

LEARNING ANALYTICS NA ZČU - PRVNÍ PŘÍSTUP

LEARNING ANALYTICS AT UWB – FIRST APPROACH

Petr Grolmus, Lucie Rohlíková

Abstrakt

Tento článek představuje první pohled a interpretaci analýzy učení z dat e-learningového systému (Learning Management System – LMS) provozovaného na Západočeské univerzitě v Plzni (ZČU). Předpokládáme, že existují tři typy granularity dat LMS. Prvním typem je nejvyšší úroveň popisující přístupy a využití LMS jakožto celku. Druhým typem jsou data na úrovni kurzu popisující chování a aktivity všech uživatelů daného kurzu. Posledním typem je uživatelská úroveň dat, která interpretuje aktivity jednotlivých uživatelů kurzu.

Příspěvek diskutuje první dva typy granularity založených na reálných datech univerzitního e-learningového systému Moodle. Inspirovali jsme se mnoha předchozími studii zaměřenými na učební systémy typu LMS, které se často zaměřují na zejména na predikci akademického úspěchu nebo identifikaci studentů ohrožených studijní neúspěšností (např. Smith a kol. 2012; Jayaprakash a kol. 2014; Baker a kol. 2015).

Tato zjištění tvoří základ pro další výzkum, který se zaměřuje na identifikaci chování uživatelů na kurzu a na vyhledání studentů ohrožených studijní neúspěšností.

Klíčová slova: *Learning Innovation, Computers in Human Behavior, Higher Education, Educational Data Mining, Learning Analytics, Student Interaction Analysis, Learning Management System, Prediction Study Success*

Abstract

The focus of this paper is the first look and interpretation of learning analytics data from learning management system (LMS) at the University of West Bohemia in Pilsen (UWB). We claim that there are three types of granularity of LMS data. The first type is top-level which describes approaches and usage of LMS as a whole. The second one is course-level which deals with the behavior and activities of all users as a whole at a specific course. And the last user-type that interprets the activities of users in the course and looking for common patterns of behavior.

This paper presents the first two types of granularity based on real data from the university LMS. We are inspired by many previous studies focusing on learning systems of the LMS, that often pay attention especially to academic success prediction or at-risk student identification (e.g. Smith et al. 2012; Jayaprakash et al. 2014; Baker et al. 2015).

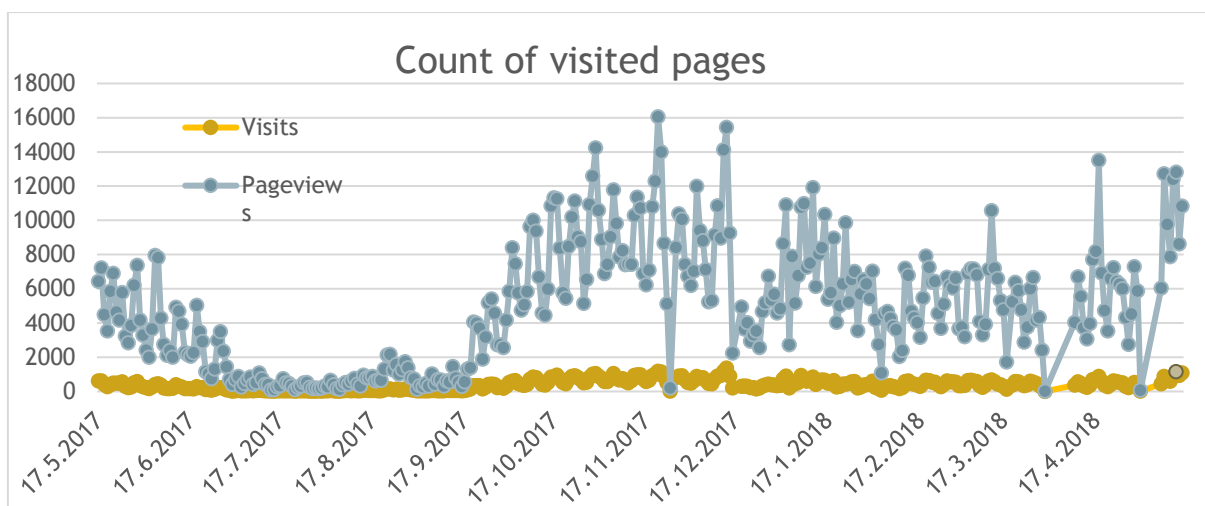
These findings form the basis for further research on identifying user behavior on the course and identifying students at risk of learning failure.

Key words: Learning Innovation, Computers in Human Behavior, Higher Education, Educational Data Mining, Learning Analytics, Student Interaction Analysis, Learning Management System, Prediction Study Success

1 Learning Analytics na ZČU

Na Západočeské univerzitě (ZČU) v Plzni máme za poslední dvě dekády bohaté zkušenosti s provozem různých Learning Management Systems (LMS). Lze říci, že v interním souboji e-learningových systémů zvítězil Moodle, který byl původně “nadšenecky” provozován již od roku 2006 (rok založení prvního kurzu). Roku 2010 se Moodle stal oficiální součástí univerzitního výpočetního prostředí a byl plně provázán jak s mechanismem jednotného přihlašování uživatelů Single Sign-On, tak i se studijní agendou STAG, obsahující veškeré studijní výsledky našich studentů.

Aktuálně (květen 2018) máme za celou dobu provozu v Moodle uloženo cca 1900 kurzů. Z důvodu zachování persistence neprobíhá v tomto LMS promazávání starých kurzů a výsledky nebo odevzdané úkoly studentů jsou tak kdykoliv zpětně k dohledání. Že se Moodle na ZČU stal nedílnou součástí výuky, zcela jednoznačně dokládá i graf návštěvnosti za poslední rok na přiloženém obrázku 1.



Obrázek 1 – Celkový počet návštěv za rok (květen 2017 – květen 2018)

Z obrázku 1 jsou zřetelně patrné také útlumy aktivity uživatelů v období hlavních prázdnin a ve zkuškovém období, resp. těsně před začátkem dalšího semestru. Lokální, pravidelně opakující se poklesy návštěvnosti odpovídají víkendům. Důkazem o neustále se zvyšující oblibě LMS na ZČU, jako podpoře běžné výuky, je i viditelný skokový nárůst návštěvnosti mezi dvěma semestry, končícím letním semestrem akademického roku 2016/2017 (v grafu květen-červen 2017) a zimním semestrem akademického roku 2017/2018 (od září do prosince).

Vysoká míra návštěvnosti e-learningových kurzů generuje velké množství dat - logů, jak v LMS samotném, tak i ve službě webového serveru, který zajišťuje komunikaci s koncovými uživateli.

Dlouhou dobu tyto informace o chování uživatelů během procesu učení a konání testů zůstávaly nevyužité a v podstatě jen zabíraly cenné zdroje serveru. To se změnilo až v posledních letech s rozvojem výzkumných oblastí Educational Data Mining

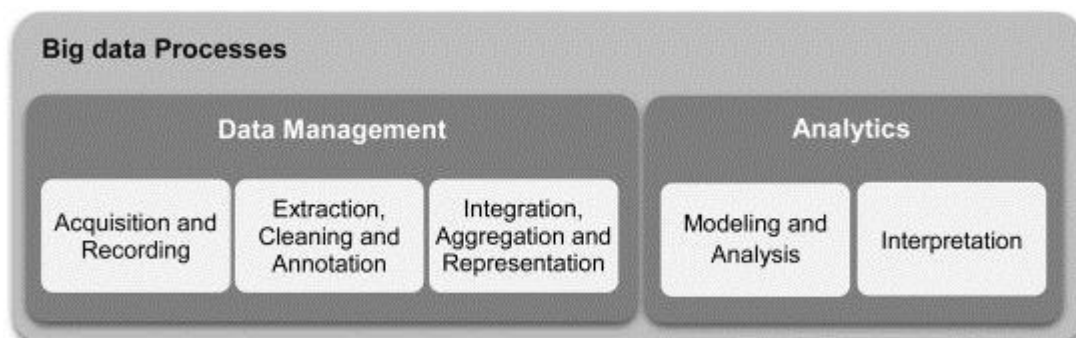
(EDM) a o něco později i Learning Analytics (LA). Oblast EDM byla na vzestupu v letech 2008-2009 (Romero a Ventura 2013), přestože její počátky lze datovat již do roku 2005 (Romero a Ventura 2007). Mladší oblast LA lze datovat do let 2010-2011 (Ferguson 2012; Juhaňák a Zounek 2016).

Obě výzkumné oblasti EDM a LA se v mnoha ohledech liší, co však mají společné je, že se snaží vyzískat maximum informací a souvislostí z dat generovaných a uložených v rámci samotného LMS. Za tímto účelem používají různé analytické a data mining metody a postupy, které umožňují získat důležité informace a poznatky o tom, jak se studenti v těchto systémech chovají, učí, plní úkoly a skládají testy (Juhaňák et al. 2017).

(Siemens et al. 2011) definuje oblast LA jako:

“Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. Learning analytics are largely concerned with improving learner success.”

Stejný zdroj uvádí, že LA je speciální reprezentací aplikace “big data” a analýzy v oblasti vzdělávání. Činnosti spojované s Big data bývají obecně děleny na dvě základní části: data management a samotnou analýzu (Gandomi a Haider 2015). Data management je dále členěn na získávání a zaznamenávání dat, extrakci a čištění dat a jejich následnou integraci, agregaci a reprezentaci. S takto získanými daty se pak provádí modelování, analýza a následná interpretace výsledků – viz obrázek 2.



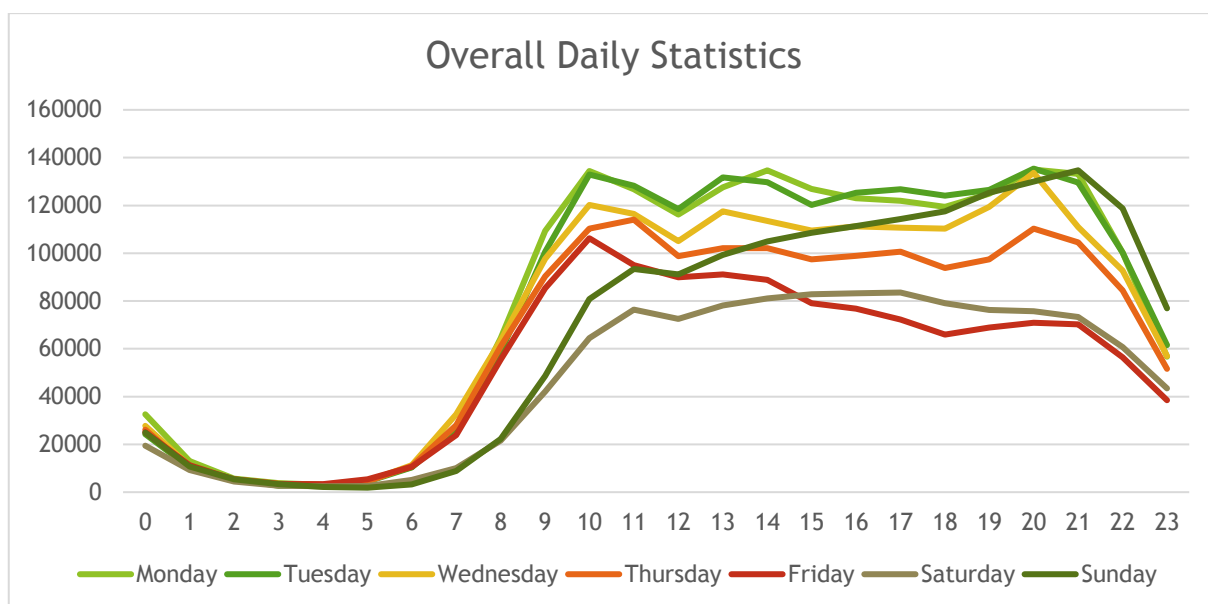
Obrázek 2 – Zpracování Big Data (zdroj: Gandomi a Haider 2015)

V rámci zvyšujících se potřeb na kvalitu a evaluaci výuky na ZČU jsme se primárně zaměřili na oblast e-learningu, právě z dříve popsanych důvodů -- v logách LMS a webserveru již vlastníme obrovské množství dat, které lze snadno použít pro výzkum na poli EDM, resp. LA. Tento článek popisuje první zkušenosti ZČU v oblasti LA a jako takový může být východiskem i pro ostatní zájemce a organizace.

2 Celkový pohled na data

V našem LMS máme nastřádaná data za několik let zpětně a tato data lze v podstatě ihned využít k vytvoření celkové statistiky. Prvním vhladem do těchto dat může být výše uvedený obrázek 1 vyjadřující míru návštěvnosti za poslední kalendářní rok.

Další zajímavou statistikou z celkových dat je úhrnná návštěvnost v jednotlivých dnech v týdnu – viz Obrázek 3 – Celkové denní statistiky.



Obrázek 3 – Celkové denní statistiky

Již v tomto hrubém náhledu na celková data lze vyvodit několik zajímavých faktů. Předně, průběh grafu v jednotlivých dnech týdne je víceméně podobný, liší se pouze v absolutních počtech přístupů. Minimální, téměř nulový provoz v nočních hodinách pro nás nebyl příliš překvapivý a pouze potvrdil správnost rozhodnutí provádět pravidelné zálohy LMS systému ve 4h ráno. Žádným překvapením nebyl cca o 2 hodiny zpožděný nárůst provozu během víkendu.

Více překvapivé bylo u celkových studijních návyků studentů zjištění, že mezi 10. a 21. hodinou všech dní je provoz vyrovnaně stálý. Drobný pokles mezi 11. a 12. hodinou u všech dní odpovídá času oběda. Dalším překvapením patrným z grafu je celodenní pozvolný nárůst nedělního provozu, který kulminuje kolem 21. hodiny a vyrovnává se provozním špičkám v pracovních dnech. Nedělní provoz mezi 21. a 23. hodinou dokonce v celkových hodnotách převyšuje provoz v ostatních dnech.

3 Statistiky kurz

V případě ZČU jde o úplně první počín v oblasti LA z dat našeho dlouhodobě provozovaného LMS Moodle. Prvním krokem je tedy nutnost stanovení rámcové koncepce a prvotní strategie. Náš první krok jsme po zralé úvaze minimalisticky nadefinovali následovně:

- zvolíme pro zpracování pouze jeden kurz,
- máme kompletní logy akcí studentů na kurzu,
- kurz musí mít větší počet studentů, ideálně více než 100 z důvodu statisticky významného vzorku,
- kurz musí mít odevzdávané úkoly hodnocené učitelem (zpětná vazba),
- kurz se musí pravidelně opakovat každý akademický rok, abychom mohli srovnávat výsledky a případně predikovat studenty ohrožené studijní neúspěšností na základě dřívějších dat,

- známe výsledek studia předmětu -- tj. buď zápočet s hodnocením splnil/nesplnil nebo známka z předmětu 1 (nejlepší), 2, 3 nebo 4 (nesplnil předmět).

Z užšího výběru jsme zvolili kurz "Informační technologie ve výuce" zakončený zápočtem, vyučovaný Katedrou výpočetní a didaktické techniky při Fakultě pedagogické ZČU. V akademickém roce 2016/2017 mělo tento předmět zapsáno podle studijní agendy celkem 293 studentů. Z toho bylo 51 studentů kombinované formy studia a 242 studentů prezenční formy studia.

Jelikož lze předpokládat, že chování studentů prezenční a kombinované formy studia bude odlišné a navíc, studenti kombinované formy studia nevyužívali vybraný kurz, proto další zpracování se týká výhradně 242 studentů prezenční formy studia.

Zajímavým a nečekaným bonusem u zvoleného kurzu byl fakt, že studenti nebyli zapisováni do kurzu přes funkční webservice vazby Moodle-STAG, která běžně udělí přístup všem studentům, kteří dle studijní agendy mají předmět zapsaný. Učitel na prvním prezenčním setkání se studenty předmětu zveřejnil přístupové heslo, kterým se studenti do kurzu sami zapisovali. Získali jsme tak další možný ukazatel pro chování studentů.

Do zvoleného kurzu se zapsalo jen 227 studentů. Rozdíl mezi 242 studenty zapsanými ve studijní agendě STAG a 227 zapsanými v kurzu jsou studenti, kteří přestali studovat z různých důvodů. Úspěšně předmět zakončilo zápočtem 199 studentů. Bohužel, 28 studentů v tomto předmětu neuspělo.

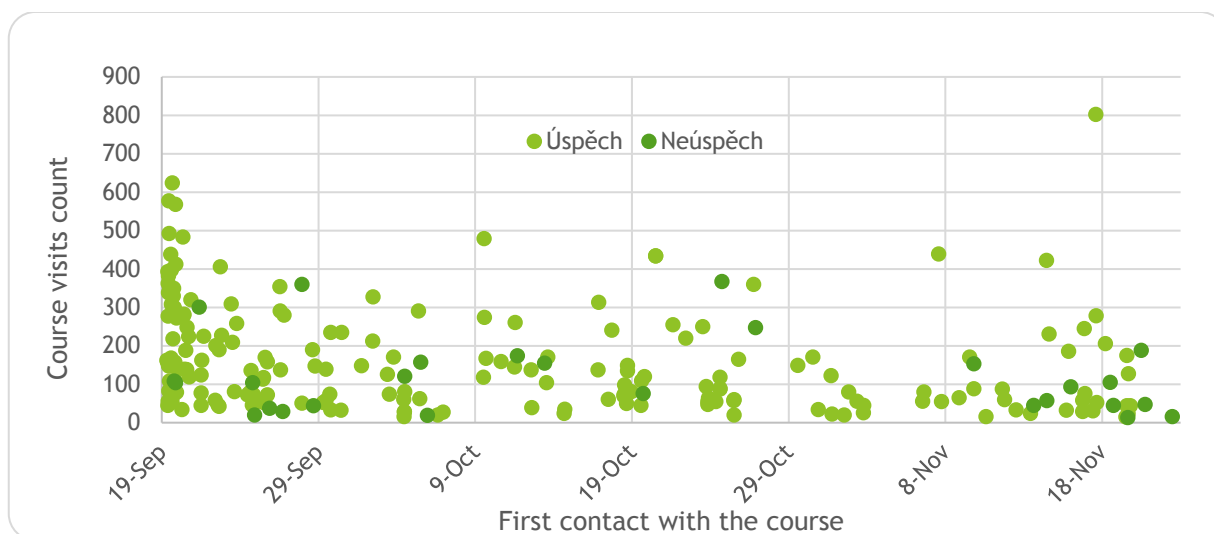
Dalším sledovaným ukazatelem byl počet návštěv sledovaného kurzu v systému Moodle, dle jednotlivých studentů zapsaných v kurzu. Celkem bylo zaregistrováno 35359 návštěv tohoto kurzu studenty. Z tabulky 1 jednoznačně vyplývá, že úspěšní studenti se vraceli do kurzu častěji, než studenti neúspěšní.

Tabulka 1 – Návštěvnost kurzu

| | 199 úspěšných studentů | 28 neúspěšných studentů |
|----------------------|-------------------------------|--------------------------------|
| Počet návštěv | 32174 | 3185 |
| Průměr | 161,68 | 113,75 |

Jelikož jsme znali u jednotlivých studentů skutečnost, zda v kurzu uspěli, jejich počet návštěv a i datum prvního kontaktu s kurzem, zkusili jsme tyto údaje zobrazit v grafu a sledovat, zda budou patrné nějaké základní charakteristiky – viz obrázek 4 - Studenti v kurzu. Předpoklad, že úspěšní studenti začnou kurz studovat dříve a neúspěšní později se nepodařilo jednoznačně prokázat. Studenti svůj první kontakt s kurzem rozložili do celého semestru. Nicméně v úplném počátku je patrná převaha úspěšných studentů, kteří zápočet získali. Taktéž na úplném konci lze sledovat poměrně častý výskyt neúspěšných studentů.

Za zmínku rozhodně stojí student (či studentka), který se do kurzu poprvé přihlásil 17. listopadu 2016 – těsně před zápočtovým testem 21. listopadu 2016, ale provedl celkem největší počet návštěv kurzu – 802 zobrazení kurzu. V grafu jej naleznete jako bod vpravo, zcela nahoře.



Obrázek 4 – Studenti v kurzu

4 Závěr a budoucí výzkum

V naší organizaci dlouhou dobu panuje snaha o zajištění kvalitní výuky s cílem připravit studenty do nadcházejícího profesního života v maximální možné míře. Každoročně zpracovává Komise pro kvalitu na ZČU zpětnou vazbu o vyučovaných předmětech přímo od studentů. Ti anonymně v přehledných formulářích “známkuji” např. srozumitelnost výkladu, přínos předmětu pro zvolený obor, jeho využitelnost i zkušenosti získané při praktických cvičeních. Vyplnění dotazníku není však povinnost a tak zpětná vazba závisí na dobré vůli hodnotitelů, kteří navíc mohou odpovídat náhodně nebo dokonce s cílem poškodit výsledky hodnocení konkrétního předmětu. Dle posledních statistik Komise pro kvalitu poskytne zpětnou vazbu jen 15-20% studentů ze všech devíti fakult ZČU.

V podobě LA nyní dostáváme do rukou další využitelnou metriku hodnocení odvozenou přímo z chování a výsledků studentů daného předmětu. Metriku získanou nezávisle na vůli studentů z dat uložených v LMS. Metriku, kterou nelze snadno znehodnotit zavádějícími odpověďmi v dotazníku.

Prozatím jsme se zabývali pouze daty získanými z jednoho předmětu. Díky vhodně zvolené vstupní podmínce -- předmět musí být vyučován pravidelně i v následujících akademických rocích -- snadno v budoucnu získáme srovnání stávajících výsledků s výsledky novými. Navíc, díky znalosti chování neúspěšných studentů z předchozích let, dokážeme s jistou mírou přesnosti vytipovat studenty ohrožené studijní neúspěšností v právě probíhající výuce stejného kurzu. Učitel předmětu, resp. tutor kurzu, se může díky této znalosti více soustředit na skupinu těchto “ohrožených” studentů a pokusit se je více motivovat k úspěšnému zvládnutí předmětu.

Zajímavým výsledkem by mohlo být i srovnání u shodného předmětu studentů prezenčního studia se studenty kombinovaného studia. Vycházíme z předpokladu, že styl učení obou skupin se bude do velké míry lišit. Skupina statistiky věkově mladších prezenčních studentů se de facto věnuje jen studiu, zatímco druhá skupina studentů kombinovaného studia již vykonává vlastní profesi. Tito navíc často mají již založené rodiny a čas ke studiu ukrajují z velmi cenného volného času anebo ze spánku.

Na výše uvedeném příkladu jsme si vyzkoušeli metody získávání dat relevantních k jednomu kurzu v LMS. U toho prvního pokusu nám dlouho trvalo i rozklíčování některých vazeb v rámci databáze LMS, která má více jak 300 různě propojených entit. Získané zkušenosti nyní půjde výrazně rychleji využít pro větší skupinu kurzů splňujících výše definovaná kritéria.

Použitá literatura

1. Baker, R. S.J.D., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education e-Learning Courses. In *Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain, Jun 26-29, 2015*.
2. Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317.
3. Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, volume 35, issue 2, p. 137-144.
4. Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6-47.
5. Juhaňák, L., Zounek, J. (2016). Analytika učení: nový přístup ke zkoumání učení (nejen) ve virtuálním prostředí. *Pedagogická orientace*, 26(3), 560–583.
6. Juhaňák, L., Zounek, J., Rohlíková, L. (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. In *Computers in Human Behavior*, Elsevier, currently in press.
7. Romero, C., Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, 3, 12–27.
8. Romero, C., Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 1UWB35–146.
9. Siemens, G. et al. (2011). Open Learning Analytics: an integrated & modularized platform, <http://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf>.
10. Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive Modeling to Forecast Student Outcomes and Drive Effective Interventions in Online Community College Courses. *Journal of Asynchronous Learning Networks*, 16(3), 51-61.

Kontaktní údaje

Ing. Petr Grolmus

Západočeská univerzita v Plzni, Centrum informatizace a výpočetní techniky, Univerzitní 20,
301 00 Plzeň

Tel: 377 632 851

e-mail: indy@civ.zcu.cz

PhDr. Lucie Rohlíková, Ph.D.

Západočeská univerzita v Plzni, Fakulta pedagogická
Klatovská tř. 51, 306 19 Plzeň

E-mail: lrohlik@kvd.zcu.cz