowi użytkownicy, którzy nie różnią się istotnie pod względem potrzeb).
- 6 Proponowane dotychczas metody personalizacji nie spełniają celów wyznaczonych przez strategię odraczania, charakterystyczną dla nowoczesnych zwinnych biznesów. Klient otrzymuje usługę zaprojektowaną dla typowego użytkownika. Istniejące metody personalizacji dostosowują usługę do preferencji użytkownika już w trakcie korzystania z niej. Brakuje metod personalizacji od pierwszego kontaktu z serwisem.

- 7 Brakuje badań stanowiących adekwatny punkt odniesienia dla badań i metod proponowanych w rozprawie. Różnice to: inne dane, binarne klasy, inne metody zbierania danych lub inne modele osobowości.
- 8 Niewiele uwagi w badaniach poświęcono również personalizacji opartej na profilu osobowości użytkownika. Brakuje też przykładów wykorzystania personalizacji w działaniach przynoszących korzyści użytkownikowi, a nie tylko w dopasowaniu komunikacji marketingowej. Rozwój technologii umożliwi znacznie bardziej niezawodne i przyjazne ludziom rozwiązania, niż proponowane obecnie, co było jedną z motywacji dla zaprojektowanych badań.

## Zaproponowane Rozwiązanie

Na podstawie przeglądu literatury zidentyfikowano potrzebę opracowania metody umożliwiającej automatyczną personalizację doświadczeń użytkownika (UX) i interfejsu użytkownika (UI) w oparciu o osobowość. Osobowość użytkownika wg modelu Wielkiej Piątki wydaje się cenną informacją o jego potrzebach. Jednak klasyfikacja taka musi uwzględniać rzeczywiste różnice między jednostkami (wyróżniając niespecyficzne potrzeby). Wydaje się, że model Wielkiej Piątki odpowiada wymaganiom stawianym poprzez cele personalizacji i klasyfikowania klientów wg potrzeb. Dodatkowo wyznaczenie osobowości z danych cyfrowych, w większości istniejących badań, nie zapewnia pełnej prywatności użytkownika (tzn. profil obliczany jest z danych wrażliwych, które są przetwarzane i utrwalane poza urządzeniem końcowym). Jak już wspomniano, przechowywanie i bezpieczeństwo danych oraz wytworów z tych danych rzadko były omawiane w literaturze dotyczącej określania osobowości użytkownika na podstawie cyfrowych śladów. W związku z tym w badaniu zaproponowano eksperymenty, które mogłyby wykazać użyteczność metody wyznaczania osobowości z danych dostępnych na smartfonie (**UISPP - User's Initial Smartphone Personality Profile**). Zaprezentowano też metodę bezpiecznego użycia takiego profilu do automatycznej personalizacji interfejsu i doświadczeń użytkownika (**PAPA - Personality based Automatic Personalisation of App**).

Proponowana nowa metoda personalizacji PAPA, opiera się na danych statystycznych związanych z korzystaniem ze smartfona dostępnych w momencie instalacji usługi. Podczas instalacji usługi, na podstawie statystyk z danych wyliczany jest UISPP.

Dostępność profilu od momentu instalacji daje możliwość dostosowania serwisu do psychologicznych i osobistych potrzeb użytkownika. Dzięki temu użytkownicy mogą otrzymać tę usługę w różnej formie, wyglądzie, treści i konfiguracji w zależności od potrzeb przypisanych do profilu osobowości. Profil pozostaje częścią aplikacji i pozostaje prywatny, dostępny tylko dla użytkownika. Posiadanie tego wstępnego profilu nie wyklucza wykorzystania innych metod personalizacji, a nawet metoda zakłada weryfikację wstępnego profilu w trakcie używania usługi.

## **Weryfikacja zaproponowanego rozwiązania**

Zaprojektowane badania eksperymentalne, podporządkowane celom dysertacji, wymagały stworzenia dedykowanych narzędzi, takich jak: inwentarz psychometryczny z określeniem norm pomiarowych dla pięciu cech osobowości oraz aplikacja do przetwarzania i zbierania statystyk z telefonu komórkowego. W psychometrycznych badaniach wstępnych wzięło udział ponad 1300 osób (ochotników). Posłużyły one do skonstruowania wy standardyzowanego narzędzia do pomiaru osobowości.

Azucar et al. (2018) zdefiniowali trzy uniwersalne kryteria oceny badań nad budowaniem modeli osobowości na podstawie danych: dane muszą być powiązane z wynikami osobowości na poziomie indywidualnym, osobowość powinna być mierzona w oparciu o standaryzowane narzędzie do pomiaru osobowości (niekoniecznie diagnostyczne), a dane użytkowników muszą być zbierane w sposób zautomatyzowany.

UISPP został stworzony na anonimowych statystykach z telefonu zebranych w sposób zautomatyzowany od 2202 osób, w tym ok 25% stanowili ochotnicy, a ok 75% to osoby z płatnej rekrutacji z profesjonalnego ogólnopolskiego panelu badawczego. Badania wymagały samodzielnego zainstalowania aplikacji oraz udzielenia pięciu dodatkowych zgód na przetwarzanie danych z telefonu. Biorąc pod uwagę tak trudne kryteria rekrutacyjne, finansowanie badań wydaje się koniecznością i stanowi o ich unikalności na tle innych badań akademickich. UISPP to zestaw 5 modeli, po jednym dla każdego wymiaru Wielkiej Piątki (E, A, S, SE, O). Model miał na celu przewidzenie wyniku użytkownika zdefiniowanego poprzez 3 klasy: niską, średnią i wysoką. Klasa średnia odpowiada typowemu użytkownikowi zdefiniowanemu zgodnie z właściwościami rozkładu normalnego  $+ - 2$  SD od średniej. Wyniki były normalizowane do skali

stenowej i w tej skali klasa średnia odpowiadała przedziałowi od 4 do 7 stena. Klasa niskich wyników to steny 1-3, a wysokich to steny 8-10.

Podczas kreowania modelu użyto trzech referencyjnych modeli wyjściowych: losowego, personalizacyjnego i pochodzącego z podobnych badań (SoTA). Wyjściowy model losowy stworzono na podstawie predykcji losowej trzech klas dla każdej z cech, na podstawie rozkładu etykiet w posiadanym zbiorze danych,. Wyjściowy model biznesowy dotyczył oceny zysku z dopasowania usługi personalizowanej w porównaniu do usługi z brakiem personalizacji. Opierał się na biznesowym założeniu, że przy braku personalizacji tylko klasa modalna jest „poprawnie” zidentyfikowana, ponieważ tylko ona otrzymuje typowy produkt dostosowany do jej niespecyficznych potrzeb. Trzecim poziomem odniesienia były policzone średnie poziomy dokładności dla innych modeli liczących osobowość z danych. Zdecydowano się na te referencje pomimo braku porównywalności (różnice w metodach, inny materiał badawczy, klasy binarne). Stanowią one jednak istotną informację jaki poziom dokładności jest możliwy dla technik uczenia maszynowego w przypadku przewidywania zmiennych psychologicznych.

Z początkowych ponad 200 zarejestrowanych statystyk z telefonu, wygenerowano 477 zmiennych opisujących ludzkie zachowanie. Po etapie kreowania cech predykcyjnych do modelu i inżynierii tych cech, dokonano wyboru technik uczenia maszynowego, które tworzyły modele o najlepszych parametrach miar dokładności (ang. accuracy), precyzji (ang. precision) i *F1score*. Weryfikowano różne metody udoskonalania działania klasyfikatorów takie jak soft i hard voting, metody automatyczne (auto ML) i metody łączące różne klasyfikatory (stacking, ensemble). Celem było znalezienie techniki i hiperparametrów, które najlepiej pasują do użytych danych i najlepiej przewidują znormalizowane wyniki użytkowników w kwestionariuszu osobowości. Kolejny etap miał na celu zredukowanie modelu do wersji, która mogłaby zostać zaimplementowana na urządzeniu końcowym o ograniczonej mocy obliczeniowej. Zastosowano łączone metody selekcji cech, odrębnie dla każdej z 5 cech osobowości z wykorzystaniem metody *Mutual Information*, *Permutation Feature Importance* oraz uwzględniając cechy o wysokich współczynnikach korelacji z przewidywaną zmienną. Najlepsze parametry w eksperymentach uzyskały modele z użyciem technik Random Forrest (RF) i Extra Tree Classifier (ETC). W procesie implementacji klasyfika-

torów ML bezpośrednio do kodu android (ang. transpiling) liczba drzew musiała zostać ograniczona do 100 drzew na cechę.

W przeprowadzonych badaniach zostały spełnione wszystkie trzy kryteria postulowane przez Azucar et al. (2018). Przedstawiono procedurę psychometryczną oraz parametry psychometryczne dla wystandaryzowanego narzędzia do pomiaru osobowości, jak i określono normy interpretacyjne dla wyników. Dane do badań zostały zebrane w sposób zautomatyzowany za pomocą aplikacji Dr Charakter. Dzięki automatyzacji dane z ankiety osobowości dla każdej badanej osoby były łączone ze statystykami ze smartfonów i zapisane w postaci już zanonimizowanej.

W pracy przedstawiono i omówiono logiczną architekturę wykorzystania metody PAPA. W metodzie tej na podstawie statycznych danych z telefonu można przewidzieć profil osobowości użytkownika (UISPP). Następnie, UISPP można wykorzystać do zautomatyzowanego mechanizmu personalizacji aplikacji. Przedstawiona metoda PAPA umożliwia automatyczną personalizację, zapewniając lokalne przetwarzanie danych i poufność zarówno danych jak i profilu bez konieczności współdzielenia ich z dostawcą usługi. Co do zasady, dane służące do wyliczenia profilu, jak i sam profil, należy traktować jako dane wrażliwe i należy je chronić (Finck, 2021).

Jedną z motywacji niniejszej rozprawy było stworzenie metody personalizacji z pełną ochroną prywatności danych i profili użytkowników. Zastosowano etyczne korzystanie z danych wrażliwych i pełną ochronę prywatności badanych również w procesie badania i rozwoju rozwiązania. Do badania wykorzystano wyłącznie w pełni anonimowe dane i statystyki, przy pełnej przejrzystości ich gromadzenia. Wydaje się to spełniać warunki etycznego wykorzystywania danych osobowych i danych wrażliwych do celów naukowych.

## **Wyniki przeprowadzonych badań**

W badaniach potwierdzono, że zarówno UISPP jak i PAPA są przydatne z punktu widzenia personalizowania usług. Z proponowanym rozwiązaniem związane są niektóre krytyczne wymagania biznesowe, które bezpośrednio wiążą się ze zidentyfikowanymi w przeglądzie literatury lukami dotyczącymi istniejących rozwiązań. Te wymagania są w pewnym sensie także ograniczeniami proponowanego rozwiązania.



Podsumowując (odwołania w nawiasach dotyczą oznaczenia luk zidentyfikowanych po przeglądzie literatury):

- Profil osobowości jest wyliczony na urządzeniu końcowym użytkownika, aby możliwa była pełna ochrona danych, a co za tym idzie, dostępność profilu jest ograniczona do właściciela profilu. (Ad 1,2,4 )
- Profil jest obliczony na podstawie danych dostępnych w momencie instalacji usługi, aby automatyczne profilowanie usługi mogło odbywać się od samego początku. W przypadku aplikacji mobilnej można to zrobić zaraz po zainstalowaniu aplikacji na telefonie i uzyskaniu wymaganych prawem zgód użytkownika. (Ad 1,6,8)
- Nie ma potrzeby posiadania szczegółowego profilu osobowości do wstępnego profilowania. Oszacowanie na profilu na 3 poziomach (niski - średni - wysoki) wydaje się wystarczające do personalizacji. Personalizacja jest uruchamiana w przypadku zidentyfikowanych osób, które mają szczególne, nietypowe potrzeby i preferencje (poziom niski i wysoki cech). (Ad 3,4,5,8 )
- Ponieważ różne elementy usługi są dopasowywane do poziomu poszczególnych wymiarów osobowości niezależnie, nie ma konieczności liczenia całego profilu dla wszystkich usług. Modele są niezależne dla każdego wymiaru i mogą być również używane niezależnie, np. tylko do różnicowania realizacji usług dla Introwertyków i Ekstrawertyków. Takie podejście pozwala na eliminację zbędnych, nieistotnych informacji o użytkowniku oraz zwiększenie technicznych parametrów wydajnościowych. (Ad 3,4,8).

Proces badań ewaluacyjnych miał na celu walidację metody personalizacji PAPA zaproponowanej w rozprawie, wykorzystującej profil osobowości UISPP wyliczony na podstawie danych ze smartfona. Podczas tworzenia UISPP rozwiązanie było na bieżąco oceniane według zdefiniowanych kryteriów. Budowa ostatecznego modelu zakładała maksymalizację parametrów takich jak Dokładność, Precyzja czy F1 score (C.1 dla oceny UISPP). Miary jakościowe modelu odzwierciedlały stopień zgodności przewidywanego profilu z profilem uzyskanym w badaniu osobowości. Uzyskane parametry jakościowe modelu UISPP porównano z różnie zdefiniowanymi poziomami wyjściowymi (ang. baseline) (Tabela 1).

**Table 1. Wyniki testów dla zbudowanych modeli i porównanie finalnego modelu UISPP z wartościami wyjściowymi (baseline). P oznacza precyzję , A oznacza dokładność, F1 to F1 score.**

	Ekstrawersja			Sumiennosc			Ugodowosc			Stabilnosc			Otwartosc		
	P	A	F1	P	A	F1	P	A	F1	P	A	F1	P	A	F1
Baseline 1 losowy	0.56	0.34	0.39	0.60	0.35	0.41	0.54	0.34	0.38	0.57	0.35	0.40	0.63	0.35	0.41
Baseline 2 personalizacyjny	0.52	0.72	0.61	0.55	0.74	0.63	0.49	0.70	0.58	0.52	0.72	0.60	0.58	0.76	0.66
Baseline 3 SOTA	A: 0.61- 0.69														
S1. Holdout	0.66	0.7	0.68	0.55	0.74	0.63	0.62	0.69	0.60	0.52	0.72	0.60	0.67	0.76	0.68
S1. Holdout z walidacją krzyżową	0.65	0.73	0.65	0.52	0.72	0.60	0.68	0.71	0.60	0.52	0.72	0.60	0.59	0.77	0.67
S2. Soft Voting	0.65	0.74	0.65	0.55	0.74	0.63	0.58	0.70	0.59	0.52	0.72	0.60	0.59	0.76	0.66
S2. Hard Voting	0.65	0.74	0.65	0.55	0.73	0.62	0.67	0.71	0.60	0.52	0.72	0.6	0.59	0.76	0.66
S3. Stacking Classifier	0.70	0.74	0.64	0.55	0.74	0.63	0.67	0.71	0.60	0.52	0.72	0.60	0.59	0.76	0.66
S.4. Auto ML	0.72	0.77	0.69	0.68	0.75	0.65	0.49	0.69	0.57	0.67	0.72	0.62	0.59	0.74	0.65
UISPP (optymalna ilość drzew)	0.76	0.74	0.73	0.72	0.76	0.63	0.69	0.57	0.62	0.69	0.58	0.62	0.73	0.64	0.68
Final UISPP (ilość drzew <100)	0.76	0.74	0.73	0.71	0.62	0.61	0.70	0.61	0.64	0.71	0.65	0.63	0.74	0.65	0.69
UISPP vs baseline 1	0.20	0.40	0.34	0.11	0.27	0.2	0.16	0.27	0.26	0.14	0.3	0.23	0.11	0.3	0.28
UISPP vs baseline 2	0.24	0.02	0.12	0.16	-0.12	-0.02	0.21	-0.09	0.06	0.19	-0.07	0.03	0.16	-0.11	0.03
UISPP model vs baseline 3 dolny		0.13			0.01			0			0.04			0.04	
UISPP vs baseline 3 górny		0.05			-0.07			-0.08			-0.04			-0.04	
UISPP vs najlepszy model	0.04	-0.03	0.04	0.03	-0.13	-0.04	0.02	-0.1	0.04	0.04	-0.07	0.01	0.15	-0.12	0.02

Ostateczny UISPP uzyskał parametry przekraczające większość założonych wartości wyjściowych, w tym tych pochodzących z innych modeli wyznaczających osobowość z danych (wg. danych z publikacji). Zakresy dla 5 cech to: Dokładność: 0,70-0,76, Precyzja: 0,61-0,76 i F1: 0,69-0,71.

Kolejnym krytycznym elementem jest szybkość, z jaką obliczany jest profil, a więc minimalny czas po jakim jest dostępny do celów personalizacji. Aby móc wdrożyć personalizację na jak najwcześniejszym etapie usługi, czas potrzebny na jej obliczenie nie może być dłuższy niż kilka sekund (C.2 dla oceny UISPP). W oparciu o całkowitą próbkę w przypadku 95% użytkowników, model UISPP będzie dostępny do użycia w metodzie PAPA w ciągu 4,6 sekundy; dominanta to 1,4 sekundy. Ten wynik jest wystarczająco dobry, aby profil mógł być policzony podczas procedury instalacji aplikacji. Szybkość procesorów w smartfonach stale się poprawia, a patrząc na wyniki pomiaru telefony najnowszej generacji będą w stanie policzyć UISPP w czasie poniżej 1 sekundy.

Podczas realizacji testu technicznego, poproszono użytkowników o ocenę dopasowania profilu osobowości, policzonego na podstawie ich danych (C.3 dla oceny UISPP). 67% testerów oceniło dopasowanie profilu pozytywnie. Ocena PAPA została dokonana w oparciu o wykonalność metody UISPP (C.4), ponieważ działa ona automatycznie w środowisku rzeczywistym (aplikacja mobilna na Androida używana do testów użytkowników). Wyniki testów potwierdziły również zdolność metody do działania, pomimo wykorzystania wyłącznie lokalnego przetwarzania danych, które nie wymaga udostępniania danych i profili użytkowników usługodawcy (C.5). Ponadto, podobnie jak w przypadku UISPP, rozwiązanie to zostało ocenione przez uczestników testów (C.6). 81% testerów zadeklarowało, że zezwoliłoby na personalizowanie usługi na podstawie profilu osobowości policzonego z danych w telefonie.

Wyniki oceny UISPP i PAPA można podsumować w następujących punktach:

- W cyklu rozwoju i oceny wybrano model UISPP o najlepszych parametrach jakościowych. Celem było uzyskanie modelu nie gorszego od innych modeli predykcyjnych, ale spełniającego techniczne kryteria przetwarzania danych lokalnych (prywatności) i przetwarzania tylko danych dostępnych w telefonie w momencie instalacji usługi. Przeprowadzone eksperymenty wykazały, że pomimo zastosowanych celowych ograniczeń w konstrukcji modelu (dane dostępne na telefonie) oraz ograniczeń w kształcie samego modelu (liczba drzew), parame-

try jakości predykcji nie są gorsze od tych uzyskanych z innych danych (bez ograniczeń przyjętych dla UISPP). Najlepsze wyniki osiągnęły modele dla E, SE i U. Słabsze, choć akceptowalne wyniki, uzyskały modele S i O.

- Wykazano wykonalność implementacji UISPP w aplikacji na smartfony. Wyniki uzyskane dają podstawę do stwierdzenia, że niezależnie od parametrów technicznych smartfona możliwe jest policzenie UISPP w wystarczająco krótkim czasie podczas instalacji aplikacji. Profil osobowości użytkownika użyteczny dla personalizacji (model PAPA) można uzyskać bez udostępniania danych osobowych i poufnych. Profil może być również chroniony i dostępny tylko dla danej aplikacji. Co więcej, profil można obliczyć na podstawie stosunkowo niewielkiej ilości danych bez konieczności monitorowania zachowania w czasie.
- Zweryfikowano również opinie testerów odnośnie prezentowanego profilu osobowości oraz koncepcji wykorzystania automatycznie wyliczanego profilu do personalizacji aplikacji. Zdecydowana większość testerów zaakceptowała lub oceniła neutralnie zarówno UISPP, jak i koncepcję personalizacji aplikacji w oparciu o ten profil. Należy zauważyć, że taki pomiar jest sztuczny i nie odbywa się w sytuacji naturalnego kontaktu użytkownika z aplikacją, gdzie użytkownik nie zostanie poproszony o akceptację profilu, a raczej o zgodę na wykorzystanie danych wymaganą przepisami prawa.
- Podczas testów na użytkownikach potwierdzono, że model UISPP potrafi wykrywać osoby o specjalnych potrzebach, dla których personalizacja może mieć większe znaczenie niż dla około 70% typowych użytkowników. W pomiarze testowym zidentyfikowano 45 osób ze specjalnymi potrzebami spośród 170 uczestników testu.

Reasumując, w oparciu o weryfikację na przykładowej aplikacji zainstalowanej na smartfonach, przeprowadzona ocena potwierdza tezę niniejszej rozprawy, że automatyczna identyfikacja profilu osobowości użytkownika na podstawie danych dostępnych podczas instalacji usługi elektronicznej jest możliwa i użyteczna. Co więcej, identyfikacja ta może być również wykorzystana jako klasyfikator ułatwiający personalizację tej usługi. Przeprowadzone badania miały charakter interdyscyplinarny, wykorzystując metody psychometryczne oraz uczenia maszynowego w aplikacji mobilnej i przyczyniły

się do wzbogacenia bazy wiedzy w dziedzinie nauk społecznych w szczególności nauki o zarządzaniu.

## Referencje

- Anand, P. (2009 (dostęp z dnia 30, sierpnia 2019)). The Reality Behind Voice Shopping Hype.
- Angleitner, A. (1991). Personality psychology: Trends and developments. *European Journal of Personality*, 5, 185–197.
- Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis.
- Becker, B. W., & Connor, P. E. (1981). Personal values of the heavy user of mass media. *Journal of Advertising research*.
- Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting personality traits using eye-tracking data. *Conference on Human Factors in Computing Systems - Proceedings*.
- Bin Tareaf, R., Alhosseini, S. A., & Meinel, C. (2019). Facial based personality prediction models for estimating individuals private traits. *2019 IEEE, ISPA,BDCloud,SustainCom,SocialCom 2019*.
- Blom, J. (2000). *Personalization: A Taxonomy*. Association for Computing Machinery.
- Brdulak, H. (2012). *Logistyka przyszłości (red)*. Polskie Wydawnictwo Ekonomiczne, Warszawa.
- De Raad, B. (2012). Structural models of personality. *The Cambridge Handbook of Personality Psychology*.
- DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*.
- Fan, H., & Poole, M. (2006). What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. *Journal of Organizational Computing and Electronic Commerce - J ORGAN COMPUT ELECTRON COMME*, 16, 179–202.
- Feliksiak, M. (2016). Inwigilacja w internecie. *CBOS report*, 72.
- Finck, M. (2021). The Limits of the GDPR in the Personalisation Context. *Forthcoming in: U. Kohl, J. Eisler (eds), Data-Driven Personalisation in ...*

- Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Šnajder, J. (2020). PANDORA talks: Personality and demographics on reddit.
- Hevner, A., & Chateerjee, S. (2012). Integrated Series in Information Systems Volume 28.
- Hinds, J., Williams, E. J., & Joinson, A. N. (2020). “It wouldn’t happen to me”: Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, 143, 102498.
- Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8), 56–59.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*.
- Kalimeri, K., Lepri, B., & Pianesi, F. (2013). Going beyond traits: Multimodal classification of personality states in the wild. *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, 27–34.
- Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arif, A., & Khalid, H. A. (2020). Personality classification from online text using machine learning approach. *International Journal of Advanced Computer Science and Applications*, 11(3).
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.
- Kotler, P., & Turner, R. E. (1997). *Marketing management: Analysis, planning, implementation, and control* (Vol. 9). Prentice hall Upper Saddle River, NJ.
- Li, Y., Kazameini, A., Mehta, Y., & Cambria, E. (2021). Multitask Learning for Emotion and Personality Detection.
- Liu, L., Preoțiu-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. (2016). Analyzing personality through social media profile picture choice. *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 211–220.
- Matthews, G. (2008). Personality and information processing: A cognitive-adaptive theory. *The SAGE Handbook of Personality Theory and Assessment: Volume 1 - Personality Theories and Models*.

- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the USA*, 114(48), 12714–12719.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*.
- Montjoye, Y.-a. D., Quoidbach, J., & Robic, F. (2013). Predicting personality using novel phone-based metrics. *Social Computing, Behavioral-Cultural Modeling and Prediction Lecture Notes in Computer Science*, 48–55.
- Ning, H., Dhelim, S., & Aung, N. (2019). PersoNet: Friend Recommendation System Based on Big-Five Personality Traits and Hybrid Filtering. *IEEE Transactions on Computational Social Systems*, 6(3).
- Stajner, S., & Yenikent, S. (2021). Why Is MBTI Personality Detection from Texts a Difficult Task? *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 3580–3589.
- Strelau, J., & Zawadzki, B. (2012). Activity as a temperament trait. *Handbook of temperament*.
- Sundar, S. S., & Marathe, S. S. (2010). Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. *Human Communication Research*, 36(3), 298–322.
- Swaminathan, J. M., & Lee, H. L. (2003). Design for Postponement.
- Trzecieliński, S. (2020). *Przedsiębiorstwo zwinne - Agile Enterprise*.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems*, 25(1), 77–89.
- Watson, I. (2014). IBM Watson: How it works.
- Wedel, M., & Kamakura, W. A. (2002). Introduction to the special issue on market segmentation. *Intern. J. of Research in Marketing*, 19, 181–183.
- Xu, R., Frey, R. M., & Ilic, A. (2016). Individual Differences and Mobile Service Adoption: An Empirical Analysis. *Proceedings - 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications, BigDataService 2016*, 234–243.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.