



Analytika učení a data mining ve vzdělávání v kontextu systémů pro řízení výuky

Libor Juhaňák

MASARYKOVA
UNIVERZITA



#525

OPERA FACULTATIS PHILOSOPHICAE
UNIVERSITATIS MASARYKIANAE

SPISY FILOZOFICKÉ FAKULTY
MASARYKOVY UNIVERZITY

MUNI
ARTS



Analytika učení a data mining ve vzdělávání v kontextu systémů pro řízení výuky

Libor Juhaňák

**MASARYKOVA
UNIVERZITA**

BRNO 2023

KATALOGIZACE V KNIZE – NÁRODNÍ KNIHOVNA ČR

Juhaňák, Libor, 1986-

Analytika učení a data mining ve vzdělávání v kontextu systémů pro řízení výuky / Libor Juhaňák.
– Vydání první, elektronické. – Brno : Masarykova univerzita, 2023. – 1 online zdroj. – (Opera
Facultatis philosophicae Universitatis Masarykianae = Spisy Filozofické fakulty Masarykovy univerzity,
ISSN 2787-9291 ; 525)

Anglické resumé

Obsahuje bibliografii a bibliografické odkazy

ISBN 978-80-280-0185-8 (online ; pdf)

* 37.012 * 004.9:37.018.43 * 004.659 * 004.62 * (048.8)

- pedagogický výzkum
- systémy pro řízení výuky
- dolování dat
- analýza dat
- analytika učení
- monografie

004.4/.6 - Programování. Software [23]

Recenzovali: doc. RNDr. Jana Straková, Ph.D. (Univerzita Karlova)
PhDr. Lucie Rohlíková, Ph.D. (Západočeská univerzita)

© 2023 Masarykova univerzita, Libor Juhaňák

ISBN 978-80-280-0185-8

ISBN 978-80-280-0184-1 (brožováno)

ISSN 1211-3034 (print)

ISSN 2787-9291 (online)

<https://doi.org/10.5817/CZ.MUNI.M280-0185-2023>

Obsah

1 ÚVOD	9
2 VZNIK A HISTORICKÉ KOŘENY ANALYTIKY A DATA MININGU VE VZDĚLÁVÁNÍ	13
2.1 Historické kořeny data miningu ve vzdělávání	17
2.2 Historické kořeny analytiky učení	19
2.3 Další oblasti formující analytiku učení a data mining ve vzdělávání	21
3 VYMEZENÍ ANALYTIKY A DATA MININGU VE VZDĚLÁVÁNÍ A JEJICH ROZDÍLY	25
3.1 Vymezení data miningu ve vzdělávání	25
3.2 Vymezení analytiky učení	27
3.3 Odlišnosti mezi analytikou učení a data miningem ve vzdělávání	29
3.3.1 Rozdílný vznik a vývoj v čase	30
3.3.2 Rozdílné využívání analytických metod a technik	32
3.3.3 Rozdílné zaměření pozornosti	33
4 PROCES ANALÝZY A TYPY VYUŽÍVANÝCH DAT	35
4.1 Proces aplikace analytiky učení a data miningu ve vzdělávání	38
4.2 Typy dat využívaných v oblasti analytiky učení a data miningu ve vzdělávání	43
5 METODY A TECHNIKY ANALÝZY DAT	49
5.1 Prediktivní metody	49
5.2 Shlukování	51
5.3 Metody redukce dimenzí	52
5.4 Dolování vztahů	54
5.5 Analýza sociálních sítí	55
5.6 Dolování textů a zpracování přirozeného jazyka	57
5.7 Objevování pomocí modelů	58
6 DOSAVADNÍ SMĚRY VÝZKUMŮ A ŘEŠENÁ TÉMATA	59
6.1 Analytické nástroje pro sumarizaci a vizualizaci dat	61
6.2 Predikce úspěšnosti a identifikace rizikových studentů	63
6.3 Modelování studentů	65
6.3.1 Modelování znalostí a dovedností studentů	66
6.3.2 Modelování chování studentů	68

6.3.3 Modelování emocí a afektivních stavů	70
6.4 Adaptivní a doporučovací systémy	72
6.5 Sociální analytika učení	75
6.5.1 Analytika sociálních sítí	76
6.5.2 Analytika obsahu	78
6.5.3 Analytika diskurzu	79
6.6 Multimodální analytika učení	80
6.7 Etické aspekty analytiky učení a data miningu ve vzdělávání	84
7 ANALÝZA 1: PREDIKCE ÚSPĚŠNOSTI STUDENTŮ V BLENDED LEARNINGOVÝCH KURZECH	89
7.1 Přehled dosavadních výzkumů	91
7.1.1 Predikce úspěšnosti studentů v kontextu blended learningu a LMS	92
7.1.2 Temporální charakteristiky návštěvnosti jako prediktory úspěšnosti	95
7.1.3 Otázky zobecnitelnosti a přenositelnosti prediktivních modelů	97
7.2 Řešený problém a otázky	98
7.3 Metodologie	100
7.3.1 Metodologický přístup – klasifikace	100
7.3.2 Analyzovaná data, sběr a předzpracování dat	102
7.3.3 Použité proměnné	103
7.3.4 Postup analýzy dat	107
7.4 Výsledky	109
7.4.1 Základní posouzení predikčních schopností analyzovaných klasifikátorů	110
7.4.2 Hodnocení přínosnosti použitých prediktorů	115
7.4.3 Predikční testování externí validací	118
7.5 Diskuze a závěr	120
7.5.1 Shrnutí odpovědí na stanovené otázky	120
7.5.2 Důsledky pro praxi a další výzkum	123
8 ANALÝZA 2: INTERAKCE STUDENTŮ A UČITELŮ V ONLINE DISKUZNÍCH FÓRECH	125
8.1 Přehled dosavadních výzkumů	126
8.1.1 Participace studentů v online diskuzích a její výzkum	128
8.1.2 Analýza sociálních sítí v kontextu e-learningu a LMS	131
8.1.3 Analýza sociálních sítí a interakce v online diskuzních fórech v LMS	132
8.2 Řešený problém a otázky	135
8.3 Metodologie	136

8.3.1 Metodologický přístup – analýza sociálních sítí	136
8.3.2 Postup sběru a předzpracování dat	138
8.3.3 Analýza dat	140
8.4 Výsledky	142
8.5 Diskuze a závěr	162
8.5.1 Shrnutí odpovědí na stanovené otázky	162
8.5.2 Limity a možnosti dalšího výzkumu	165
9 ANALÝZA 3: CHOVÁNÍ STUDENTŮ V PRŮBĚHU PLNĚNÍ ONLINE TESTŮ	167
9.1 Přehled dosavadních výzkumů	168
9.1.1 Přístupy založené na person-fit indexech	170
9.1.2 Přístupy založené na rychlosti odpovídání	172
9.1.3 Přístupy vycházející z data miningu ve vzdělávání a analytiky učení	173
9.2 Řešený problém a otázky	175
9.3 Metodologie	176
9.3.1 Metodologický přístup – dolování procesů	176
9.3.2 Analyzované případy	177
9.3.3 Sběr a předzpracování dat	179
9.3.4 Analýza dat	181
9.4 Výsledky	181
9.5 Diskuze a závěr	187
9.5.1 Shrnutí odpovědí na stanovené otázky	187
9.5.2 Limity, možnosti dalšího výzkumu a důsledky pro praxi	189
10 ZÁVĚR	191
SUMMARY	195
BIBLIOGRAFIE	198
SEZNAM TABULEK	220
SEZNAM OBRÁZKŮ	222

1 ÚVOD

*It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts.*

Sherlock Holmes
Arthur Conan Doyle (1891)
A Scandal in Bohemia

Problematiku informačních a komunikačních technologií (ICT) můžeme v kontextu současného pedagogického výzkumu považovat za jedno z dominantních témat. Přestože by bylo možné argumentovat, že český pedagogický výzkum této problematice nevěnuje až tolik pozornosti (srov. Zounek & Tůma, 2014), množství existujících zahraničních odborně zaměřených pedagogických časopisů specializujících se přímo na téma digitálních technologií ve vzdělávání nás nenechá na pochybách, že jde o problematiku, jíž se zabývá značné množství výzkumníků a odborníků, a to nejen přímo v kontextu vzdělávání a pedagogického výzkumu. Pozornost věnovaná moderním technologiím přitom není nijak zvlášť překvapivá, uvědomíme-li si, že se informační a komunikační technologie postupem času staly přirozenou součástí společnosti a každodenního života většiny lidí. Digitální technologie jsou tak v současnosti běžně využívány v dlouhé řadě odvětví lidské činnosti, včetně vzdělávání a školství.

Jedním z důvodů, proč si problematika informačních a komunikačních technologií získala v pedagogickém výzkumu tak velkou pozornost, je jistě i to, že jde o téma velmi různorodé a široké, a tudíž v řadě dílčích oblastí stále relativně nové a neprobádané. Výzkumníci se tak mohou této problematice věnovat ve velkém množství kontextů, v odlišných souvislostech a z rozdílných úhlů pohledu. Ať už v rámci jednotlivých pedagogických disciplín (obecná pedagogika, speciální pedagogika, andragogika, pedagogická evaluace a další), v souvislosti s různými aktéry vzdělávání (student, učitel) či třeba s ohledem na různé typy technologií, které mohou být ve vzdělávání využívány (od počítačů přes tablety a chytré telefony až třeba po online výukové hry nebo tzv. rozšířenou či virtuální realitu). S neutuchajícím vývojem v oblasti digitálních technologií, který souvisí mimo jiné

s exponenciálním růstem výpočetního výkonu počítačů (Moore, 1965), se navíc objevují stále nové technologie a možnosti jejich využití ve vzdělávání.

Přesto i v rámci výzkumu digitálních technologií ve vzdělávání existují oblasti, kterým je pozornost věnována poměrně soustavně a v současnosti lze již říci, že i relativně dlouhodobě. Jednou z takových oblastí jsou systémy pro řízení výuky (*learning management systems*)¹, běžně označované zkratkou LMS. Je to dáno tím, že systémy typu LMS se v kontextu vzdělávání staly prakticky „technologickým standardem“. Přestože je situace v oblasti využívání ICT ve školách a vzdělávání obecně přirozeně poměrně odlišná (a to jak napříč zeměmi, tak i napříč jednotlivými vzdělávacími institucemi), systémy typu LMS lze zvláště v případě institucí terciárního vzdělávání považovat v současnosti za široce až univerzálně rozšířené (srov. Ferguson, 2012a; Macfadyen & Dawson, 2012; Poulová, 2010).

Implicitní součástí většiny systémů typu LMS je přitom to, že automaticky sbírají a uchovávají data o svých uživateliích či o tom, jakými způsoby uživatelé daný systém využívají, přičemž tato data mohou být zároveň poměrně bohatá. V případě studentů, jakožto jedněch z hlavních uživatelů těchto systémů, může jít například o informace o kurzech, které studují či které již absolvovali. Může jít také o studijní výsledky v jednotlivých kurzech, a to i v rámci dílčích výukových aktivit daného kurzu. Stejně tak mohou systémy typu LMS uchovávat informaci o komunikaci studentů s vyučujícím a s ostatními studenty kurzu či např. informaci o tom, jak často a jak aktivně se vlastně studenti přihlašují do odpovídajícího kurzu v LMS. Až do relativně nedávné doby však zůstávaly tyto online výukové systémy v určitém ohledu černými skříňkami. Přestože řadu let automaticky shromažďovaly a ukládaly výše nastíněné typy informací a dat, jen ojedinele byla tato data systematicky využívána pro účely výzkumu či v souvislosti se zkvalitňováním vzdělávání prostřednictvím těchto výukových systémů.

Výraznější změna nastala teprve v posledních letech spolu s rozvojem specifických výzkumných oblastí, jež začaly být označovány jako data mining ve vzdělávání (*educational data mining*) a analytika učení (*learning analytics*). Oblasti data miningu ve vzdělávání začala být věnována zvýšená pozornost na přelomu let 2008 a 2009 (Romero & Ventura, 2013), oblast analytiky učení pak zaznamenala výraznější rozvoj zvláště po roce 2012, přestože za její „zrod“ lze považovat přelom let 2010 a 2011 (Ferguson, 2012a, 2012b; Juhaňák & Zounek, 2016, 2019; Siemens, 2013). Přestože jsou mezi data miningem ve vzdělávání a analytikou učení dílčí rozdíly, hlavním cílem obou těchto výzkumných oblastí je snaha co nejvíce využít potenciál unikátních dat, která jsou generována a ukládána různými online vzdě-

1 Systémy typu LMS lze chápat jako specifický typ online vzdělávacích systémů, který je charakteristický tím, že se pomocí integrace různých typů nástrojů (např. nástroje pro tvorbu a prezentaci obsahu, nástroje pro komunikaci, nástroje pro hodnocení, nástroje pro administraci studia atd.) snaží podporovat vzdělávací proces v jeho celistvosti (srov. Kats, 2010; Kitsantas & Dabbagh, 2010; Zounek & Sudický, 2012).

lávacími prostředím a systémy, mezi než spadají mimo jiné právě systémy typu LMS. Pro tyto účely jsou přitom vyvíjeny a následně používány různé analytické a data miningové metody a techniky umožňující získat důležité informace a poznatky o tom, jak se studenti v těchto online výukových prostředích chovají a učí.

Tato monografie vychází z mého dlouhodobého zájmu o problematiku digitálních technologií ve vzdělávání, a zvláště pak o systémy typu LMS. Těmto systémům se věnuji již více než deset let a při jejich využívání jsem postupně získával zkušenosti nejen z pozice běžného uživatele, ale také z pozice člena uživatelské podpory, e-learningového metodika, testera, administrátora či výzkumníka. A ačkoli jsem při psaní využil zkušenosti ze všech zmiňovaných rolí, primárně v této publikaci přistupuji k systémům typu LMS z pozice výzkumníka. Monografie vychází z mé disertační práce obhájené na Ústavu pedagogických věd Filozofické fakulty Masarykovy univerzity, v níž jsem se zaměřoval na možnosti využití metod analytiky učení a data miningu ve vzdělávání pro účely analýzy dat pocházejících z LMS Moodle. V kontextu českého pedagogického výzkumu totiž analytika učení a data miningu ve vzdělávání doposud nebyla věnována téměř žádná či jen velmi ojedinělá pozornost. A to i přesto, že v zahraničí jde postupně o stále více akcentované výzkumné téma a sílící trend. Hlavním cílem této publikace je tak podat základní vhled do těchto „nových“ výzkumných oblastí a metod, a přispět tak k obohacení dosavadního pedagogického výzkumu.

S ohledem na výše naznačený cíl byly koncipovány i jednotlivé kapitoly této monografie. Po úvodní kapitole tak následuje kapitola 2 mapující vznik a historické kořeny analytiky učení a data miningu ve vzdělávání včetně nástinu dalších relevantních výzkumných oblastí, které se podílely na vzniku a formování analytiky učení a data miningu ve vzdělávání jakožto specifických výzkumných směrů. V další kapitole (kapitola 3) je pak podáno podrobnější vymezení obou oblastí a jsou identifikovány hlavní rozdíly mezi data miningem ve vzdělávání a analytikou učení.

Kapitola 4 podrobněji nastiňuje proces aplikace analytických a data miningových metod a technik v kontextu vzdělávání a představuje základní typy dat, která bývají v těchto oblastech využívána pro účely analýz. Následně navazuje kapitola 5, v níž jsou přiblíženy nejčastěji používané metody a techniky analýzy dat v oblasti analytiky učení a data miningu ve vzdělávání. Kapitola 6 pak poskytuje podrobný přehled témat či směrů výzkumu, jež jsou v oblasti analytiky učení a data miningu ve vzdělávání řešeny a kterým je věnována nejvýraznější výzkumná pozornost.

Následující tři kapitoly (tj. kapitoly 7, 8 a 9) pak představují příklady tří různých zaměřených analýz využívajících různé metody a techniky z oblasti analytiky učení a data miningu ve vzdělávání. První analýza (kapitola 7) směřuje k využití vybraných klasifikačních algoritmů pro účely predikce úspěšnosti studentů v blended learningových kurzech realizovaných v rámci LMS, kdy jsou pro predikaci využity různé charakteristiky týkající se návštěvnosti e-learningových opor kurzů v LMS. Druhá analýza (kapitola 8) přesouvá pozornost k interakci studentů a učitelů

1 Úvod

v online diskuzních fórech, přičemž metodologicky staví na tzv. analýze sociálních sítí. Třetí realizovaná analýza (kapitola 9) pak zkoumá možnosti využití metody dolování procesů pro účely analýzy a detekce různých forem chování studentů v průběhu plnění online testů v LMS.

Závěrečná desátá kapitola shrnuje představenou problematiku jako celek a diskutuje hlavní výsledná zjištění realizovaných analýz. Zaujímá přitom širší pohled na řešenou problematiku a poskytuje zhodnocení toho, nakolik a v jakých ohledech mohou být metody analytiky učení a data miningu ve vzdělávání přínosné pro studium učení a výuky v systémech pro řízení výuky a v online vzdělávacích prostředích obecně.

2 VZNIK A HISTORICKÉ KOŘENY ANALYTIKY A DATA MININGU VE VZDĚLÁVÁNÍ

Zaměříme-li se nejprve na výzkumnou oblast označovanou jako data mining ve vzdělávání (*educational data mining – EDM*), lze říci, že o vzniku této oblasti se nejčastěji hovoří ve spojení s lety 2008 a 2009. Je to dáno tím, že v roce 2008 proběhla první mezinárodní konference zaměřená specificky na data mining ve vzdělávání (*International Conference on Educational Data Mining*) a v roce 2009 pak vyšlo první číslo časopisu *Journal of Educational Data Mining*, zaměřeného výhradně na tuto oblast. Jak ovšem naznačuje jeden z prvních článků publikovaných ve zmiňovaném časopise (Baker & Yacef, 2009), historické kořeny této oblasti jsou výrazně starší. Pokud bychom vyšli ze studie Romera a Ventury (2007), pak můžeme spatřovat první snahy o využití data miningu ve vzdělávání již od roku 1995². V kontextu e-learningu a technologiemi podporovaného učení se však začala data miningu věnovat větší pozornost spíše až na přelomu tisíciletí. To se následně projevilo i v tom, že se téma data miningu ve vzdělávání začalo postupně objevovat v podobě specializovaných workshopů a seminářů na několika tematicky příbuzných konferencích³ (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018). Rostoucí zájem o téma data miningu v kontextu vzdělávání pak v roce 2005 vyústil v publikaci *Data Mining in E-learning* (2005), která se tak stala první knihou věnující se tomuto specifickému tématu. Zároveň bylo v roce 2005 poprvé použito označení

2 Zde však šlo ještě o využití data miningu v kontextu „tradičního“ vzdělávání nikoli v souvislosti s online vzdělávacími systémy typu LMS či VLE, jelikož ty se začaly více rozšiřovat až o několik let později.

3 Šlo např. o konference *International Conference on Artificial Intelligence in Education* (AIED), *International Conference on Intelligent Tutoring Systems* (ITS) či *Annual Conference of the Association for the Advancement of Artificial Intelligence* (AAAI).

2 Vznik a historické kořeny analytiky a data miningu ve vzdělávání

Educational data mining přímo v názvu odborného workshopu zaměřujícího se na tuto oblast (Romero & Ventura, 2013).

Přesto lze právě v souvislosti s lety 2008 a 2009 hovořit o určitém počátku, a to v tom smyslu, že přibližně v této době se začíná o data miningu ve vzdělávání hovořit jako o samostatné a svébytné výzkumné oblasti. Vedle již zmiňované první konference a prvního časopisu věnujících se tomuto tématu lze za další důležitý milník považovat knihu *Handbook of Educational Data Mining* (Romero, Ventura, Pechenizkiy, & Baker, 2010), která se jako první zaměřuje explicitně na oblast EDM. V následujícím roce pak byla založena mezinárodní společnost pro data mining ve vzdělávání (*The International Educational Data Mining Society*). Zároveň se tou dobou začínala vyčleňovat druhá z oblastí, jimž se věnuje tato publikace, totiž analytika učení (*learning analytics – LA*). V souvislosti s oběma zmiňovanými oblastmi se pak přibližně od roku 2014 postupně začíná objevovat hned několik odborných knih, které jsou pro přehlednost shrnuty v následující tabulce (viz tabulka 1).

Tabulka 1: Přehled základních odborných knih věnujících se specificky problematice data miningu ve vzdělávání či analytice učení

Rok	Oblast	Název knihy	Odkaz na zdroj
2005	EDM	<i>Data Mining in E-learning</i>	(Romero & Ventura, 2005)
2010	EDM	<i>Handbook of Educational Data Mining</i>	(Romero, Ventura, Pechenizkiy, & Baker, 2010)
2014	EDM	<i>Educational Data Mining: Applications and Trends</i>	(Peña-Ayala, 2014a)
2014	LA	<i>Learning Analytics: From Research to Practice</i>	(Larusson & White, 2014)
2016	LA	<i>Developing Effective Educational Experiences through Learning Analytics</i>	(Anderson & Gavan, 2016)
2016	EDM i LA	<i>Data Mining and Learning Analytics: Applications in Educational Research</i>	(ElAtia, Ipperciel, & Zaiane, 2016)
2016	EDM	<i>Educational Data Mining with R and Rattle</i>	(Kamath & Kamat, 2016)
2017	LA	<i>Big Data and Learning Analytics in Higher Education: Current Theory and Practice</i>	(Daniel, 2017)
2017	EDM i LA	<i>Big Data in Education: The digital future of learning, policy and practice</i>	(Williamson, 2017)
2017	LA	<i>Handbook of Learning Analytics</i>	(Lang, Siemens, Wise, & Gašević, 2017)
2017	LA	<i>Learning Analytics: Fundamentals, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning</i>	(Peña-Ayala, 2017)
2017	LA	<i>Learning Analytics Explained</i>	(Sclater, 2017)

2 Vznik a historické kořeny analytiky a data miningu ve vzdělávání

Rok	Oblast	Název knihy	Odkaz na zdroj
2017	LA	<i>Learning and Knowledge Analytics in Open Education</i>	(Lai & Lehman, 2017)
2018	LA	<i>Learning Analytics Goes to School: A Collaborative Approach to Improving Education</i>	(Krumm, Means, & Bienkowski, 2018)
2018	LA	<i>Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers</i>	(Lodge, Horvath, & Corrin, 2018)
2018	LA	<i>Learning Analytics in Higher Education: Current Innovations, Future Potential, and Practical Applications</i>	(Lester, Klein, Johri, & Rangwala, 2018)
2019	LA	<i>Emerging Trends in Learning Analytics: Leveraging the Power of Education Data</i>	(Khine, 2019)
2019	EDM	<i>Utilizing Educational Data Mining Techniques for Improved Learning: Emerging Research and Opportunities</i>	(Bhatt, Sajja, & Liyanage, 2019)
2019	LA	<i>Utilizing Learning Analytics to Support Study Success</i>	(Ifenthaler, Mah, & Yau, 2019)
2020	LA	<i>Adoption of Data Analytics in Higher Education Learning and Teaching</i>	(Ifenthaler & Gibson, 2020)
2020	EDM i LA	<i>Big Data on Campus: Data Analytics and Decision Making in Higher Education</i>	(Webber & Zheng, 2020)
2020	LA	<i>Learning Analytics Cookbook: How to Support Learning Processes Through Data Analytics and Visualization</i>	(Jaakonmäki et al., 2020)
2020	LA	<i>Radical Solutions and Learning Analytics: Personalised Learning and Teaching Through Big Data</i>	(Burgos, 2020)
2021	LA	<i>A Beginner's Guide to Learning Analytics</i>	(Srinivasa & Kurni, 2021)
2021	EDM i LA	<i>Advancing the Power of Learning Analytics and Big Data in Education</i>	(Azevedo, Azevedo, Uhoimobhi, & Ossiannilsson, 2021)
2021	LA	<i>Innovative Learning Analytics for Evaluating Instruction: A Big Data Roadmap to Effective Online Learning</i>	(Frick, Myers, Dagli, & Barrett, 2021)
2021	EDM	<i>Modern Approach to Educational Data Mining and Its Applications</i>	(Sweta, 2021)
2021	LA	<i>Visualizations and Dashboards for Learning Analytics</i>	(Sahin & Ifenthaler, 2021)

O analytice učení jako svébytné oblasti zabývající se zkoumáním učení ve virtuálním prostředí se začíná hovořit na přelomu let 2010 a 2011, a to především díky tomu, že v roce 2010 publikoval Siemens na svém blogu zřejmě první definici analytiky učení, která se později stala všeobecně přijímanou v řadě odborných

textů (srov. např. Ferguson, 2012a; Chatti et al., 2012; Siemens, 2013). V roce 2011 pak proběhla první mezinárodní konference zaměřená specificky na analytiku učení (*1st International Conference on Learning Analytics and Knowledge*) a zároveň byla založena společnost pro výzkum v oblasti analytiky učení (*Society for Learning Analytics Research*). Ve stejném roce byla také analytika učení poprvé zmíněna ve výroční zprávě *Horizon Report* (Johnson, Smith, Willis, Levine, & Haywood, 2011), čímž se tato oblast začala dostávat do širšího povědomí odborníků v kontextu e-learningu i vzdělávání obecně.

Analytice učení pak byla věnována výrazná pozornost i v následujících letech. V uvedené výroční zprávě *Horizon Report* se analytika učení nadále objevovala jako jeden z klíčových trendů ve vzdělávání (Johnson, Adams, & Cummins, 2012; Johnson et al., 2013; Johnson, Adams Becker, Estrada, & Freeman, 2014). Od roku 2012 se pak na analytiku učení začaly více zaměřovat i odborné časopisy a postupně vzniklo několik speciálních čísel v prestižních odborných časopisech věnovaných právě tomuto tématu. Jejich základní přehled podává následující tabulka (viz tabulka 2). V roce 2014 pak vychází první číslo nového časopisu zabývajícího se přímo analytikou učení (*Journal of Learning Analytics*) a vychází také první monografie zaměřená specificky na téma analytiky učení (Larsson & White, 2014). V roce 2017 se následně objevila řada dalších publikací zaměřujících se právě na analytiku učení, jak je patrné z tabulky výše (viz tabulka 1).

Tabulka 2: Vybraná speciální čísla odborných časopisů zaměřující se na analytiku učení

Rok	Název časopisu	Číslo	Téma speciálního čísla
2012	<i>Journal of Asynchronous Learning Networks</i>	Vol. 16, No. 3	Learning Analytics
2012	<i>Journal of Educational Technology & Society</i>	Vol. 15, No. 3	Learning and Knowledge Analytics
2013	<i>American Behavioral Scientist</i>	Vol. 57, No. 10	Learning Analytics
2013	<i>Journal of the Learning Sciences</i>	Vol. 22, No. 4	Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning
2015	<i>Journal of Universal Computer Science</i>	Vol. 21, No. 1	Learning Analytics
2015	<i>Computers in Human Behavior</i>	Vol. 47	Learning Analytics, Educational Data Mining and data-driven Educational Decision Making
2016	<i>Journal of Computer Assisted Learning</i>	Vol. 32, No. 3	Learning Analytics in Massively Multi-User Virtual Environments and Courses

Rok	Název časopisu	Číslo	Téma speciálního čísla
2016	<i>Educational Technology Research and Development</i>	Vol. 64, No. 5	Exploring the Relationship of Ethics and Privacy in Learning Analytics and Design: Implications for the Field of Educational Technology
2017	<i>Interactive Learning Environments</i>	Vol. 25, No. 2	Learning Analytics
2017	<i>Zeitschrift für Hochschulentwicklung [Journal for Higher Education Development]</i>	Vol. 12, No. 1	Learning Analytics: Implications for Higher Education
2017	<i>New Directions for Higher Education</i>	Vol. 2017, No. 179	Learning Analytics in Higher Education
2017	<i>Journal of Educational Technology & Society</i>	Vol. 20, No. 2	Learning Analytics in Technology Enhanced Language Learning
2017	<i>ASHE Higher Education Report</i>	Vol. 43, No. 5	Learning Analytics in Higher Education
2017	<i>Journal of Research in Innovative Teaching & Learning</i>	Vol. 10, No. 2	Learning analytics in primary, secondary and higher education
2018	<i>International Journal of Innovation and Learning</i>	Vol. 23, No. 1	User Modelling and Learning Analytics
2018	<i>Journal of Educational Technology & Society</i>	Vol. 21, No. 2	Trends and Research Issues of Learning Analytics and Educational Big Data
2019	<i>Interactive Learning Environments</i>	Vol. 27, No. 5-6	The new potentials for Intelligent Tutoring with learning analytics approaches
2019	<i>User Modeling and User-Adapted Interaction</i>	Vol. 29, No. 4	Learning Analytics and Personalised Support Across Spaces
2020	<i>Australasian Journal of Educational Technology</i>	Vol. 36, No. 6	Learning Analytics: Pathways to Impact

2.1 Historické kořeny data miningu ve vzdělávání

Data mining jako takový, někdy označován také jako dobývání znalostí z databází (*knowledge-discovery in databases – KDD*)⁴, se začal výrazněji rozvíjet začátkem 90. let

⁴ Nutno podotknout, že ačkoli mnozí autoři dnes chápou pojmy data mining a dobývání znalostí z databází jako vesměs synonymní (Gorunescu, 2011), jiní autoři je rozlišují a chápou data mining pouze jako jeden z kroků v rámci procesu dobývání znalostí z databází. Toto rozlišování obou pojmů vychází především z prací Fayyada (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), podle nějž KDD odkazuje k celkovému procesu získávání užitečných znalostí z dat, kdežto pojem data mining odkazuje pouze ke konkrétnímu kroku v tomto procesu, kdy jsou aplikovány specifické algoritmy pro extrakci určitých vzorců z dat. V této publikaci se přikláním k první skupině autorů a považuji oba pojmy za synonymní.

minulého století a postupem času se začal prosazovat v mnoha oblastech od akademického výzkumu přes soukromou sféru až po různé specifické oblasti využití jako např. lékařství, geografie či internetová bezpečnost.

Data mining lze přitom definovat různými více či méně odlišnými způsoby. Například Gorunescu (2011) uvádí následující možné definice:

- Automatické hledání vzorců (*patterns*) v obrovských databázích s využitím výpočetních technik z oblasti statistiky, strojového učení a rozpoznávání vzorců (*pattern recognition*).
- Netriviální extrakce implicitních, dříve neznámých a potenciálně použitelných informací z dat.
- Věda o získávání užitečných informací z velkých datových souborů nebo databází.
- Proces automatického objevování informací, identifikace vzorců a vztahů „skrytých“ v datech.

Vidíme, že všechny uvedené definice mají jedno společné, totiž že v rámci data miningu jde o práci s velkými datovými soubory, v nichž jsou s využitím specifických výpočetních technik automaticky odhalovány doposud neznámé informace, vzorce či vztahy. Data mining přitom čerpá především ze tří hlavních oblastí, které propojuje a jejichž techniky a metody využívá. V první řadě se jedná o „klasickou“ statistiku, ve které data mining navazuje především na oblast exploratorní analýzy dat⁵. Zde data mining čerpá především základní popisné a explorační statistické a vizualizační techniky. Druhou oblastí, na níž data mining staví, je umělá inteligence (*artificial intelligence*) a strojové učení (*machine learning*). Tyto oblasti propůjčují data miningu především různé techniky zpracování dat založené na pokročilém matematickém aparátu, jakož i sadu algoritmů umožňujících počítači „učit se“ z dat. V neposlední řadě pak data mining navazuje na oblast databázových systémů (*database systems*), kde využívá především technické poznatky o způsobech ukládání a uchovávání dat, struktuře dat a databází, data managementu apod.

Oblast data miningu ve vzdělávání, jakožto oblast zaměřující se na využití data miningu specificky ve vzdělávacím kontextu, pak vznikla především na základě snah o využití data miningu pro analýzu tzv. logů⁶ či záznamů o interakci studentů v online vzdělávacích systémech (Baker & Yacef, 2009). Postupně pak byly rozvíjeny stále nové možnosti využití data miningu v oblasti e-learningu a již v roce

5 Exploratorní analýzou dat odkazují k rozlišení, které provedl a propagoval Tukey (1962, 1977), jenž se vymezoval vůči klasické statistice a raději používal pojem analýza dat (*data analysis*), který zároveň vnímal jako širší. V rámci analýzy dat pak rozlišoval mezi tzv. konfirmační analýzou dat (*confirmatory data analysis*) a exploratorní analýzou dat (*exploratory data analysis*).

6 Termín „log“ resp. „logy“ má v češtině dva významy. V prvním významu se tímto slovem označuje soubor či tabulka obsahující záznamy o činnosti (např. o činnosti uživatelů určitého systému), v druhém významu se slovem log odkazuje ke každému jednotlivému záznamu o činnosti, který je uložen ve zmiňovaném souboru.

2007 hovoří Castro a spol. (2007) o řadě směrů využívajících data mining v oblasti e-learningu. Zmiňují např. využití data miningu v oblasti hodnocení výkonu studentů, při evaluaci učebních materiálů a online kurzů či data mining zaměřený na detekci atypického učebního chování studentů. Tyto stále se rozšiřující možnosti využití nakonec vedly k ustanovení data miningu ve vzdělávání jakožto svébytné výzkumné oblasti.

2.2 Historické kořeny analytiky učení

Analytiku jakožto obecný pojem je poměrně obtížné přesně vymezit. S tím, jak se tento pojem v průběhu posledních patnácti až dvaceti let dostal do širšího povědomí, začal totiž být používán v mnoha různých oblastech a kontextech (Van Barneveld, Arnold, & Campbell, 2012). Pojem analytiky tak najdeme nejen ve spojení s učením (*learning analytics*), ale i ve spojení s řadou dalších oblastí (např. *health analytics*, *news analytics*, *cultural analytics* apod.). Časté je také rozlišování analytiky podle účelu prováděných analýz (např. *predictive analytics*, *prescriptive analytics* či *action analytics*), nebo na základě spojení s konkrétní sadou analytických nástrojů (*google analytics*, *twitter analytics*, *facebook analytics* apod.). Někteří autoři se proto přiklánějí k názoru, že kvůli nadužívání tohoto pojmu je jakýkoli pokus o podání jeho dostatečně podrobné, ale netriviální definice předem odsouzen k nezdaru (Cooper, 2012a).

Pro účely této publikace vycházím z obecně pojaté definice autorek Bienkowski, Feng a Means (2012), které vymezují analytiku v širokém smyslu jako využívání technik z počítačových věd, matematiky a statistiky pro extrakci užitečných informací z velkých objemů dat. Všimněme si přitom tří zásadních aspektů:

1. Analytika čerpá ze znalostí a dovedností jak z oblasti informatiky, tak z oblasti matematiky, a primárně statistiky a analýzy dat.
2. Cílem analytiky je získat informace, které pro nás budou užitečné při dalším plánování či jednání. Jde tedy o informace, jež vedou k nějaké další akci (Cooper, 2012a; Picciano, 2012). Van Barneveld, Arnold a Campbell (2012) dokonce vyzdvihují pouze tento aspekt a definují analytiku jednoduše jako daty řízené rozhodování (*data-driven decision making*).
3. Analytika se zabývá velkými objemy dat, kdy nejsou neobvyklé ani datové sady o desítkách či stovkách milionů záznamů. Zde pak lze spatřovat určitý rozdíl oproti „tradiční“ statistice, která příliš často nepracuje s takto velkými daty.

Do širšího povědomí se pojem analytiky dostal především na základě dvou oblastí: webové analytiky (*web analytics*) a podnikové analytiky (*business analytics* či *business intelligence*). Z nich zároveň čerpaly další specifické oblasti analytiky při vlastním vzniku a formování. Není proto divu, že i u analytiky učení uvádějí mnozí

autoři právě tyto dvě oblasti jako jedny z jejích hlavních historických kořenů (např. Buckingham Shum, 2012; Cooper, 2012b, Ferguson, 2012a, 2012b; Macfadyen & Dawson, 2012).

O webové analytice se začíná hovořit poté, co se rozšířilo komerční využití webových stránek, resp. webových technologií obecně. Jakmile začaly nejrůznější firmy postupně využívat webové stránky ke svému podnikání, ukázalo se jako užitečné (tj. výdělečné) sledovat návštěvnost stránek. Firmy proto začaly věnovat pozornost tomu, jakým způsobem se na jejich webových stránkách uživatelé pohybují a jak je využívají. Na základě těchto znalostí pak firmy mohly nechat své stránky upravit a optimalizovat, což jim následně přinášelo větší zisky. Webová analytika se tak ustanovila jako oblast zabývající se sběrem, měřením, analýzou a vyhodnocováním webových dat za účelem porozumění chování uživatele a optimalizace webových stránek (Kaushik, 2011).

V průběhu 90. let a na přelomu tisíciletí (viz Ferguson, 2012a; Macfadyen & Dawson, 2012) se pak technologie webu začala využívat i pro účely výuky, kdy se mezi vzdělávacími institucemi rozšířilo využívání systémů pro řízení výuky (LMS). Po tomto masovém rozšíření webových výukových systémů se tak i v kontextu vzdělávání začalo uvažovat o možnostech analýzy návštěvnosti online vzdělávacích prostředí s cílem jejich následné optimalizace a zlepšení procesu učení se v nich. Jedním z hlavních pramenů analytiky učení je tak myšlenka aplikace webové analytiky ve specifickém kontextu online vzdělávacích systémů založených na webových technologiích.

Stejně výrazným příspěvkem do formování analytiky učení však byla oblast podnikové analytiky, resp. *business intelligence*. V případě podnikové analytiky jde o využití analytiky především v souvislosti se zvyšováním efektivity firemních procesů a podporou strategického managementu a rozhodování. Tento směr využití analytiky je dokonce starší než webová analytika. Rovněž v kontextu vzdělávání se uvažování o analytice ve vzdělávání ubíralo nejprve tímto směrem. Již v roce 2005 Goldstein a Katz (2005) hovoří o akademické analytice jako o využití *business intelligence* v akademickém sektoru za účelem zlepšení organizačních procesů a zvýšení efektivity v rámci univerzity. Podobně Campbell, DeBlois a Oblingerová (2007) či Fritz (2011) přibližují akademickou analytiku jako oblast zabývající se primárně administrativními a provozními záležitostmi vzdělávacích institucí. Zde může jít například o záležitosti týkající se rozpočtu a plánování, řízení lidských zdrojů, financí a fundraisingu nebo správy institucionálního výzkumu.

Lze tedy říci, že dvě výše zmíněné oblasti využití analytiky ve firemní sféře daly vzniknout dvěma různým směrům využití analytiky ve vzdělávání. Na jedné straně stojí analytika učení inspirována primárně webovou analytikou a zaměřující se na samotné procesy učení a vyučování, na druhé straně se pak nachází akademická analytika, která vychází z podnikové analytiky a souvisí především s problematikou řízení vzdělávacích institucí.

2.3 Další oblasti formující analytiku učení a data mining ve vzdělávání

Vedle výše uvedených disciplín či výzkumných oblastí však můžeme identifikovat ještě několik dalších oblastí výzkumu, které se významně podílely na vzniku a formování analytiky učení a data miningu ve vzdělávání. Nejdříve je ale nutno podotknout, že v případě analytiky učení lze za zřejmě nejdůležitější inspirační zdroj považovat právě data mining ve vzdělávání. Jak již bylo zmíněno výše, komunita výzkumníků v rámci data miningu ve vzdělávání se ustanovila o několik let dříve, než vůbec začala být věnována pozornost oblasti označované jako analytika učení. Jakmile se pak začala ustanovovat analytika učení jako svébytná výzkumná oblast, mohla již výrazně čerpat z existující komunity výzkumníků zaměřujících se na data mining ve vzdělávání. Proto je většina níže zmiňovaných oblastí důležitým inspiračním zdrojem nejen pro data mining ve vzdělávání, ale i pro analytiku učení.

Pokud by měl být vybrán jeden hlavní inspirační zdroj pro data mining ve vzdělávání, pak by šlo zřejmě o oblast modelování uživatelů (*user modeling*). Jde o oblast výzkumu v rámci multidisciplinárního oboru studujícího interakci mezi člověkem a počítačem⁷. Cílem výzkumníků v této oblasti je porozumět tomu, jak různí uživatelé interagují s určitou aplikací či uživatelským rozhraním. Následně je možné vytvořit model, na jehož základě může daná aplikace či rozhraní reagovat odlišnými způsoby na různé formy interakce uživatele s aplikací či rozhraním. A to například tím, že se vzhled aplikace přizpůsobí podle potřeb uživatele nebo je uživateli doporučen pro něj relevantní obsah apod. V kontextu data miningu ve vzdělávání se pak často hovoří o modelování studentů (viz např. Romero, Ventura, Pechenizkiy, & Baker, 2010; Peña-Ayala, 2014; Papamitsiou & Economides, 2014), kam může spadat např. modelování chování studentů, modelování jejich emočních či afektivních stavů, modelování motivace či nálady studentů apod. Dále je třeba doplnit, že řada autorů (např. Ferguson, 2012a, 2012b; Chatti et al., 2012; Siemens, 2013) zmiňuje modelování uživatelů zároveň jako jeden z důležitých inspiračních zdrojů i v souvislosti s formováním analytiky učení.

Zásadní důležitost široké oblasti uživatelského modelování pro data mining ve vzdělávání i analytiku učení lze spatřovat především v tom, že určitá forma modelování uživatelů se stala jejich inherentní součástí. Řada projektů a výzkumů v oblasti analytiky a data miningu ve vzdělávání tak zahrnuje alespoň nějaký aspekt modelování uživatelů. Zároveň je vliv modelování uživatelů patrný také v tom, že vedle vlastního modelování uživatelů existuje hned několik specifitější zaměřených výzkumných oblastí, které jsou na modelování uživatelů postaveny a které lze rovněž považovat za inspirační zdroje analytiky a data miningu ve vzdělávání.

7 V angličtině se pro tento obor vžil označení *Human-Computer Interaction* (HCI), které se do češtiny překládá jen velmi zřídka.