

Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra aplikované ekonomie

**MANIFESTACE PSYCHOLOGICKÝCH
CHARAKTERISTIK UŽIVATELŮ FACEBOOKU
A JEJICH VYUŽITÍ**



Bakalářská diplomová práce

Autor: Mgr. Ondřej Krupčík

Vedoucí práce: Mgr. et Mgr. Jan Stoklasa, Ph.D.

Olomouc

2019

Poděkování

Děkuji panu Mgr. et Mgr. Janu Stoklasovi, Ph.D. za laskavé a vstřícné vedení této práce.

Prohlášení

Místopřísežně prohlašuji, že jsem bakalářskou diplomovou práci na téma: „Manifestace psychologických charakteristik uživatelů Facebooku a jejich využití“ vypracoval samostatně pod odborným dohledem vedoucího bakalářské diplomové práce a uvedl jsem všechny použité podklady a literaturu.

V Olomouci, 26. listopadu 2019

Podpis

„Můžete vytvářet zprávy, které některým lidem nedávají smysl, ale mají fakt velký význam pro ostatní lidi (Wylie, 2018)...“

Obsah

Úvod	5
1 Segmentace, targeting a osobnost člověka	7
1.1 Segmentace v marketingu	7
1.2 Targeting v marketingu.....	9
1.3 Chování spotřebitele a jeho osobnost	10
2 Facebook a jeho využití	13
2.1 Populace uživatelů Facebooku.....	13
2.2 Segmentace a targeting na Facebooku.....	16
2.3 Manifestace psychologických charakteristik	18
2.4 Využití psychologických charakteristik	22
3 On-line data a osobnost člověka.....	25
3.1 On-line data a „reálný svět“.....	25
3.2 On-line data a algoritmizovaná predikce osobnosti	26
3.3 Validita algoritmizované predikce	26
4 Zkoumaný problém, cíle a použité metody	29
4.1 Cíl práce.....	29
4.2 Data a použité metody	30
5 Výsledky.....	32
5.1 Deskripce zkoumaného souboru	32
5.2 Regresní analýz a predikce psychologických charakteristik.....	34
Diskuze.....	36
Závěr.....	39
Summary.....	40
Seznam literatury.....	41
Seznam zkratk	45

Úvod

Počet lidí, kteří se během prosince roku 2018 alespoň jednou připojili k sociální síti Facebook, dosáhl hodnoty 2,3 mld. V průběhu téhož měsíce byl průměrný počet denních uživatelů 1,5 mld. (Facebook, Inc., 2019). Společnost v daném roce dosáhla výnosů v hodnotě přibližně 55,8 mld. a zároveň hospodářského výsledku 22,1 mld. amerických dolarů. Jednalo se o zhruba 38% nárůst v porovnání s předchozím rokem. Podle serveru Investopedia (Sharma, 2019) mohlo být 89 % výnosů (na základě dat z roku 2017) tvořeno prodejem internetové reklamy (tedy asi 49,6 mld.). Pro srovnání můžeme uvést, že celkové reálné výdaje na reklamu v České republice za rok 2018 dosáhly podle Asociace komunikačních agentur v přepočtu na dolary hodnoty přibližně 4,9 mld. (Hospodářské noviny, 2019). Částka je tedy zhruba ekvivalentní jedné desetíně výnosů Facebooku z prodeje reklamy na internetu.

Můžeme říci, že přibližně jeden a půl miliardy lidí v současnosti sdílí vlastní emoce, myšlenky nebo přesvědčení pomocí jedné internetové služby. V podobě textů, obrázků, videí nebo různých druhů reakcí její provozovatel zaznamená pozoruhodné množství informací. Přestože oficiální údaje o současném stavu nejsou k dispozici, na základě sdělení z minulých let (Facebook, Inc., 2014) odhadujeme, že v současnosti může společnost shromažďovat data v řádech jednotek petabytů každý den. Pomocí matematických modelů je možné v těchto datech nacházet struktury popisující vlastnosti lidí, kteří tato data vytvářejí. Na základě explorační statistiky můžeme popisovat nové kategorie a segmenty, které mohou být využity například k optimalizaci uživatelské zkušenosti. Pomocí regresních modelů je například možné aktivitu uživatelů vztáhnout k existujícím psychologickým konstruktům jako je Big Five.

Facebookovou aplikaci MyPersonality, která byla v roce 2007 vytvořena na University of Cambridge a svým uživatelům dala možnost podstoupit několik desítek psychodiagnostických testů, využilo více než 6 mil lidí. Přibližně 40 % z nich umožnilo přístup k datům na jejich uživatelském

účtu (The Psychometrics Centre, University of Cambridge, 2013). V rámci jednoho výzkumného projektu tedy vznikl vzorek o rozsahu přibližně 2,5 mil osob. U každého z těchto lidí zároveň existují reprezentativní psychometrické záznamy.

Pro službu jako je Facebook, tedy takovou, která je přímo založená na aktivitě uživatelů, je rozhodující pozitivní uživatelská zkušenost. Jinými slovy – pro Facebook je výhodné generovat a také vybírat obsah, který je pro jeho uživatele atraktivní a který v nich vyvolá žádoucí odezvu. Cílem Facebooku nepochybně je pomocí nejrůznějších prostředků zlepšovat uživatelskou zkušenost a tím prodlužovat dobu, kterou uživatelé na síti stráví. To by zároveň mělo přispívat k maximalizaci zisku. Jedním z důležitých faktorů, které tuto zkušenost utvářejí, je vhodně zacílená reklama. Optimalizace v této oblasti lze považovat za dynamické a aktuální téma (Chen, Cheng & Chen, 2019), které může být předmětem výzkumu nebo marketingové praxe. Podle Facebooku lidé rádi vidí reklamy, které jsou pro ně relevantní. Doslova uvádí, že reklamy, které jsou relevantnější, stojí méně a dosahují lepších výsledků a lidé i firmy si v takovém případě odnáší lepší dojmy.

Tato práce je rozčleněna do pěti hlavních kapitol. V první z nich se věnujeme tématu segmentace a targetingu. V kontextu psychografie naznačujeme také problematiku osobnosti člověka a uvádíme základní termíny, ke kterým se vracíme v dalších částech práce. Ve druhé kapitole části se dostáváme k deskriptivnímu popisu populace Facebooku, zabýváme se také způsoby, kterými se manifestují psychologické charakteristiky uživatelů. Věnujeme se tradičnímu pojetí segmentace a targetingu v prostředí této sítě a naznačujeme možné využití psychografické segmentace. Třetí kapitola je věnována tématu on-lin dat, algoritmizované predikci osobnosti a její validitě. V posledních dvou kapitolách nejprve stanovíme cíle a následně zpracováváme volně dostupnou datovou sadu, která obsahuje základní informace o uživatelích a jejich chování v podobě lajků. Replikujeme regresní

modely, které slouží k odhadu hodnot jako věk, pohlaví, nebo rysy osobnosti podle Big Five. Pomocí křížové validace hodnotíme jejich přesnost.

1 Segmentace, targeting a osobnost člověka

Podle Kotlera a Armstronga (2017) se kupující na jakémkoli trhu liší ve svých potřebách, zdrojích, poloze, nákupních postojích a zvyklostech. Jak autoři uvádějí, prostřednictvím marketingové segmentace marketéři rozdělují velké, heterogenní trhy na malé segmenty, které mohou být dosaženy efektivněji a účinněji s produkty a službami, které odpovídají jejich unikátním potřebám. Kotler a Armstrong (2017) uvádějí, že na segmentaci navazuje targeting, což je proces výběru určitých segmentů.

1.1 Segmentace v marketingu

Dacko (2008) definuje marketingovou segmentaci jako proces rozdělování trhu do vymezených skupin zákazníků, kteří mají relativně odlišné chování, potřeby a jiné charakteristiky. Jinými slovy se tedy jedná o proces agregování zákazníků do skupin založených na společných charakteristikách a potřebách. Jak uvádějí Kotler a Armstrong (2017), existuje větší počet způsobů, pomocí kterých je možné trh segmentovat. Prodejci podle nich zkoušejí různé segmentační proměnné, ať už jednotlivě anebo v kombinaci. Nejčastěji jsou zákazníci podle autorů shlukováni na základě geografických, demografických, psychografických a behaviorálních proměnných, které shrnujeme v tabulce (1).

Tabulka 1. Schéma segmentace podle Kotlera a Armstronga (2017)

Segmentační proměnné	Příklady
Geografické	národy, regiony, státy, okresy, města, čtvrti, hustota obyvatelstva (městská, příměstská, venkovská), klima
Demografické	věk, fáze životního cyklu, pohlaví, příjem, povolání, vzdělání, náboženství, etnický původ, generace
Psychografické	životní styl, osobnost
Behaviorální	příležitosti, benefity, stav uživatele, míra využití, zákaznická věrnost

Podle Dacka (2008) označuje marketingový segment zákazníky, kteří budou na marketingový podnět reagovat podobným způsobem. Kotler a Armstrong (2017) uvádějí, že například zákazníci, kteří preferují velké, komfortní a luxusní automobily budou tvořit jiný segment, než zájemci o levnou a úspornou alternativu. Bylo by náročné vyrobit jeden vůz, který by byl atraktivní pro obě skupiny. Proto společnosti zaměřují své výrobky na jednotlivé segmenty.

Segmentace může být prospěšná do té míry, v jaké obchodníkovi poskytuje přehled a příležitost nabídnout jeden nebo více odlišných tržních segmentů nabídka na trhu (Dacko, 2008). Nabídky odpovídající určitým segmentům se mohou lišit v mnoha aspektech – ať už jde o různý typ produktu či služby nebo jiných prvků marketingového mixu jako je cena propagace nebo distribuce (Dacko, 2008).

Podle Kotlera a Armstronga (2017) má každý trh segmenty, ale ne všechny způsoby segmentace jsou podle něho vždy stejně přínosné. Například na některé generické léky mohou podle autorů stejným způsobem reagovat nízkopříjmové i vysokopříjmové skupiny zákazníků.

Psychografická segmentace

Jak uvádějí autoři Kotler a Armstrong (2017), psychografická segmentace umožňuje zpřesnění té demografické – lidé patřící do jedné demografické skupiny se mohou významně lišit například z pohledu osobnostních charakteristik nebo životního stylu. Jako příklad další psychografické proměnné autoři Belch a Belch (2018) uvádějí hodnoty. Souhrn uvádíme v tabulce (2).

Tabulka 2. Schéma psychografické segmentace (Belch & Belch, 2018)

Psychografické segmentační proměnné	Příklady
Osobnost	Společenský, nutkavý, introvertní, agresivní, ambiciózní
Hodnoty	Actualizer, fulfiller, achiever, experiencer, believer, striver, maker, struggler
Životní styl	Typ bílého plaňkového plotu atd.

1.2 Targeting v marketingu

Po tom, co marketér definuje marketingové segmenty, může se zaměřit na jeden nebo větší počet z nich (Kotler & Armstrong, 2017). Podle Dacka (2008) zahrnuje marketingový targeting proces definování specifických potřeb a profilů jednotlivých segmentů a výběr těch, na které může být zacílena adekvátní nabídka, strategie, nebo marketingový obsah. Jak dále uvádí autor, je možné zvolit různé typy strategií, které se liší v míře specifičnosti. Marketingová komunikace podle autora nemusí být cílená, ale také může být zacílena na konkrétní skupiny spotřebitelů s konkrétními zprávami, které jsou specificky určeny vzhledem k charakteristikám toho kterého segmentu (Dacko, 2008).

Podle Kotlera a Armstronga (2017) targeting zahrnuje zhodnocení atraktivity. Společnost by se měla zaměřit na segmenty, jejichž prostřednictvím může nejlépe dosahovat svých cílů a zároveň dokáže udržet tento stav v čase. Podnik s limitovanými zdroji by se tedy podle Kotlera a Armstronga (2017) měla zaměřit na určitou část trhu, jako příklad uvádí společnost Ferrari. Tento výrobce cílí na poměrně specifickou skupinu zákazníků. Například v USA prodá za jeden rok okolo 1500 vozů, což není mnoho, jejich ceny se ale pohybují ve stovkách tisíc dolarů.

Kritika segmentace a targetingu

Podle údajů z roku 2004 je segmentace trhu široce využívána v praxi. Společnost Marakon Associates and the Economist Intelligence Unit provedla studii na vzorku 200 vedoucích pracovníků a došla k závěru, že téměř 60 % z nich využilo segmentaci v předchozích dvou letech. Zároveň pouze 14 % mělo uvést, že segmentace přinesla reálné výsledky (Yankelovich & Meer, 2006).

Vedle termínů segmentace a targetingu se můžeme setkat s pojmy jako „mikromarketing“, „hypersegmentace“, „přímý marketing“, „individuální marketing“, „marketing jeden na jednoho“ nebo „tailoring“. Pouze v obecnější rovině shrnujeme, že tyto termíny vyjadřují snahu o individualizaci nebo

personalizaci marketingových záměrů. Jednotliví zákazníci v rámci těchto pojetí nejsou chápáni jako členové masových a homogenních segmentů, ale stávají se svébytnějšími individui.

Někteří autoři uvádějí, že marketing založený na segmentaci nemusí být účinnější než ten masový a navíc možná nepřispívá k budování silné značky (O'Donohue & Addison, 2001). Překážkou může být také to, že tradiční segmentace nedokáže identifikovat dostatečně specifické skupiny spotřebitelů (Bardakci & Whitelock, 2003). Významnou roli mají také faktory jako dynamika trhu, která způsobuje nestabilitou segmentů v čase nebo strukturální změny, které vedou ke změnám segmentů a migraci členů z jednoho segmentu do druhého (Blocker & Flint, 2007).

1.3 Chování spotřebitele a jeho osobnost

Autoři Jen-Hung Huang a Yi-Chun Yang (2010) popsali, že existuje vztah mezi jednotlivými dimenzemi pětifaktorového modelu (Big Five) a nákupním chováním na internetu. Demonstrují, že otevřenost, svědomitost, nebo extraverte korespondují s různými typy motivace při nakupování.

Existují studie, které poukazují na vztah rysů osobnosti a nákupního chování – respektive vztah těchto rysů a účinnosti určitého reklamního sdělení (Moon, 2002, Hirsh et al., 2012). Podle autorů (Matz et al., 2017) se často jedná o studie, založené na datech získaných v kontrolovaných podmínkách, což může představovat limitaci ve snaze popsat chování v reálné situaci. Další překážkou může být využití subjektivního hodnocení probandů, nikoli reálné nákupní chování.

Big Five model osobnosti

Pro potřeby této práce využíváme pojetí osobnosti, na kterém je založený známý pětifaktorový model – Big Five. Tento pojem můžeme chápat jako soubor teorií osobnosti a zároveň jako skupinu psychodiagnostických nástrojů, založených na faktorové analýze přirozeného jazyka. V podstatě se jedná o redukci lidské povahy do pěti snadno srozumitelných dimenzí, které byly souhrnně označeny jako Big Five.

Hřebíčková (2010), která je autorkou českých verzí NEO, základních pět faktorů pojmenovala slovy Otevřenost vůči zkušenosti, Svědomitost, Extraverze, Přívětivost a Neuroticismus. Autorka také upozorňuje, že v různých jazycích, včetně češtiny, má Big Five různá specifika a odchylky. V rámci této práce je nepovažujeme za významné, a tedy Big Five chápeme jako jeden koncept nezávislý na konkrétním jazyku.

Neuroticismus

Podle Piedmonta (2014) tento termín označuje škálu charakterizovanou afektivní vyrovnaností na jedné straně a emoční nestabilitou na straně druhé. Autor uvádí, že jedinci, kteří v neuroticismu skórují vysoko, jsou náchylní k prožívání psychického distresu, nerealistických představ, excesivnímu dožadování, naléhání a maladaptivním copingovým odpovědím. Přestože vysoké skóre v této doméně neindikuje přítomnost psychické poruchy, jedinci s vysokým skórem mají zvýšené riziko psychiatrického onemocnění (Piedmont, 2014). Podle Costy a McCrae (1992) neuroticismus tvoří šest samostatných faktorů. Hřebíčková (2010), tyto faktory pojmenovala jako úzkostnost, hněvivost-hostilitu, depresivnost, rozpačitost, impulzivnost a zranitelnost.

Extraverze

Podle Piedmonta (2014), který odkazuje na původní práce Costy a McCrae (1985), tato doména představuje kvalitu a intenzitu interpersonálních interakcí, potřebu stimulace a sklony k prožívání radosti. Autor dále uvádí, že odlišuje jedince, kteří jsou sociabilní, aktivní a orientovaní na lidi od těch, kteří jsou rezervovaní, rozvážní, samotářští a tiší. Podle Hřebíčkové (2010) je extraverze tvořena faktory pojmenovanými jako vřelost, družnost, asertivita, aktivnost, vyhledávání vzrušení a pozitivní emoce.

Otevřenost vůči zkušenosti

Tato doména je podle Piedmonta (2014) popisována jako proaktivní vyhledávání, tendence k ocenění vlastní zkušenosti, toleranci a potřebě prozkoumávat nové. Odlišuje zvědavé, originální, netradiční a kreativní jedince od těch, kteří jsou konvenční, neuměleční a neanalytičtí. Ze všech domén je právě tato nejvíce kontroverzní, je nejméně rozvinutá a prozkoumaná. V přirozeném jazyce má nejmenší počet deskriptorů. Podle Hřebíčkové (2010) je tvořena faktory nazvanými jako fantazie, estetické prožívání, prožívání, novátorské činnosti, ideje a hodnoty

Přívětivost

Popisuje to, do jaké míry si člověk užívá společnost jiných osob. Vedle toho hodnotí postoje jedince vůči jiným lidem – mohou být velmi lidské, soucitné, plné důvěry, dobromyslné na jedné straně a velmi antagonistické, cynické, manipulativní, pomstychtivé a nemilosrdné na straně druhé. Piedmont (2014) uvádí, že škála může být reprezentována machiavelismem na jedné straně a „matko-terezovostí“ na straně druhé. Podle Hřebíčkové (2010) ji utváří důvěra, upřímnost, altruismus, poddajnost, skromnost, jemnocit.

Svědomitost

Tato doména vyjadřuje míru organizovanosti jedince, vytrvalost a motivaci při chování zaměřeném na cíl. Odlišuje spolehlivé a puntičkářské jedince od těch, kteří jsou apatičtí, nedbalí a nepořádní. Také popisuje míru sebekontroly jedince a jeho schopnost odložit gratifikaci potřeb. Podle Hřebíčkové (2010) je tvořena faktory pojmenovanými jako způsobilost, pořádkumilovnost, zodpovědnost, cílevědomost, disciplinovanost, rozvážnost.

2 Facebook a jeho využití

V této kapitole se věnujeme základní deskripci populace uživatelů Facebooku, na kterou navazujeme tématem segmentace a targetingu na této síti. Věnujeme se také způsobům, kterými se manifestují psychologické charakteristiky uživatelů a naznačujeme způsob jejich využití.

2.1 Populace uživatelů Facebooku

Jak už jsme zmínili, počet lidí, kteří se během prosince roku 2018 alespoň jednou připojili k sociální síti Facebook, dosáhl hodnoty 2,3 mld. V průběhu téhož měsíce byl průměrný počet denních uživatelů 1,5 mld. (Facebook, Inc., 2019).

Na stránkách *Facebook Audience Insights* můžeme nalézt množství aktuálních informací o různých skupinách a segmentech uživatelů Facebooku. Jako jednu skupinu si můžeme prohlížet také všechny uživatele. Například zjistíme, že na dané síti jsou významně více zastoupeni muži než ženy (tabulka 3), nebo že nejvíce uživatelů, kteří mají uvedené vzdělání, uvedlo vysokou školu (tabulka 4).

Tabulka 3. Údaje o věku a pohlaví. Založeno na informacích, které sami uživatelé uvedli na svých profilech. Informace jsou k dispozici pouze u uživatelů starších 18 let.

Věková kategorie	Ženy (44%)	Muži (56%)
18-24	24 %	27 %
25-34	32 %	36 %
35-44	19 %	18 %
45-54	12 %	10 %
55-64	8 %	5 %
65+	5 %	4 %

Tabulka 4. Nejvyšší dosažené vzdělání na základě dat, která uživatelé uvedli.

Dosažené vzdělání	Podíl uživatelů
Střední škola	28 %
Vysoká škola	66 %
Navazující vysokoškolské studium	6 %

Můžeme si také všimnout, že z pohledu rodinného stavu je největší část uživatelů ženatých, nebo vdaných (tabulka 5). Mezi profesemi je nejčastěji zastoupené zaměstnání manažera (tabulka 6).

Tabulka 5. Data od uživatelů, kteří na Facebooku uvedli svůj rodinný stav.

Rodinný stav	Podíl uživatelů
Svobodný/á	37 %
Ve vztahu	16 %
Zasnoubený/á	4 %
Ženatý/vdaná	42 %

Tabulka 6. Odvětví, v nichž uživatelé pravděpodobně pracují dle údajů, které sami uvedli. Je možné uvést více možností.

Název zaměstnání	Podíl uživatelů
Management	27 %
Administrativní služby	22 %
Prodej	20 %
Výroba	14 %
Vzdělávání a knihovny	12 %
Umění, zábava, sport a média	12 %
Zdravotnické a lékařské služby	11 %
Komunitní a sociální služby	11 %
Přeprava a stěhování	9 %
Obchod a finance	9 %

Nejpočetněji je mezi uživateli Facebooku zastoupena angličtina, popřípadě španělština (tabulka 7). Z pohledu národnosti tvoří největší část Indové (tabulka 8). Značnou část uživatelů najdeme v asijských městech (tabulka 9).

Tabulka 7. Nejpočetněji zastoupené jazyky mezi uživateli

Jazyk	Podíl uživatelů
Angličtina (USA)	29 %
Španělština	15 %
Angličtina (UK)	12 %
Portugalština	7 %
Indonéština	6 %
Francouzština	5 %
Arabština	5 %
Vietnamština	3 %
Thajština	3 %
Turečtina	2 %

Můžeme si také všimnout, že většina uživatelů přistupuje na síť pomocí mobilního telefonu, nejčastěji s operačním systémem Android. Aktivitou, kterou uživatelé nejčastěji vykonávají je lajkování. Průměrný uživatel během 30 dní označí 13 příspěvků jako „to se mi líbí“.

Tabulka 8. Nejpočetněji zastoupené státy mezi uživateli.

Stát	Podíl uživatelů
Indie	13 %
Spojené státy americké	10 %
Brazílie	6 %
Indonésie	6 %
Mexiko	4 %
Filipíny	3 %
Vietnam	3 %
Thajsko	3 %
Spojené království	2 %
Turecko	2 %

Tabulka 9. Nejpočetněji zastoupená města uživatelů.

Město, stát	Podíl uživatelů
Dháka, Bangladéš	0,6 %
Dillí, Indie	0,5 %
Mumbaj, Indie	0,5 %
Karáčí, Pákistán	0,4 %
Láhaur, Pákistán	0,4 %
Bangkok, Thajsko	0,4 %
Hanoj, Vietnam	0,4 %
Ho Či Minovo Město, Vietnam	0,4 %
Gazipur, Bangladéš	0,3 %
São Paulo, Brazílie	0,3 %

Tabulka 10. Jak uživatelé v posledních 30 dnech přistupovali na Facebook. Zohledňuje všechna používaná zařízení. Založeno na aktivitě uživatelů.

Typ zařízení	Podíl
Počítač	2 %
Počítač a mobilní zařízení	20 %
Mobilní zařízení	78 %
Platforma	
Počítač	4 %
iPhone/iPod/iPad	14 %
Android	78 %

Tabulka 11. Frekvence aktivit za posledních 30 dní. Četnost udává průměrný počet akcí na jednoho uživatele. Založeno na aktivitě.

Aktivita	Četnost
Příspěvky označené to se mi líbí	13
Reklamy, na které klikli	12
Komentáře	5
Sdílené příspěvky	1
Využité promoakce	1

2.2 Segmentace a targeting na Facebooku

Na stránkách Facebook Ads Help Center (2019a) společnost doslova uvádí: „Lidé raději vidí reklamy, které jsou pro ně relevantní. Když se reklamy zobrazí relevantním okruhům uživatelů, odrazí se to v lepších obchodních výsledcích. Z tohoto důvodu před doručením reklamy někomu zvažujeme, do jaké míry je pro tohoto člověka daná reklama relevantní. Reklamy, které jsou relevantnější, stojí méně a dosahují lepších výsledků. Lidé i firmy si tak odnáší lepší dojmy.“ Z této citace vyplývá, že společnosti Facebook záleží na tom, aby uživatelé její sítě získávali pozitivní uživatelskou zkušenost.

Abychom poukázali na ekonomický rozměr problému, připomeneme, že podle serveru Investopedia (Sharma, 2019) mohlo být 89 % výnosů (na základě dat z roku 2017) tvořeno prodejem internetové reklamy (tedy asi 49,6 mld.). Zdá se tedy logické, že Facebook má zájem pomocí tvorby a výběru obsahu optimalizovat prožívání a akce uživatelů. Jak už jsme zmínili, téma optimalizace v těchto oblastech lze považovat za aktuální téma (Chen, Cheng & Chen, 2019), které může být předmětem výzkumu nebo marketingové praxe.

Faktory ovlivňující relevanci reklamy

Mezi faktory, které ovlivňují relevanci reklamy podle Facebooku patří pozitivní reakce – kliknutí, instalace aplikace, zhlédnutí videa – nebo reakce, které je negativní – když někdo u reklamy označí „tohle nechci vidět“ (Facebook Ads Help Center, 2019b).

Při zobrazování konkrétních reklamních sdělení uživatelům Facebook využívá různé metriky relevantnosti. Zhruba v polovině roku 2019 začal využívat trojdimenzionální model zahrnující takzvané „hodnocení kvality“, „hodnocení míry projeveného zájmu“ a „hodnocení konverzního poměru“. Hodnoty těchto metrik lze vyjádřit pomocí percentilů, které reklamu srovnávají s reklamami, které jsou jí podobné.

Podle Facebooku dochází k analýze historické úspěšnosti jednotlivých sdělení v aukci reklam za vybrané časové období, přičemž platí:

- Hodnocení kvality srovnává vnímanou kvalitu reklamy s reklamami soupeřícími o ten samý okruh uživatelů.
- Hodnocení míry projeveného zájmu srovnává míru projeveného zájmu o reklamu s reklamami soupeřícími o ten samý okruh uživatelů.
- Hodnocení konverzního poměru srovnává očekávaný konverzní poměr reklamy s reklamami, které mají stejný cíl optimalizace a soupeří o stejný okruh uživatelů.

Facebook také uvádí, že na základě diagnostiky relevantnosti reklamy mohou inzerenti lépe stanovit, jestli jimi provozované reklamy byly pro oslovený okruh uživatelů relevantní. Pokud reklamy nesplňují stanovené účely, mohou pomocí diagnostiky relevantnosti reklamy lépe pochopit, jestli by se výsledky nedaly vylepšit úpravou položek obsahu, prostředí po kliknutí nebo cílení na jiný okruh uživatelů (Facebook Ads Help Center, 2019a).

Segmentace a targeting

V současnosti může cílení reklamních sdělení na sociální síti Facebook probíhat několika způsoby. Jedním z nich je v podstatě tradiční přístup provedený pomocí výběru takzvaného okruhu uživatelů, tedy například pomocí určení lokality, věku, pohlaví nebo jazyka (obrázek 1). Tyto proměnné samozřejmě můžeme kombinovat a tím v podstatě vytvářet a vybírat nové segmenty.

Jak je patrné z obrázku (1) pomocí položky „*detailed targeting*“ je možné zpřesnit okruh uživatelům, kterým se má určité sdělení zobrazit. Umožňuje zahrnout, nebo naopak vyloučit uživatele na základě demografických údajů, zájmů nebo chování. Tyto skutečnosti mohou být založeny na stránkách, s nimiž uživatelé provedli interakci, nebo na aktivitách, jimž se na Facebooku věnují – právě zde se otevírá prostor pro psychologický targeting.

Obrázek 1. Výběr tzv. okruhu uživatelů v prostředí Facebook Ads Manager

Audience
Define who you want to see your ads. [Learn more](#)

Create new Use a saved audience ▼

Custom Audiences ⓘ

[Exclude](#) | [Create new](#) ▼

Locations ⓘ **Location:**
• Czech Republic

Age ⓘ 18 - 65+

Gender ⓘ All genders

Detailed targeting ⓘ **INCLUDE** people who match at least ONE of the following ⓘ

| [Suggestions](#) | [Browse](#)

- ▶ **Demographics** ⓘ
- ▶ **Interests** ⓘ
- ▶ **Behaviours** ⓘ

2.3 Manifestace psychologických charakteristik

Existuje řada behaviorálních fenoménů pozorovatelných na sociální síti Facebook, které mají vztah k tradičním psychologickým charakteristikám. Studie, které je možné dohledat, nejčastěji rozlišují tři kategorie základní druhy aktivity, jedná se o lajky, sdílený text a fotografie. Na základě těchto vstupů řada autorů snaží pomocí metod strojového učení predikovat nejrůznější charakteristiky, například inteligenci, věk, pohlaví, sexuální orientaci, etnicitu, politické preference, rysy osobnosti a další.

Abychom získali systematičtější přehled o možných způsobech manifestace psychologických charakteristik uživatelů Facebooku, vytvořili jsme systematičtější přehled na toto téma. Zaměřili jsme se pouze na studie využívající pětifaktorový model Big Five. Získali jsme celkem deset primárních zdrojů, které se zabývají vztahem chování kvantifikovaného

pomocí uvedených prediktorů (lajky, text, fotky) a rysy osobnosti podle Big Five. V tabulce (12) uvádíme jejich přehled.

Pro účely vyhledávání bylo využito rozhraní EBSCOhost, které umožnilo přístup například k databázím PsycINFO, PsycARTICLES, Academic Search Complete, Business Source Complete nebo MEDLINE. Dále byly využity databáze Scopus, Web of Science a služba Google Scholar.

Tabulka 12. Přehled studií na téma manifestace psychologických charakteristik uživatelů Facebooku

Autor, autoři	Rok vydání	Prediktor	n
Youyou et al.	2015	lajky	86 220
Kosinski et al.	2013	lajky	58 000
Tareaf, Berger, Hennig & Meinel	2018	lajky	738 000
Segalin et al.	2017	fotografie	11 736
Celli, Bruni & Lepri	2014	fotografie	112
Park et al.	2015	text	66 732
Kulkarni et al.	2018	text	152 845
Kern et al.	2014	text	69 792
Schwarz et al.	2013	text	75 000
Tadesse, Lin, Xu & Yang	2018	text	250

Lajky

Konstrukce lajků je uživateli Facebooku využívána pro vyjádření pozitivní asociace k on-line i off-line objektům jako jsou například produkty, aktivity, sporty, hudebníci, knihy, restaurace nebo webové stránky. Podle autorů Youyou a kolektiv (2015) lajky představují nejzákladnější druh digitální stopy, která může mít jasné implikace a využití. Kvalitativně je lze připodobnit například k vyhledávacím dotazům nebo historii webového prohlížení (Kosinski et al., 2013).

Identifikovali jsme celkem tři studie, které se tímto tématem zabývají. Již zmínění autoři Youyou a kolektiv (2015) využívají vzorek 86 220 uživatelů, kterým předložili až stopoložkový dotazník IPIP. Ukazují, že například shoda měřená pomocí Pearsonova korelačního koeficientu, mezi výsledkem self-posouzení a posouzením na základě lajků nabývá hodnot $r = 0,56$ pro otevřenost, $r = 0,44$ pro svědomitost, $r = 0,47$ pro extraverci, $r = 0,41$ pro přívětivost a $r = 0,41$.

Autoři Tareaf, Berger, Hennig a Meinel (2018) provedli studii na 738 000 dobrovolnících. Jejich kategorizační model byl například schopný v 81 % případů rozlišit mezi emocionálně stabilním a emocionálně nestabilním (neurotickým) jedincem.

Kosinski a kolektiv (2013) na vzorku 58 000 participantů ukazují, že pomocí lajků může být predikována celá řada citlivých údajů. Při predikci rysu otevřenosti jejich model dosáhl $r = 0,43$. U svědomitosti $r = 0,29$, extraverze $r = 0,40$, přívětivosti $r = 0,30$ a neuroticismu $r = 0,30$. V jedné z navazujících studií (Matz et al., 2017) autoři na příkladu extraverze uvádějí lajky s největší prediktivní hodnotou. Extraverzi podle jejich zjištění dokáží nejlépe predikovat označení, jako To se mi líbí u „Parties“, „Making People Laugh“, „Meeting New People“, „Dancing“, „Shwayze“, „Entourage“, „Rebelution“ a „Slightly Stoopid“. U introverze se jednalo o „Sanctuary“, „Stargate SG1“, „Dark Angel“, „Stargate“, „Serenity“, „Stargate Atlantis“, „Battlestar Galactica“ a „Computers“.

Fotografie

Autoři Segalin a kolektiv (2017) vycházejí z předpokladu, že dojem, který se lidé snaží vytvářet na sociálních sítích, koresponduje s jejich profilovou fotografií a také s osobností. Například ukazují, že extravertní a přívětiví lidé v profilových fotografiích spíše používají teplé barvy. V jejich fotografiích se také objevuje více obličejů, což zrcadlí inklinaci k sociálnímu životu. Lidé s neurotickými rysy vedle toho tendují k fotografiím pořízeným v interiéru. Autoři sestavili automatický klasifikační model a ukazují, že jeho predikce jsou při posuzování extraverze a neuroticismu spolehlivější než pokud ji provedou lidští hodnotitelé.

Autoři vytvořili Celli, Bruni a Lepri (2014) využívají metodu *Bag-of-Visual-Word* a na základě klastrové analýzy docházejí mimo jiné k závěrům, že extraverti se na profilových fotografiích více usmívají a více se fotí ve společnosti jiných lidí. Introverti se fotí sami a využívají menší detail. Svědomití lidé spíše udržují přímý pohled.

Text

Identifikovali jsme celkem pět studií, které se zabývají vztahem jazyka (buď v podobě soukromých zpráv, nebo veřejných příspěvků) a rysů osobnosti v kontextu sociální sítě Facebook. Lze říci, že existují dva základní přístupy, jejichž prostřednictvím je dané téma v současnosti studováno. Jedná se o přístupy takzvané uzavřené, nebo otevřené slovní zásoby. První z metod v podstatě představuje postup, při kterém jsou počítány četnosti předem definovaných slov, která jsou dále kategorizována. Park a kolektiv (2015) zmiňují například metodu LIWC, která určuje 60 psychologicky validních kategorií typu „funkční slova“, mezi které patří jednotlivé slovní druhy (částice, zájmena, spojky...), afektivní procesy (šťastný, brečící, nervózní...) nebo sociální procesy (partner, kamarád, povídat...).

Metoda otevřené slovní zásoby vedle toho nestanoví apriorní kategorie, místo toho usiluje o komprehenzivní extrakci jazykových vlastností. Například je počítána relativní četnost jednotlivých slov a dalších projevů (emotikon nebo pomocných znaků a podobně), frází, slov v sémantickém vztahu nebo tematických okruhů (Kulkarni et al., 2018). Kern a kolektiv (2014) uvádějí, že tento přístup, který se v současnosti zdá být dominantní, umožňuje pracovat také s lexikálními jednotkami typu „*wat*“, „*2day*“, „*u*“, „*sooooo*“, „*xxxx*“, „*mga*“, „*ttyl*“, analogicky potom s prvky na jiných úrovních jazyka.

Podle Parka a kolektivu (2014) můžeme říci, že například lidé, kteří více skórují na škále neuroticismu, více používají singulár první osoby (já, mě, moje...), zatímco extraverti mají tendenci užívat slova s pozitivním emočním zabarvením jako například „skvělý“, „šťatný“, „úžasný“. Schwarz a kolektiv (2013) analyzovali zprávy 75 000 uživatelů, od kterých zároveň měli k dispozici validní psychometrická data. Ukazují, že například lidé s neurotickými rysy častěji používají frázi „mít něčeho po krk“ nebo „mít depku“. Dále například uvádějí, že extraverti častěji používají slova jako „párty“, „láska“, „kluci“, „holky“, zatímco introverti spíše upřednostňují výrazy se vztahem k solitérní činnosti typu „počítač“, „internet“ nebo „čtení“.

U otevřenosti vůči zkušenosti měla prediktivní hodnotu slova jako „hudba“, „umění“, „psaní“, „sen“, „vesmír“ nebo „duše“. Autoři Tadesse, Lin, Xu a Yang (2018) ukazují, že hodnotu osobnostních rysů lze s pomocí jazyka spolehlivě predikovat až v 75 % případů.

2.4 Využití psychologických charakteristik

Podle Hirshe, Kanga a Bodenhausena (2012), je persvazivní komunikace (například reklamní) zvláště účinná, pokud odpovídá psychologickým charakteristikám a motivacím jedince, na kterého je směřována. Tvorbou a distribucí netriviálně psychologicky personalizovaného sdělení se má zabývat „psychografický targeting“, nebo „psychografický mikrotargeting“. V používání těchto termínů není jasný konsenzus, proto je budeme považovat za synonyma.

Jak uvádějí autoři Bhattacharyya, Bhaumik, Mukherjee a De (2019), koncept psychografického mikrotargetingu představuje proces přizpůsobení obsahu (obrazů, videa, zvuků, textu) na základě psychologických rysů a stavů jednotlivce takovým způsobem, aby vyvolal žádoucí akce na straně příjemce.

Při porovnání s psychografickým targetingem by psychografický mikrotargeting měl být založený na jemnější segmentaci. Podle Nixe (2019) můžeme v takovém případě psychografického mikrotargetingu pracovat se stovkami segmentů. Reálně se nám ale nepodařilo pro koncept dohledat reálnou empirickou oporu. V další části se textu se věnujeme studii, která se zaměřila na dva segmenty (introverty a extraverty), nemáme tedy k dispozici podklad, který by poukázal na relevanci a funkčnost několika stovek segmentů

Využití psychografické segmentace v targetingu na Facebooku

Podařilo se nám dohledat pouze jednu studii (Matz et al., 2017), která v reálných podmínkách ověřuje fungování psychografického targetingu na sociální síti Facebook. Studie je založená na datasetu MyPersonality, který jsme krátce popsali již v úvodu. Ve třech experimentech s dosahem okolo 3, 5 milionu osob autoři demonstrují, že nabídka odpovídající psychologickým

charakteristikám individua významným způsobem ovlivňuje chování měřené počtem kliknutí a nákupů. Persvazivní sdělení, které odpovídalo úrovni extraverte a otevřenosti vůči zkušenosti daného jedince, vedlo až k 40% nárůstu kliknutí a až 50% nárůstu nákupů v porovnání s neadekvátním, nebo nepersonalizovaným obsahem. Výsledky autorů naznačují, že aplikace psychologického targetingu umožňuje významným způsobem ovlivnit chování rozsáhlých skupin spotřebitelů.

První studie demonstruje vliv persvazivního psychologického targetingu na nákupní chování. Autoři přizpůsobili reklamní sdělení jednoho z britských prodejců kosmetiky na základě míry extraverte. Vzhledem k povaze produktu byla nabídka zaměřená pouze na ženy. Autoři použili pět párů sdělení v podobě obrázku a textu buď s introvertně, nebo extravertně zabarveným sdělením (obrázek 2 a 3). Příjemce zároveň rozdělili do dvou skupin na základě hodnoty predikované extraverte. Vytvořili tedy poměrně silný kvaziexperimentální mezisubjektový design se čtyřmi skupinami. První z nich tvořili introverti, kteří v jednom případě obdrželi adekvátní sdělení, tedy takové, které bylo kongruentní s jejich povahou. Ve druhém případě obdrželi obsah, který kongruentní nebyl – byl určený pro extraverty. U extravertů autoři postupovali analogicky. Na tomto základě byly vytvořeny čtyři zmíněné skupiny.

Obrázek 2. Introvertně laděné sdělení



„Beauty doesn't have to shout“

Obrázek 3. Extravertně laděné sdělení



„Dance like no one's watching (but they totally are)“

Kampaň na sociální síti Facebook experimentátoři aplikovali po dobu sedmi dní. Během této doby získala dosah více než 3,1 milionu uživatelů (762

tisíc až 814 tisíc na skupinu), zaznamenala 10 346 kliknutí (2 426 až 2 710 na skupinu) a přinesla 390 nákupů. V případě introvertů se jednalo o četnost 121 při kongruentním a 90 při nekongruentním cílení. U extravertů šlo o 117 při kongruentním a 62 při nekongruentním. Na základě hierarchického logisticky regresního modelu autoři došli k výše uvedeným výsledkům – tedy až 50% nárůstu nákupů. Efekt zůstal robustní, pokud byl kontrolován věk a jeho interakce s reklamním sdělením. Signifikantních výsledků autoři dosáhli také při použití lineárního pravděpodobnostního modelu nebo při testu chí-kvadrát. Logicky bylo dosaženo také významných rozdílů v ukazatelích jako cena za konverzi (CPC), nebo návratnost investic (ROI).

Druhý experiment, který autoři provedli, byl založený na obdobném modelu. Byl použit jiný produkt (jednalo se o mobilní aplikaci), kampaň vedle Facebooku probíhala také na Instagramu po dobu 12 dní a targetingovou proměnnou byla v tomto případě otevřenost. Přestože rozptyl v dosahu jednotlivých sdělení byl v poměru k dosahu u jednotlivých skupin (8 926 až 29 277) vyšší, autoři opět naměřili signifikantní rozdíly. Pomocí hierarchické logistické regrese autoři popsali 38% nárůst kliknutí a 31% nárůst nákupů aplikace u kongruentního sdělení. Autoři zároveň kontrolovali věk, pohlaví a interakci s osobností.

Ve třetím případě autoři vytvořili reklamu pro mobilní hru. Targetingovou proměnnou odvodili na základě dat, která získali v rámci projektu MyPersonality. Nejprve určili podobné hry, o jejichž uživatelích měli díky lajkům informace. Zjistili, že uživatelé dané hry vykazují výraznou introverzi, tedy opět určili introverzi jako základní prediktor. Cílem sice bylo vytvořit sdělení pro existující produkt, je ale patrné, že toto nastavení v podstatě kopíruje předchozí experimenty.

Během kampaně trvající sedm dní, autoři získali dosah více než půl milionu uživatelů, 3 173 kliknutí a 1 837 instalací. Dosáhli o 30 % více kliknutí a o 20 % více instalací než standardní reklama.

3 On-line data a osobnost člověka

Podle Kosinského, Stillwella a Graepela (2013) se stále zvyšuje podíl lidských aktivit, které jsou zprostředkovány pomocí digitálních služeb a technologií. Může se jednat o sociální interakce, zábavu, nakupování nebo získávání informací. Digitálně zprostředkovaná aktivita může být snadno zaznamenávána a analyzována.

3.1 On-line data a „reálný svět“

Jedna ze základních otázek, kterou můžeme v souvislosti s on-line sběrem dat položit, se týká toho, zda jsou dané informace validní ve vztahu k off-line prostředí. Jestli rozšiřují naše deskripční, predikční a explanační možnosti a jestli s jejich pomocí můžeme efektivně řešit problémy ve světě, který tradičně označujeme jako reálný.

Studium digitálních stop je atraktivní pro řadu oborů, například informatiku, zdravotnictví nebo sociální vědy (Inkster, Stillwell, Kosinski, & Jones, 2016; Kosinski, Matz, Gosling, Popov, & Stillwell, 2015; Lazer et al., 2009). Zároveň byla v jednotlivých odvětvích podniknuta celá řada zjištění.

Příkladem ze zdravotnictví může být studie autorů (Eichstaedt et al., 2015) kteří ukazují, že lexikum používané obyvateli různých států na sociálních sítích (upozorňujeme, že se v tomto případě jedná o sociální síť Twitter) nejenže koresponduje s mírou stresu uživatelů, ale zároveň umožňuje predikovat podíl úmrtí vlivem kardiovaskulárních onemocnění v daných populacích. Další autoři Birnbaum a kolektiv (2018) demonstrují úspěšné využití dat z komunikace na sociálních sítích zpracovaných metodami strojového učení při predikci psychotických symptomů a relapsu onemocnění. Podobných výsledků dosahují autoři De Choudhuryová, Counts a Horvitz (2013) při predikci depresivního syndromu.

3.2 On-line data a algoritmizovaná predikce osobnosti

Někteří autoři (Back et al., 2010) uvádějí, že profily na sociálních sítích představují validní reprezentaci osobnosti reálných jedinců. Zároveň se ukazuje, že na základě on-line profilu dokáží lidé poměrně spolehlivě odhadnout základní osobnostní rysy daného uživatele (Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011). Zdá se, že bezprostřední a interaktivní povaha sociálních sítí spojená s neustálým růstem jejich využití vede k tvorbě masivních, přirozeně strukturovaných a ekologicky validních dat v podobě digitálních stop (Azucar, Marengo, & Settanni, 2018).

Youyou, Kosinski a Stillwell (2015) dokládají, že algoritmizovaná predikce osobnosti provedená počítačem (autoři popisují predikci provedenou pomocí regresních modelů) může být přesnější než ta, kterou provádí člověk. V obou případech úspěšnost úzce souvisí s množstvím informací, které lze k posouzení využít. Autoři například uvádějí, že stačí 100 lajků na Facebooku (uživatelem označených stránek označených jako To se mi líbí) aby na jejich základě počítačový algoritmus založený na regresním modelu dosáhl shodných výsledků, při odhadu rysů osobnosti jedince, jako člověk v „průměrně blízkém vztahu“ k jedinci (podrobněji dále). Podle autorů existuje významný vztah například mezi rysem otevřenosti a lajky stránek jako Salvator Dalí, meditace, nebo TED talks. V případě extravertů se jedná například o stránky spojené s párty životem, tancem nebo reality show.

3.3 Validita algoritmizované predikce

Autoři Youyou et al. (2015) dále uvádějí tři způsoby, v jejichž intencích přesnost člověka a počítače srovnávají. Jedná se o shodu při „*self-other*“ posouzení, shodu mezi posuzovateli a externí validitu. V následujících odstavcích krátce popíšeme všechny tyto přístupy.

„Self-other“ shoda

V první případě se jedná o shodu mezi vlastním posouzením subjektu a posouzením provedeným vnějším pozorovatelem (počítač, nebo reálný člověk). Na základě výsledků lze souhrnně říci, že lineárnímu regresnímu modelu

stačí 10 lajků na to, aby dosáhl lepší shody při *self-other* posouzení rysů Big Five než průměrný kolega v práci, 70 – než přítel nebo spolubydlící, 150 – než rodinný příslušník a 300 – než partner (Youyou et al., 2015).

Sebeposouzení bylo v tomto případě provedeno pomocí stopoložkového „*International Personality Item Pool*“ (IPIP) dotazníku, zaměřeného na známých pět faktorů osobnosti. Metodu lze považovat za alternativní, ne zcela ekvivalentní k dotazníkům NEO. Vnější posuzovatelé potom využívali zkrácenou, desetipoložkovou verzi téhož nástroje – tato skutečnost se zdá být limitující a podle našeho názoru by mohla významným způsobem ovlivnit výsledky. Ve všech případech se jednalo přátele na Facebooku vzhledem k hodnocené osobě, a zároveň byli na základě informací z profilu klasifikováni jako kolegové v práci, přátelé, spolubydlící, rodinní příslušníci nebo partneři. Data o lajcích v tomto případě využíval pouze počítač, reální posuzovatelé k využití online dat nebyli instruováni.

Nejlepších výsledků dosahoval algoritmus při počtu 500 lajků, větší počet již nepřinášel významné zlepšení. Při vyjádření pomocí Personova korelačního koeficientu se jednalo o vztah mezi *self-other* posouzením o hodnotě $r = 0,66$. Vzhledem k tomu, že průměrný počet lajků jednoho uživatele je výrazně menší, autoři uvádějí číslo 227, reálné možnosti jsou poněkud nižší. Při tomto počtu autoři dosáhli výsledku $r = 0,56$. Reální lidé přitom průměrně dosahovali výsledku $r = 0,49$ (jedná se o agregovanou hodnotu kolegů v práci, spolubydlících, přátel, rodiny a partnerů). Tyto koeficienty vyjadřují průměrnou hodnotu při predikci pěti rysů Big Five, největší shoda byla konzistentně dosahována při posuzování otevřenosti, nejnižší u neuroticismu.

Shoda mezi posuzovateli

Dalším způsobem, kterým Youyou a kolektiv (2015) posuzovali validitu svého přístupu je shoda mezi různými posuzovateli. V případě lidského posuzování použili subset téměř čtrnácti a půl tisíce participantů, kteří byli posouzeni dvěma přáteli. V případě algoritmu byly využity dva separátní regresní modely, které měly k dispozici polovinu náhodně vybraných lajků.

Také v tomto případě dosáhly regresní modely signifikantně lepších výsledků $r = 0,62$ v porovnání s $r = 0,38$ u lidí.

Zároveň je podle nás nutné k takovýmto zjištěním přistupovat kriticky. Přestože se jednalo o dva nezávislé modely, které navíc dostaly náhodně vybranou množinu lajků, zjištění nedokládá externí validitu. Navíc byli do vzorku zahrnuti všichni přátelé bez ohledu na to, zda se jednalo o kolegy v zaměstnání, nebo partnery.

Externí validita

Třetí přístup je založený na skutečnosti, že autoři měli mimo dat z IPIP přístup také k informacím z dotazníků, u kterých byla už dříve zjištěna souvislost s osobnostními rysy, a to od stejných uživatelů. Youyou a kolektiv (2015) tedy měli u rozsáhlých vzorků uživatelů Facebooku informace o self-reportu životní spokojenosti, depresivité, politické orientaci, sebeřízení, impulzivité, hodnotách, zájmech, předmětu studia, fyzickém zdraví, charakteristik sociálního života a aktivit na Facebooku. Autoři dále srovnávali, s jakou úspěšností dokáže lineární, nebo logistický regresní model predikovat shodné výsledky – a to na základě self-reportu v delším IPIP dotazníku, posouzení vnějším posuzovatelem pomocí kratší IPIP verze, nebo zmiňovaných lajků.

Souhrnně lze říci, že pouze v jednom ze třinácti případů dosáhli vnější posuzovatelé lepších výsledků (jednalo se o životní spokojenost), než predikce založená na lajkách – použití zkrácené, desetipoložkové verze IPIP by ovšem mohlo představovat limitaci. V některých položkách, jako například užívání drog, nebo aktivit na Facebooku dosáhla predikce na základě lajků dokonce lepších výsledků než self-report ve stopoložkové verzi IPIP.

4 Zkoumaný problém, cíle a použité metody

Základním tématem, na které se tato práce zaměřuje je problém manifestace psychologických charakteristik v aktivitě uživatelů na sociální síti Facebook. Předpokládáme, že zjištění v dané oblasti by mohla přispět k optimalizaci uživatelské zkušenosti například v podobě jemnější segmentace, a tedy přesnějšího cílení reklamních sdělení. Jak už jsme zmínili dříve, hledání nových možností a způsobů segmentace je považováno za aktuální téma (Chen, Cheng & Chen, 2019).

Zároveň je zjevné, že téma psychografického využití dat získaných z Facebooku je spojené s mediálními kauzami doslova celosvětového významu a je předmětem debat s globálním dosahem (Hern, 2018). Také proto považujeme za důležité věnovat se věcné rovině této problematiky.

4.1 Cíl práce

Pro tuto práci jsme stanovili dva cíle:

- Na základě odborné literatury identifikovat druhy uživatelské aktivity, jejichž prostřednictvím se manifestují psychologické charakteristiky (podle Big Five) uživatelů Facebooku. Tento cíl bude splněn v podkapitole Manifestace psychologických charakteristik.
- Pomocí běžně dostupných prostředků ověřit přesnost, s jakou lze statisticky odhadnout hodnoty těchto charakteristik. Tento cíl bude splněn v kapitole Výsledky.

4.2 Data a použité metody

Využili jsme datovou sadu o uživatelích Facebooku, která je volně dostupná na internetu (Kosinski, 2016). Byla vytvořená v rámci projektu MyPersonality.org, pravděpodobně mezi lety 2007-2012. Jedná se pouze o americkou populaci.

Dataset se skládá ze tří souborů. První z nich obsahuje informace o 110 728 uživatelích. Jedná se o celkem osm proměnných – pohlaví, věk, politickou orientaci a hodnoty v pěti dimenzích Big Five. Hodnoty byly vypočítané na základě stopoložkové verze dotazníku IPIP. Záhlaví uvádíme v tabulce (13).

Tabulka 13. Záhlaví datasetu o uživatelích. Součástí je anonymizované jméno uživatele, pohlaví, věk, politická orientace (demokrat, nebo republikán) a pět položek Big Five (otevřenost, svědomitost, extraverte, svědomitost a neuroticismus).

	userid	gender	age	political	ope	con	ext	agr	neu
1	54f34605aebd63f7680e37ffd299af79	0	33	0	1,26	1,65	1,17	-1,76	0,61
2	86399f8c44ba54224b2e60177ca89fa9	1	35	0	1,07	0,17	-0,14	1,49	0,3
3	84fab50f3c60d1fdc83aa91b5e584a78	1	36	0	0,89	1,28	0,86	1,07	0,99
4	f3b8fdaccc12ef6352bfad4d6052fe9	0	39	NA	0,33	-1,01	-0,33	-0,68	0,92
5	8b06ea5e9cb87c61da387995450607f7	0	31	NA	0,15	0,47	1,17	-1,01	-0,32
6	7a2ec9c1de4ec137367e66ad759ec848	0	38	NA	1,26	0,76	-0,46	-1,76	-0,76

Další soubor obsahuje data o objektech na Facebooku a s nimi spojených lajkách. Objektem v tomto případě označujeme jakýkoli obsah, se kterým uživatel může interagovat pomocí lajku. Typicky se může jednat o stránku nebo příspěvek. Záhlaví je uvedeno v tabulce (14).

Tabulka 14. Záhlaví datasetu o lajkách a objektech na Facebooku. Objektem může být například stránka, sport, kniha, fotografie nebo příspěvek přítele.

	likeid	name
1	3c1636c878e6eb2acfd00c6b61086e38	REIGN by Paul Gibson
2	feca46ddb8ef04f86172ace0cb7e004c	Cupcake Wishes & Birthday Dreams
3	b65f46d64c688fe98bdbcf93a76a71fc	Yo también me rei de la caída de otro jejeje
4	9c5c8bb82d2cd46fbd7582f944fe370e	Abraham Joshua Heschel Day School- Alumni Network
5	2d82fa84ad79b085dc516dde154327a2	Kennesaw Farmer's Market
6	0a7a01c82143347fc4703faa4e7f415c	Karlsruher SC

Poslední soubor obsahuje data o párech uživatel–lajk. Tyto páry umožňují vytvořit asociaci mezi uživatelem a objektem v podobě lajku. V tabulce (15) opět uvádíme záhlaví.

Tabulka 15. Data zachycující pár uživatel-lajk.

	userid	likeid
1	71bc7c0901488aec6d30f0add257e7c5	3c1636c878e6eb2acfd00c6b61086e38
2	978ab8e90c4d6ad1a48ef5c973b62f4d	feca46ddb8ef04f86172ace0cb7e004c
3	85123b0e358907725cf19a2cb0ec3983	b65f46d64c688fe98dbcf93a76a71fc
4	ce110562b3e2f7e5cad3775b32d9caa5	b65f46d64c688fe98dbcf93a76a71fc
5	8188d20745471273fa69ba44a5b28473	b65f46d64c688fe98dbcf93a76a71fc
6	e9c621f322640e5f3b3842d19b38aaa8	b65f46d64c688fe98dbcf93a76a71fc

Použité metody

K analýze byl využit standardní kancelářský notebook vybavený programem RStudio ve verzi 1.1.456. Do analýzy byly zapojeny knihovny dplyr, Matrix, ROCR, topicmodels a irlba. Ze statistických metod jsme využili základní deskriptivní ukazatele jako četnost, podíl, průměr, medián, směrodatná odchylka, minimum a maximum. Z pokročilejších nástrojů jsme využili metodu singulárního rozkladu (SVD – *singular value decomposition*), rotaci varimax, logistickou a lineární regresi a také křížovou k -fold validaci. Tento postup jsme převzali od Kosinského (2016), v podstatě se tedy jedná o replikaci.

5 Výsledky

V této části práce jsme replikovali analýzu, jejíž popis publikovali Kosinski a kolektiv (2016) a která je k nalezení také na webových stránkách autora (Kosinski, 2016). Pracovní postup jsme mírně modifikovali, proto jsme dospěli k lehce odlišným výsledkům.

Na základě výše zmíněných dat a za použití uvedených metod jsme nejprve vytvořili datovou matici reprezentující uživatele, objekty a lajky mezi nimi. Tato matice měla více než sto tisíc řádků, více než milion a půl sloupců a více než deset milionů hodnot (v našem případě jedniček reprezentujících lajk). Tuto matici jsme pomocí cyklu upravili tak, aby obsahovala pouze uživatele s alespoň padesáti lajky, u objektů jsme zvolili dvojnásobek, tedy sto. Tímto způsobem vznikla významně menší matice, která byla více saturovaná daty. I po této úpravě se její „hustota“ blížila pouze dvěma procentům.

5.1 Deskripce zkoumaného souboru

V tabulce (16) uvádíme vybrané vlastnosti zkoumaného vzorku (upravená matice), k dispozici jsou také údaje za původní soubor. Součástí zkoumaného vzorku je více než 22 tisíc uživatelů a více než 12 tisíc objektů.

Tabulka 16. Základní charakteristiky původní a upravené matice

Veličina	Původní matice	Upravená matice
Uživatelé (n)	110 728	22 144
Objekty (n)	1 580 284	12 620
Páry uživatel-lajk (n)	10 612 326	5 525 719
Hustota matice	0,01 %	1,98 %
Objekty na uživatele		
Průměr	96	200
Medián	22	108
Minimum	1	50
Maximum	7973	2812
Směrodatná odchylka	1	264
Uživatelé na objekt		
Průměr	7	351
Medián	1	209
Minimum	1	100
Maximum	19998	9099
Směrodatná odchylka	70	440

V tabulce (17) uvádíme deskriptivní přehled a srovnání spojitých proměnných u zkoumaného vzorku a toho původního. U některých položek, věku a otevřenosti, bychom pravděpodobně zjistili statisticky významné rozdíly. Pro naši analýzu to ovšem nepovažujeme za důležité.

Tabulka 17. Původní matice (n = 110 728), upravená matice (n = 22 144), O = otevřenost, S = svědomitost, E = extraverte, P = Přívětivost, N = neuroticismus

Veličina	Věk	O	S	E	P	N
Původní matice						
Průměr	29,7	0,0	0,0	0,0	0,0	0,0
Medián	27,0	0,2	0,0	0,1	0,1	-0,1
Minimum	18,0	-5,5	-3,7	-3,1	-4,3	-2,1
Maximum	80,0	1,9	2,2	1,9	2,3	2,8
Směrodatná odchylka	7,9	1,0	1,0	1,0	1,0	1,0
Upravená matice						
Průměr	27,6	0,2	-0,1	0,0	0,0	0,1
Medián	26,0	0,3	-0,1	0,1	0,1	0,0
Minimum	18,0	-5,2	-3,7	-3,1	-4,3	-2,1
Maximum	80,0	1,9	2,2	1,9	2,3	2,8
Směrodatná odchylka	6,7	0,9	1,0	1,0	1,0	1,0

Je evidentní, že zkoumaný soubor se z pohledu pohlaví a věku významně liší od současné populace Facebooku. V tabulce (18) uvádíme současná data a také hodnoty za zkoumaný vzorek. Hlavními důvody může být obecně větší míra participace žen v online šetřeních (vyplnění dotazníku bylo dobrovolné), z pohledu věku by mohl hrát roli také přibližně než desetiletý odstup (data z MyPersonality jsou přibližně deset let stará).

Tabulka 18. Charakteristiky pohlaví a věku u současné populace F., původní a upravené matice

Věková kategorie	Podíl podle pohlaví a věkové kategorie pohlaví	
Populace Facebooku	Ženy (44%)	Muži (56%)
18-24	24 %	27 %
25-34	32 %	36 %
35-44	19 %	18 %
45-54	12 %	10 %
55-64	8 %	5 %
65+	5 %	4 %
Původní matice (n = 110 728)	Ženy (56%)	Muži (44%)
18-24	24 %	25 %
25-34	54 %	60 %
35-44	14 %	10 %
45-54	6 %	4 %
55-64	2 %	1 %
65+	0 %	0 %
Upravená matice (n = 22 144)	Ženy (58%)	Muži (42%)
18-24	38 %	34 %
25-34	50 %	56 %
35-44	8 %	7 %
45-54	3 %	2 %
55-64	1 %	1 %
65+	0 %	0 %

5.2 Regresní analýz a predikce psychologických charakteristik

Pro popis vztahu mezi lajky a závislými proměnnými, především rysy podle Big Five, jsme využili modely lineární a také logistické regrese. Pro tento účel bylo nutné provést další úpravy matice. Za účelem redukce její dimenzionality jsme využili metodu singulárního rozkladu (SVD). Dílčím a pravděpodobně známějším případem této metody je analýza hlavních komponent (PCA). Tímto způsobem jsme získali 50 faktorů, které byly rotovány pomocí rotace varimax.

Při konstrukci regresních modelů jsme lajky uživatelů (redukované výše popsaným způsobem) uvažovali jako nezávislou proměnnou, s jejíž pomocí jsme odhadovali hodnotu nezávisle proměnné – pohlaví, politickou orientaci, věk a pět dimenzí Big Five. Vzhledem k tomu, že pohlaví i politická orientace jsou binárními veličinami, využily jsme logistický model. V případě jiných proměnných byla využita lineární regrese.

Pomocí k -fold křížové validace jsme získali informace o přesnosti našich modelů. Vzhledem k rozsáhlejšímu charakteru celé analýzy prezentujeme pouze tento výsledek. Přesný postup výpočtu včetně dalších vlastností modelů přikládáme na CD jako přílohu této práce.

V rámci křížové validace byla data rozdělena na deset subsetů o stejné velikosti ($k = 10$). Následně byl model trénován na $k - 1$ subsetech a zvalidován na jednom zbývajícím subsetu. Tento proces byl desetkrát zopakován a výsledkem je průměrná přesnost za všechny pokusy. Hodnoty pro jednotlivé proměnné uvádíme v tabulce (19).

*Tabulka 19. Přesnost logistických a lineárních modelů.
AUC = plocha pod ROC křivkou*

Proměnná	Přesnost modelu
Pohlaví (AUC)	.94
Politická orientace (AUC)	.86
Věk	.58
Otevřenost	.43
Svědomitost	.26
Extraverze	.30
Přívětivost	.26
Neuroticismus	.33

Je zjevné, že největší přesnosti dosáhl logistický model při klasifikaci pohlaví a politické orientace. U spojitých proměnných bylo pomocí modelu lineární regrese dosaženo nejlepších výsledků u věku a otevřenosti. Vysoká hodnota u otevřenosti potvrzuje psychometrickou praxi. U zbývajících dimenzí se přesnost konstantně pohybuje okolo 30 %, zde bychom očekávali spíše větší rozdíly.

Diskuze

Výběr obsahu, který se zobrazí konkrétnímu uživateli, je klíčový pro pozitivní uživatelskou zkušenost. Jak už jsme uvedli v úvodu, cílem Facebooku je pomocí nejrůznějších prostředků tuto zkušenost zlepšovat a tím prodlužovat dobu, kterou uživatelé na síti stráví, případně uživatele motivovat k chování určitého typu.

Jeden z přístupů, který by mohl k optimalizaci uživatelské zkušenosti přispět, je spojený s otázkami po způsobu a využití manifestací psychologických charakteristik uživatelů prostřednictvím jejich aktivit. Pomocí matematických modelů je možné v datech získaných z Facebooku hledat struktury popisující vlastnosti lidí, kteří tato data vytvářejí. Pomocí regresních modelů jsme schopni aktivitu uživatelů odhadovat hodnoty rysů podle Big Five, u kterých již byla popsána jejich souvislost se spotřebitelským chováním.

Jako správné hodnotíme zaměření aktivitu uživatelů v podobě lajků. Jedná se o fenomén, se kterým se relativně snadno pracuje – není mimořádně náročný z pohledu metod ani výpočetního výkonu. Na základě datasetu obsahujícího záznamy o lajcích několika desítek tisíc uživatelů se nám pomocí lineární regrese a křížové validace podařilo dosáhnout přesnosti 26–43 % u všech pěti dimenzí. To je závěr, který odpovídá zjištěním jiných autorů. Jistou překážku by mohla představovat skutečnost, že zkoumaný soubor se z pohledu pohlaví i věku zjevně liší od současné populace uživatelů Facebooku. To může být dáno různou ochotou zapojit se do výzkumného šetření a vyplnit stopoložkový dotazník. V případě „samovýběru“ v internetovém prostředí nemůžeme mluvit reprezentativnosti zkoumaného souboru. Data navíc podléhají času – může se vyvíjet sama společnost a její chování, ale také řada internetových objektů – například blogů – dnes vůbec nemusí existovat. Otázkou je, jestli je reprezentativnost pro náš účel potřeba a další otázkou může být, k jakým závěrům nás takové okolnosti opravňují.

Byli bychom spíše obezřetní, pokud jde o vyvozování závěrů. Například Nix (2019) uvádí, že to, co označujeme jako algoritmizovaný psychografický

mikrotargetingem představuje významnou změnu ve způsobu, jakým mohou být vytvářeny a distribuovány marketingové obsahy. Uvádí, že místo toho, aby se sešly „expertní mozky“ a vytvořily jednotné a globálně mířené sdělení, které přivede adresáta k nákupu, nebo rozhodne o výsledku voleb, vznikne model, který určitým skupinám doručí specifické vysoce personalizované sdělení.

Výzkumy založené na datech ze sociálních sítí se podle nás zároveň potýkají z řadou problémů. Jedním z nich by mohla být skutečnost, že předložené výsledky budou v současnosti jen těžko naplňovat jeden ze základních předpokladů vědeckého zkoumání, kterým je **replikovatelnost**. Společnost Facebook, Inc. v minulých měsících výrazně omezila možnosti legálního využití dat pro jiné než byznysové účely. Tato skutečnost by mohla výzkumné aktivity podobného typu zcela eliminovat. Sám provozovatel největší sociální sítě v napjaté atmosféře ochrany citlivých dat z našeho pohledu sám ustoupil od sociálně-vědného výzkumu zaměřeného na uživatele a zaměřuje se spíše na techničtější aplikace v oblastech umělé inteligence nebo strojového učení.

Další problém by podle nás mohl nastat při validizaci byť sebelepších modelů v kontextu reálného fungování sociální sítě. V souvislosti s výzkumy Matze a kolektivu (2017), kteří uživatelům Facebooku prezentovali introvertně a extrovertně laděné reklamní sdělení, autoři Eckles, Gordon, a Johnson (2018) upozorňují, že sama sociální síť může validitu podobných výsledků výrazně snižovat. Pokud uvažujeme záměr výzkumníků a zároveň efekt sociální sítě (sama síť pracuje s algoritmy pro personalizaci obsahů), vy výsledku nejsme schopni tyto veličiny oddělit. A ani v případě A/B testování se podle autorů nemůžeme předpokládat větší míru objektivity a validity.

Jistou riziko podle nás může představovat **využití výzkumných souborů různé velikosti**. Využití takzvaných velkých dat musí být nutně spojeno s opatrností a řemeslnou poctivostí, pokud se chceme vyvarovat zjištění, jejichž reálný význam nebo velikost účinku jsou zanedbatelné.

Samostatným tématem pro diskuzi by mohlo být **využití faktorových modelů** lidské osobnosti v prostředí výpočetních technologií. Tyto modely jsou vytvořeny především pro přehlednost, srozumitelnost a snadné pochopení. Jinými slovy – jejich cílem je zjednodušit lidskou individualitu do faktorů, které umožňují snadnou deskripci, případně predikci chování. Tato potřeba ovšem přestává být aktuální při využití výkonné výpočetní techniky. Redukce člověka na pět proměnných v tomto kontextu pravděpodobně nemá odůvodnění. Například samotný Facebook při výběru obsahu, který se zobrazí jednotlivému uživateli, využívá až několik stovek proměnných. Zároveň není nutné omezovat zájem pouze na dílčí **zdroje informací**, jako jsou například zmiňované lajky.

Téma psychologického targetingu také velmi bezprostředně vyvolává **otázky etické**. Spíše se nekloníme k názoru, že psychologický targeting by mohl představovat bezprecedentní průnik do lidské individuality. Hrubé psychologické rysy nepokládáme za nic více osobního, citlivého a méně veřejného než jiné atributy, které korespondují s naší individualitou.

Navíc Facebook neumožňuje cílení reklamy na základě osobnostních rysů. Mimo jiné umožňuje cílení na základě chování vyjadřující příslušnost k určité sociální skupině definované například pomocí reakcí v podobě lajků vůči reálným objektům. Zjišťování, o jaké objekty jde, může být předmětem standardního marketingového výzkumu. Například údajná facebooková micro-targetingová kampaň Donalda Trumpa, nebo její část, měla být založena na psychografickém modelu několika tisíc uživatelů, jehož výsledkem měly být preference automobilových značek. Z práce Kosinského, Stillwella a Graepela (2013) víme, že například pozitivní vztah ke značce Harley-Davidson může významně negativně korelovat s inteligenčním kvociem.

Preference bezpochyby můžeme pokládat za osobní. Zároveň se jedná o tradiční doménu marketingu, jejíhož zkoumání se určitá část populace může zcela svobodně zúčastnit. Zároveň se domníváme, že získávání dat, které je

prováděno třetími stranami pomocí nejrozličnějších facebookových aplikací, může představovat etická rizika.

Závěr

Na základě odborné literatury jsme identifikovali tři druhy uživatelské aktivity, jejichž prostřednictvím jsou zkoumány možnosti manifestace psychologických charakteristik uživatelů Facebooku. Jedná se lajky, text a fotografie.

Pomocí volně dostupného datasetu a také statistického modelu jsme vyhodnotili přesnost, s jakou lze pomocí těchto aktivit, konkrétně lajků odhadovat hodnoty rysů podle Big Five. Přesnost při odhadu všech dimenzí nabývala hodnot 26–43 %. K těmto hodnotám je zároveň třeba přistupovat obezřetně, vzhledem k omezené reprezentativnosti.

Summary

Choosing content that is specifically tailored to each person is key to the positive user experience. One approach that could contribute to the optimization of user experience is related to questions about the way and use of manifestations of the psychological characteristics of users. Based on the literature we have identified three types of user activity through which the possibilities of manifestation of psychological characteristics of Facebook users are examined. These are likes, text and photos. Using a freely available dataset as well as a statistical model, we evaluated the accuracy with which these activities, namely likes, can estimate the values of features according to Big Five. The accuracy of all dimensions was 26-43%. At the same time, these values should be treated with caution, given the limited representativeness.

Seznam literatury

- Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124, 150–159. <https://doi.org/10.1016/j.paid.2017.12.018>
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science* (0956-7976), 21(3), 372–374. <https://doi.org/10.1177/0956797609360756>
- Bardakci, A., & Whitelock, J. (2003). Mass-customisation in marketing: the consumer perspective. *Journal of Consumer Marketing*, 20(5), 463. <https://doi.org/10.1108/07363760310489689>
- Belch, G. E., & Belch, M. A. (2018). *Advertising and promotion: An integrated marketing communications perspective*. New York, NY: McGraw-Hill Education.
- Bhattacharyya, S., Bhaumik, H., Mukherjee, A., & De, S. (2019). *Machine learning for big data analysis*. Berlin: De Gruyter.
- Birnbaum, M., Rizvi, A., Choudhury, M. D., Ernala, S., Cecchi, G., & Kane, J. (2018). O9.2. Identifying Psychotic Symptoms And Predicting Relapse Through Social Media. *Schizophrenia Bulletin*, 44(Suppl_1). doi:10.1093/schbul/sby015.246
- Blocker, C. P., & Flint, D. J. (2007). Customer segments as moving targets: Integrating customer value dynamism into segment instability logic. *Industrial Marketing Management*, 36(6), 810–822. <https://doi.org/10.1016/j.indmarman.2006.05.016>
- Celli, F., Bruni, E., & Lepri, B. (2014). Automatic personality and interaction style recognition from Facebook profile pictures. Proceedings of the 22nd ACM international conference on multimedia (pp. 1101–1104). ACM. <http://dx.doi.org/10.1145/2647868.2654977>.
- Dacko, S. G. (2008). *The advanced dictionary of marketing: Putting theory to use*. Oxford: Oxford University Press.
- De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. Proceedings of the 5th annual ACM web science conference (pp. 47–56). ACM.
- Eckles, D., Gordon, B. R., & Johnson, G. A. (2018). Field studies of psychologically targeted ads face threats to internal validity. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 115(23), E5254–E5255. <https://doi.org/10.1073/pnas.1805363115>
- Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... Seligman, M. E. P. (2015). Psychological language on

- Twitter predicts county-level heart disease mortality. *Psychological Science*, 26(2), 159–169. <https://doi.org/10.1177/0956797614557867>
- Facebook Ads Help Center. (2019a). About Ad Relevance Diagnostics. Získáno z <https://www.facebook.com/business/help/403110480493160>
- Facebook Ads Help Center. (2019b). About Relevance Score. Získáno z <https://www.facebook.com/business/help/768598746552352>
- Facebook, Inc. (2019). *Facebook Reports Fourth Quarter and Full Year 2018 Results*. Získáno z <https://investor.fb.com/investor-events/default.aspx>
- Facebook, Inc. (2014). *Scaling the Facebook data warehouse to 300 PB*. Získáno z <https://code.fb.com/core-data/scaling-the-facebook-data-warehouse-to-300-pb/>
- Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *CyberPsychology, Behavior & Social Networking*, 14(9), 483–488. <https://doi.org/10.1089/cyber.2010.0087>
- Hern, A. (2018). Cambridge Analytica: how did it turn clicks into votes? The Guardian. Získáno z <https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>
- Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science*, 23(6), 578–581. <https://doi.org/10.1177/0956797611436349>
- Hospodářské noviny. (2019). Reálné investice do reklamy loni vzrostly na 113,5 miliardy korun. Získáno z www.byznys.ihned.cz/c1-66454390-realne-investice-do-reklamy-loni-vzrostly-na-113-5-miliardy-korun-nejvice-utracel-kaufland-druhy-byl-lidl
- Hřebíčková, M. (2010). Nové přístupy ke zkoumání rysů: Pětifaktorový model osobnosti. In M. Blatný (Ed.), *Psychologie osobnosti: hlavní témata, současné přístupy* (43–70). Praha: Grada.
- Chen, T., Cheng, H., & Chen, Y. (2019). Developing a personal value analysis method of social media to support customer segmentation and business model innovation. *Expert Systems*, 36(3), N.PAG. <https://doi.org/10.1111/exsy.12374>
- Inkster, B., Stillwell, D., Kosinski, M., & Jones, P. (2016). A decade into Facebook: Where is psychiatry in the digital age? *The Lancet Psychiatry*, 3(11), 1087–1090. [https://doi.org/10.1016/S2215-0366\(16\)30041-4](https://doi.org/10.1016/S2215-0366(16)30041-4)
- Issenberg, S. (2016). How Obama's Team Used Big Data to Rally Voters. Získáno z <https://www.technologyreview.com/s/509026/how-obamas-team-used-big-data-to-rally-voters/>

Jen-Hung Huang, & Yi-Chun Yang. (2010). The Relationship between Personality Traits and Online Shopping Motivations. *Social Behavior & Personality: An International Journal*, 38(5), 673–679. Získáno z <http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=50874498&lang=cs&site=ehost-live>

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., ... Seligman, M. E. P. (2014). The online social self: An open vocabulary approach to personality. *Assessment*, 21(2), 158–169. <https://doi.org/10.1177/1073191113514104>

Kosinski, M. (2016). Data Mining Tutorial. Získáno z <https://www.michalkosinski.com/data-mining-tutorial>.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543–556. <https://doi.org/10.1037/a0039210>

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, 21(4), 493–506. <https://doi.org/10.1037/met0000105>

Kotler, P., & Armstrong, G. (2017). *Principles of marketing global edition*. Boston: Pearson.

Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., ... Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. *Plos One*, 13(11), e0201703. <https://doi.org/10.1371/journal.pone.0201703>

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 114(48), 12714–12719. <https://doi.org/10.1073/pnas.1710966114>

Moon, Y. (2002). Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality. *Journal of Consumer Psychology (Taylor & Francis Ltd)*, 12(4), 313–325. <https://doi.org/10.1207/15327660260382351>

Nix, A. (2019, April 31). *Cambridge Analytica - The Power of Big Data and Psychographic*. Speech presented at 2016 Concordia Annual Summit in New York, New York. Získáno z <https://www.youtube.com/watch?v=n8Dd5aVXLCc>

- O'Donohue, M., & Addison, T. (2001). Understanding the Customer's Relationship With a Brand: The Role of Market Segmentation in Building Stronger Brands. Získáno z <https://www.warc.com/fulltext/MRS/49705.htm>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934–952. <https://doi.org/10.1037/pspp0000020.supp> (Supplemental)
- Piedmont, R. L. (2014). *The Revised Neo Personality Inventory Clinical and Research Applications*. Springer Verlag.
- Segalin, F., et al. (2017). What your Facebook profile picture reveals about your personality. *Proceedings of the 2017 ACM on Multimedia Conference*, 2017.
- Sharma, R. (2019). How Does Facebook Make Money? Získáno z www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality Predictions Based on User Behavior on the Facebook Social Media Platform. *IEEE Access*, 6, 61959-61969. doi:10.1109/access.2018.2876502
- Tareaf, C., Berger, P., Hennig, P., & Meinel, Ch. (2018). *Towards Automatic Personality Prediction Using Facebook Like Categories* Raad
- The Psychometrics Centre, University of Cambridge. (2013). *MyPersonality database*. Získáno z www.psychometrics.cam.ac.uk/productsservices/Mypersonality
- Yankelovich, D., & Meer, D. (2006). Rediscovering Market Segmentation. *Harvard Business Review*, 84(2), 122–131. Získáno z <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=19406199&lang=cs&site=eds-live>
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 112(4), 1036–1040. <https://doi.org/10.1073/pnas.1418680112>

Seznam zkratek

API – application programming interface

AUC – plocha pod křivkou

CPC – cena za konverzi

EU – Evropská unie

GDPR – obecné nařízení o ochraně osobních údajů

IPIP – International Personality Item Pool

LIWC – Linguistic Inquiry and Word Count

mil. – milion

mld. – miliarda

ROI – návratnost investic