

Juhaňák, Libor

Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech

In: Juhaňák, Libor. *Analytika učení a data mining ve vzdělávání v kontextu systémů pro řízení výuky*. Vydání první Brno: Masarykova univerzita, 2023, pp. 89-124

ISBN 978-80-280-0184-1; ISBN 978-80-280-0185-8 (online ; pdf)

Stable URL (handle): <https://hdl.handle.net/11222.digilib/digilib.77695>

Access Date: 03. 12. 2024

Version: 20230228

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

7 ANALÝZA 1: PREDIKCE ÚSPĚŠNOSTI STUDENTŮ V BLENDED LEARNINGOVÝCH KURZECH

Jak již bylo naznačeno v kapitole 6.2 Predikce úspěšnosti a identifikace rizikových studentů, problematika predikce úspěšnosti studentů je v data miningu ve vzdělávání a v analytice učení jednou z hlavních oblastí zájmu výzkumníků. Podobně lze toto téma považovat za poměrně tradiční i v kontextu e-learningu jako takového, kde byly už od samotných počátků využívání moderních technologií ve vzdělávání zároveň kladeny otázky na to, zda a případně nakolik využívání těchto technologií přispívá k úspěšnosti studentů (srov. Dawson, Gašević, Siemens, & Joksimović, 2014; Romero & Ventura, 2013). Data mining ve vzdělávání a analytika učení jakožto výzkumné oblasti na toto obecnější téma pak více méně navazují, ovšem s jedním zásadním rozdílem. Kladou větší důraz na samotnou predikci (ideálně pak na dostatečně včasnou predikci), pro kterou se snaží využít především data o chování studentů v online vzdělávacích prostředích (Gašević, Dawson, Rogers, & Gasevic, 2016).

S důrazem na analýzu chování studentů v online vzdělávacích prostředích pak souvisí i to, že je v kontextu systémů pro řízení výuky (LMS) směřována hlavní pozornost na online learning a online kurzy. Naopak spíše ojediněle jsou realizovány výzkumy zaměřující se na predikování úspěšnosti studentů v kontextu blended learningu, resp. kurzů, které neprobíhají výhradně v online prostředí, nýbrž kombinují vzdělávání v online prostředí se vzděláváním skrze tradiční (offline) výuku (např. AlJarrah, Thomas, & Shehab, 2018; Gitinabard, Xu, Heckman, Barnes, & Lynch, 2019; González Nespereira, Dai, Díaz Redondo, & Fernández Vilas, 2014; González Nespereira, Fernández Vilas, & Díaz Redondo, 2015). Zaměření pozornosti právě na online kurzy přitom není nijak zvlášť překvapivé. V rámci online kurzů totiž veškerá výuková aktivita probíhá v online vzdělávacím prostředí, a tudíž data o chování studentů v tomto prostředí zachycují mnohem

komplexnější obraz o průběhu výuky, resp. vzdělávání jako celku. Naopak v případě blended learningu probíhá v online prostředí pouze větší či menší část výuky, a tudíž je možné skrze analýzu dat o chování studentů v online prostředí získat vždy pouze dílčí obraz o výuce jako celku. Na druhou stranu však blended learning sehrává v současnosti poměrně velkou roli. Například v kontextu univerzit a vysokých škol, kde se dnes problematika data miningu a analytiky učení rozvíjí asi nejvíce, a zvláště pak v souvislosti s využíváním LMS v rámci univerzitní výuky, je blended learning zastoupen výrazně více než online learning (Švaříček & Zounek, 2008).

V souvislosti s predikcí úspěšnosti studentů se v analytice učení a data miningu ve vzdělávání využívá řada různých analytických metod (viz kapitolu 5.1 Prediktivní metody). Hlavní snahy přitom směřují především k využití pokročilých a v tradičním pedagogickém výzkumu ojediněle využívaných prediktivních technik, jako jsou rozhodovací stromy, náhodný les, podpůrné vektory, umělé neuronové sítě apod. Zatím však stále zůstává otázkou, zda je využití těchto „nových“ analytických metod skutečně výrazným přínosem v oblasti predikce úspěšnosti studentů. Případně pak, které z široké škály dostupných prediktivních metod a technik dávají v tomto kontextu nejlepší výsledky. Prozatím téměř neřešenou otázkou je pak to, nakolik jsou tyto metody využitelné přímo v kontextu blended learningu.

V návaznosti na obecnou otázku po možnostech predikce úspěšnosti studentů, a to jak v kontextu online learningu, tak i v souvislosti s blended learningem, se pak nabízí několik doplňujících či rozšiřujících otázek. V rámci analytiky učení a data miningu ve vzdělávání se jeví jako důležité především následující dvě otázky. První se ptá na konkrétní indikátory či ukazatele, které jsou využitelné jako prediktory úspěšnosti studentů. V dosavadním výzkumu se přitom jako prediktory úspěšnosti využívají především tyto ukazatele: celkový čas strávený v LMS, míra aktivity v LMS, frekvence návštěv v LMS apod. (srov. AlJarrah, Thomas, & Shehab, 2018; Jo, Kim, & Yoon, 2015; González Nespereira, Dai, Díaz Redondo, & Fernández Vilas, 2014; You, 2016). Druhá otázka, která je akcentována především v posledních letech, se týká zobecnitelnosti a přenositelnosti dosavadních výsledků napříč různými kurzy či dokonce napříč různými online vzdělávacími systémy. Konkrétně jde především o to, do jaké míry a s jakou prediktivní silou jsou výše uvedené prediktivní metody a zmíněné prediktory úspěšnosti aplikovatelné nejen v rámci jednoho konkrétního kurzu, ale také obecně napříč různými kurzy (srov. Gašević, Dawson, Rogers, & Gasevic, 2016; Gitinabard, Xu, Heckman, Barnes, & Lynch, 2019; Huang, Lu, Huang, Yin, & Yang, 2020).

Tato analýza proto věnuje pozornost výše naznačeným otázkám a problémům. Oproti většině dosavadního výzkumu v této oblasti se však zaměřuje na blended learning a úspěšnost studentů v blended learningových kurzech. V kontextu blended learningu pak srovnává několik prediktivních technik a jejich predikční schopnosti. Pro predikci úspěšnosti studentů jsou přitom využity různé temporální

indikátory návštěvnosti e-learningových opor blended learningových kurzů a je provedeno posouzení přínosu jednotlivých prediktorů. V neposlední řadě pak tato analýza usiluje o zhodnocení toho, nakolik jsou srovnávané prediktivní techniky a použité prediktory generalizovatelné a použitelné při predikci úspěšnosti studentů napříč různými blended learningovými kurzy.

7.1 Přehled dosavadních výzkumů

Romero a Ventura (2019) zmiňují, že predikce výkonu či úspěšnosti studentů je jedním z nejstarších a nejvíce studovaných problémů v kontextu analytiky učení a data miningu ve vzdělávání. Napříč realizovanými výzkumy jsou pak s větší či menší úspěšností využívány různé regresní či klasifikační prediktivní techniky, přičemž často dochází i k praktickému využití získaných prediktivních modelů, a to především v rámci systémů včasného varování (EWS), jejichž cílem je co možná nejvčasnější identifikace studentů, jež mohou mít problém s dokončením kurzu (viz Romero & Ventura, 2013). Řada výzkumů v této oblasti se zaměřuje na využití dat ze systémů typu LMS, kterým věnuje pozornost i tato publikace, případně věnují pozornost datům ze studentských informačních systémů (SIS) či kombinaci dat z LMS i SIS (srov. Gašević, Dawson, Rogers, & Gasevic, 2016). Výrazným impulzem pro výzkum v této oblasti bylo rovněž značné rozšíření kurzů typu MOOC, ke kterému došlo především v průběhu roku 2012 a v letech následujících a které vedlo k dalšímu prohloubení zájmu o problematiku predikce úspěšnosti studentů, zvláště pak v kurzech tohoto typu. Jak ale upozorňují Gitiňabardová et al. (2019), přestože existuje relativně extenzivní výzkum predikce úspěšnosti v MOOC či obecně online kurzech, co se týče blended learningových kurzů, jsme stále relativně na počátku. Predikce úspěšnosti studentů v blended learningových kurzech je tak zatím velmi omezená a jen obtížně se daří dosáhnout dostatečně včasné a zároveň přesné predikce. Není zatím zřejmé, jaké indikátory jsou nejlepšími prediktory úspěšnosti, a dosavadní modely často nebývají použitelné u jiných kurzů, než na kterých byly vytvořeny.

V této přehledové části nejprve mapují dosavadní snahy výzkumníků o predikci úspěšnosti studentů v blended learningových kurzech v rámci systémů pro řízení výuky (LMS). Následně se zaměřují na temporální charakteristiky návštěvnosti jakožto jedny z nejvíce využívaných prediktorů úspěšnosti v online kurzech. První výzkumy zaměřující se na tento typ prediktorů v kontextu online learningu naznačují potenciál k využití těchto prediktorů i při predikci úspěšnosti v blended learningových kurzech (viz např. Jo, Park, Kim, & Song, 2014; Yu & Jo, 2014). Závěrečnou podkapitolu věnují problematice zobecnitelnosti či přenositelnosti prediktivních modelů, což lze považovat za jedno z neaktuálnějších témat současného výzkumu v této oblasti (srov. Romero & Ventura, 2019).

7.1.1 Predikce úspěšnosti studentů v kontextu blended learningu a LMS

Zaměříme-li se (s ohledem na omezený prostor) pouze na zdroje z posledních deseti let, lze mezi prvními zmínit exploratorní studii Macfadyena a Dawsona (2010), kteří zkoumali možnosti využití dat z LMS (konkrétně systém *Blackboard*) pro účely vývoje systému včasného varování (EWS). Analyzovaná data pocházela z pěti různých kurzů, ovšem ve všech případech šlo o online kurzy. Romero a kol. (viz Romero, Espejo, Zafra, Romero, & Ventura, 2013; Romero, López, Luna, & Ventura, 2013) se již zaměřují přímo na LMS Moodle a současně provádějí srovnání různých klasifikačních metod a algoritmů. Bohužel však explicitně nespécifikují, zda data pocházejí z online či blended learningových kurzů, nýbrž hovoří spíše technicistně o „Moodle kurzech“. Naopak Agudo-Peregrina et al. (2014) se již zaměřují explicitně i na blended learningové kurzy, když vedle dat ze šesti online kurzů analyzují rovněž data ze dvou blended learningových kurzů. Analyzovaná data pocházejí opět z LMS Moodle, ovšem autoři již záměrně usilují o to, aby byl jejich postup přenositelný, resp. nezávislý na konkrétním systému, a proto se zaměřují spíše na obecné typy interakce v LMS⁴⁹. Konkrétně navazují na Moorovu (1989) typologii interakcí a rozlišují tři základní typy interakcí v LMS: interakce mezi studenty navzájem, interakce mezi studentem a učitelem a interakce mezi studentem a obsahem. Výsledky však naznačují, že ačkoli má tento přístup určitý potenciál v kontextu online kurzů, v případě blended learningových kurzů se zdá být neplodný. Autoři nenacházejí žádný vztah mezi četností jednotlivých typů interakcí v LMS a výsledky studentů v blended learningových kurzech. Zároveň sami upozorňují na možnost, že je využití tohoto typu dat pro účely predikce úspěšnosti použitelné pouze u online kurzů, resp. že v případě blended learningových kurzů je zřejmě nutné využít pokročilejší metody a nástroje analýzy či jiné přístupy k řešení predikce.

Přímo na blended learningové kurzy se pak v souvislosti s predikcí úspěšnosti zaměřují např. Jo, Park, Kim a Song (2014) či Yu a Jo (2014). V obou případech je věnována pozornost především proměnným, které se týkají návštěvnosti kurzů a míry interakce studentů v nich. I přes určitou slibnost výsledků však obě studie trpí velmi malým množstvím dat, a to jak v počtu studentů, tak i v počtu sledovaných kurzů. Na včasnou predikci úspěšnosti studentů v blended learningových kurzech v rámci LMS se dále zaměřili Sheshadri et al. (2018), kteří se opět věnují především temporálním charakteristikám návštěvnosti jakožto potenciálním prediktorům. Analýzy jsou však prováděny na studentech pouze tří různých kurzů (resp. dvou běžích téhož kurzu a jednoho jiného kurzu). Lu et al. (2018) pak věnují pozornost pouze jedinému blended learningovému kurzu, mezi prediktory však

49 Autoři v rámci studie používají primárně pojem *virtuální vzdělávací prostředí* (VLE), ovšem využívají ho ve stejném smyslu, v jakém je používán pojem LMS v této publikaci.

používají i proměnné týkající se výuky mimo LMS (např. výsledky průběžných „papírových“ testů či účast na doplňující výuce). Poměrně nepřekvapivě pak autoři uzavírají, že využití jak proměnných týkajících se online prostředí, tak proměnných týkajících se tradičního (offline) prostředí přináší lepší výsledky. Určitý příspěvek k predikci úspěšnosti studentů v blended learningových kurzech pak nacházíme i v českém pedagogickém prostředí, kde se této problematice věnovala Vaculíková (2018) v souvislosti se seberegulovaným učením (SRL). Bohužel se však studie zaměřuje pouze na 54 studentů jednoho konkrétního kurzu, což zjištěné výsledky výrazně limituje, co se týče jejich zobecnění.

Z hlediska nejčastěji využívaných prediktorů úspěšnosti můžeme učinit několik základních rozlišení. Jak již bylo naznačeno výše, vedle dat z LMS jsou v některých studiích (např. Estacio & Raga, 2017; Gašević, Dawson, Rogers, & Gasevic, 2016) využívána rovněž data o studentech pocházející ze studentských informačních systémů (SIS) či jiných systémů spíše administrativního než výukového charakteru. Typicky přitom jde o data demografického či popisného charakteru (pohlaví, občanství, bydliště, semestr studia apod.) nebo o data o předchozí studijní úspěšnosti (výsledky na střední škole, výsledky přijímacího řízení, výsledky v dosavadních kurzech na vysoké škole apod.). Existují rovněž pokusy o alespoň částečné zachycení offline stránky blended learningu, kterou pochopitelně systémy typu LMS nezachytí. Příklad takové studie nabízejí již zmiňovaní Lu et al. (2018). V principu si ale lze představit i další proměnné, které by bylo možné v této souvislosti využít (např. docházka na prezenční výuku, aktivita na prezenční výuce, četnost osobních konzultací s vyučujícím apod.). Co se týče dat pocházejících z LMS, můžeme v dosavadních studiích odlišit tři základní skupiny nejčastěji využívaných prediktorů.

První skupinu tvoří obecné temporální charakteristiky návštěvnosti kurzů, kterým se podrobněji věnuje následující část. Jde o různé míry toho, jak moc je kurz navštěvován, přičemž obecnost v tomto kontextu znamená, že jde o návštěvnost kurzu jako celku a nerozlišuje se tak např. aktivita studentů v obsahově různých částech kurzu. Temporalita či časovost pak odkazuje k tomu, že jde o metriky návštěvnosti, které se v nějakém ohledu vážou na čas, např. kdy k návštěvám dochází, jak pravidelně apod. (srov. AlJarrah, Thomas, & Shehab, 2018; Jo, Park, Kim, & Song, 2014; Jo, Kim, & Yoon, 2015; You, 2016; Yu & Jo, 2014).

Druhou skupinu tvoří obvykle poměrně jednoduché četnostní metriky aktivity studentů ve specifických oblastech kurzu. Typicky jde o prostý součet záznamů o aktivitě studenta (tj. logy) v rámci konkrétního nástroje v LMS, jako jsou například nástroje pro tvorbu testů, nástroje pro prezentaci studijních materiálů, nástroje pro odevzdávání úkolů apod. (srov. Estacio & Raga, 2017; Gašević, Dawson, Rogers, & Gasevic, 2016; Lu, Huang, Lin, Ogata, & Yang, 2018; Sheshadri, Giti-nabard, Lynch, Barnes, & Heckman, 2018; Vaculíková, 2018). Určitou nevýhodou prediktorů tohoto typu je však přímá vazba na konkrétní LMS (např. Moodle), a tudíž horší možnosti přenositelnosti takových řešení mezi různými LMS.

Třetí skupinu tvoří různé metriky spojené s aktivitou studentů v diskuzních fórech. Mezi používanými metrikami pak najdeme jednak jednoduché četnostní charakteristiky jako počet vytvořených diskuzních příspěvků, počet odpovědí na příspěvky ostatních studentů apod. (viz např. Macfadyen & Dawson, 2010; Romero, López, Luna, & Ventura, 2013; Yu & Jo, 2014), jednak i pokročilejší metriky vycházející z analýzy sociálních sítí (SNA) jako např. různé míry centrality jednotlivých aktérů v síti⁵⁰ (Gitinabard, Xu, Heckman, Barnes, & Lynch, 2019; Joksimović et al., 2016).

Nutno však dodat, že existují i takové studie, které využívají specifické typy prediktorů nespádajících do žádné z výše nastíněných skupin. Ty jsou však z hlediska četnosti zastoupeny výrazně méně. Např. zmiňovaná studie Vaculíkové (2018) využívá rovněž data z dotazníku na seberegulované učení. Jiným příkladem jsou Nakayama, Mutsuura a Yamamoto (2017), kteří využívají specifické míry toho, nakolik si studenti v průběhu kurzu píší průběžné studijní poznámky.

Co se pak týče užívaných prediktivních metod, asi nejčastěji využívané zatím zůstávají lineární a logistická regrese, a to v závislosti na typu predikované proměnné (viz např. Gašević, Dawson, Rogers, & Gasevic, 2016; Jo, Park, Kim, & Song, 2014; Macfadyen & Dawson, 2010; Vaculíková, 2018). Řada autorů však využívá i jiné typy prediktivních metod a klasifikačních algoritmů, případně v rámci jedné studie využívá několik algoritmů a nabízí jejich srovnání. Tak např. Sheshadri et al. (2018) srovnávají výsledky logistické regrese s algoritmem *k*-nejbližších sousedů (*k-nearest neighbors* – *k-NN*) a s algoritmem z rodiny rozhodovacích stromů. Někteří autoři pak přistupují ke srovnávání algoritmů velmi zevrubně a srovnávají až 15 (Zafra, Romero, & Ventura, 2011), 20 (Romero, López, Luna, & Ventura, 2013) či dokonce více jak 20 různých algoritmů (Romero, Espejo, Zafra, Romero, & Ventura, 2013). Z hlediska typu přitom jde nejčastěji o algoritmy založené na rozhodovacích stromech, algoritmy založené na pravidlech, algoritmy z rodiny bayesovských klasifikátorů či algoritmy založené na umělých neuronových sítích. Srovnávání jednotlivých prediktivních postupů však alespoň prozatím (a v kontextu blended learningu obzvláště) nedává jednoznačnou odpověď ohledně jejich prediktivních schopností. Do značné míry je to však zřejmě způsobeno tím, že srovnávání jednotlivých algoritmů zatím probíhá na datech pouze z jednoho kurzu, případně na poměrně malém množství kurzů (v řádu jednotek).

50 Co se týče měr centrality aktérů v síti, jde např. o stupeň (*degree*), blízkost (*closeness*), mezilehlost (*betweenness*) apod.

7.1.2 Temporální charakteristiky návštěvnosti jako prediktory úspěšnosti

Jak bylo řečeno výše, jednou z nejčastěji využívaných skupin prediktorů úspěšnosti studentů v blended learningových (ale i online) kurzech jsou temporální charakteristiky návštěvnosti. Jde o proměnné, které se nějakým způsobem dotýkají časovosti navštěvování kurzu studenty, a zároveň se zaměřují na kurz a jeho návštěvnost jako celek, než by se týkaly dílčích výukových aktivit v kurzu. V oblasti data miningu ve vzdělávání a analytiky učení jsou pak výzkumy využívající tento typ prediktorů úspěšnosti studentů teoreticky a věcně ukotvovány nejčastěji v kontextu seberegulovaného učení (viz např. Pintrich, 2004; Pintrich, Smith, Garcia, & McKeachie, 1993; Winne, 2010; Zimmerman & Martinez-Pons, 1988) a zvláště v těch aspektech seberegulovaného učení, které se týkají plánování a time managementu (srov. AlJarrah, Thomas, & Shehab, 2018; Jo, Kim, & Yoon, 2015; Kim, Park, Yoon, & Jo, 2016; Kim, Yoon, Jo, & Branch, 2018; You, 2016). Dosavadní výzkum v oblasti e-learningu totiž naznačuje, že právě seberegulace a volba odpovídajících učebních strategií je jedním z důležitých faktorů úspěchu nejen v online, ale i v blended learningových kurzech (Broadbent, 2017; You, 2016).

Podle některých autorů (např. Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; You, 2016) však jednoduché souhrnné počty záznamů o aktivitě studentů v kurzu (tj. logy) vypovídají jen naprosté minimum (pokud vůbec něco) o jejich učebních strategiích, plánování, organizaci času apod. Proto se řada autorů zaměřuje na hledání sofistikovanějších metrik a proxy indikátorů, které by více vypovídaly o výše uvedených obecných konceptech. Někteří autoři (např. AlJarrah, Thomas, & Shehab, 2018) tak místo celkových počtů logů pracují s počtem logů za určitou časovou jednotku (např. týden, den, první polovina kurzu apod.), což už může poskytovat doplňující informace o tom, jestli jsou např. studenti v kurzu aktivní od samotného začátku, případně o tom, jak se jejich aktivita v čase vyvíjí. González Nespereira et al. (2014; 2015) pak jdou v průzkumu tohoto přístupu ještě o krok dál a za využití analýzy časových řad a dekompozice dílčích trendů jsou schopni odlišit různou míru aktivity a její vývoj v průběhu konání kurzu mezi úspěšnými a neúspěšnými studenty.

Řada autorů (např. Jo, Park, Kim, & Song, 2014; Sheshadri, Gitinabard, Lynch, Barnes, & Heckman, 2018; Yu & Jo, 2014) místo počtu logů využívá počet návštěv, resp. počet přihlášení do kurzu, případně využívají jak počet jednotlivých logů, tak i počet návštěv. To opět může poskytovat podrobnější obrázek o aktivitě studentů v kurzu, jelikož stejné množství záznamů o aktivitě může být na jedné straně rozděleno do pěti různých návštěv, zatímco na druhé straně může proběhnout v jediné návštěvě. Počet návštěv tak více vypovídá o tom, jak moc či jak často studenti skutečně „přichází“ do určitého kurzu, přičemž některé výsledky naznačují, že četnost návštěv je významným prediktorem úspěchu v kurzu (např. You, 2016).

Na druhou stranu samo o sobě zůstává výzkumným problémem vymezení toho, co je to vlastně návštěva kurzu, resp. identifikace jednotlivých návštěv. V online prostředí obecně, tedy i v systémech typu LMS, totiž uživatelé obvykle nijak explicitně neukončují svoji návštěvu (tj. např. neklikají na tlačítko pro odhlášení ze systému), a tudíž je vždy nutné nějakým způsobem odvozovat, kdy jednotlivé návštěvy kurzu skutečně končí (srov. Jo, Kim, & Yoon, 2015; Kim, Park, Yoon, & Jo, 2016).

Relativně dlouhou dobu je výzkumná pozornost rovněž věnována celkovému množství času stráveného v kurzu či při řešení určitého úkolu (tzv. *time-on-task*), jak ukazují Kovanović et al. (2015a, 2015b). Podobně jako v případě počtu návštěv však nejde o triviální problém, nýbrž se vždy musí přistupovat k určité formě odhadu na základě dostupných dat. V kontextu predikce úspěšnosti studentů v online či blended learningových kurzech v LMS se nejčastěji pracuje s celkovým množstvím času, který student stráví prací v kurzu. Dosavadní výsledky využití celkového času stráveného v kurzu pro predikci úspěšnosti však nedávají jednoznačnou odpověď. Výsledky autorů Yu a Jo (2014) naznačují, že celkový čas strávený v kurzu může být důležitým prediktorem. Dle výsledků Joa et al. (2014) však může záležet na typu či formě kurzu. Zatímco v kurzu založeném na komunikaci v online diskuzních fórech byl celkový čas druhým nejsilnějším prediktorem úspěšnosti, v případě kurzu založeném na zpřístupňování studijních materiálů a plnění průběžných úkolů nehrál celkový čas strávený v kurzu vesměs žádnou roli. V novějším výzkumu Kima et al. (2018) se navíc již celkový čas strávený v kurzu ukazuje jako nevýznamný z hlediska predikce úspěšnosti.

Poměrně často je věnována pozornost i pravidelnosti (resp. nepravidelnosti) s jakou studenti přistupují do kurzu v LMS. Podle řady autorů (např. Kim, Yoon, Jo, & Branch, 2018; You, 2016; Yu & Jo, 2014) jde totiž o jeden z velmi důležitých prediktorů úspěšnosti studentů, a to jak v online kurzech, tak i v blended learningových kurzech. Jak ale naznačují Jo et al. (2014), může opět záležet na tom, o jaký kurz se jedná. Oproti celkovému času strávenému v LMS se totiž pravidelnost přístupů zdá být důležitá spíše v kurzech nediskuzního charakteru. Mimo výše uvedené temporální charakteristiky se pak v některých výzkumech věnuje pozornost i dalším ukazatelům, případně kombinacím jednotlivých ukazatelů. Například Miyamoto et al. (2015) se zaměřili na tzv. „spacing effect“⁵¹ a poukázali na důležitý vztah mezi celkovým časem stráveným v kurzu a počtem návštěv v kurzu. Konkrétně zjistili, že mezi studenty, jež stráví v kurzu stejné množství času, dosahují lepších výsledků ti, kteří navštíví kurz vícekrát (tj. čas strávený v kurzu mají rozložený do více návštěv). Především to pak platí pro ty studenty, kteří celkově tráví v kurzech méně času.

⁵¹ Jedná se o jev zdokumentovaný již Ebbinghausem (1885) při výzkumu paměti. Obecně jde o to, že tzv. rozložené učení, kdy k učení dochází po menších částech a s průběžným opakováním, je z hlediska retence učební látky lepší než učení tzv. nakupené, kdy se student učí jednorázově velké množství obsahu.

7.1.3 Otázky zobecnitelnosti a přenositelnosti prediktivních modelů

I přes to, že byl výše představen určitý rozvoj v oblasti predikce úspěšnosti studentů v kontextu blended learningových kurzů, je nutné jedním dechem dodat, že je výzkum této problematiky stále do značné míry v počátcích. Vesměs všechny výše citované studie totiž trpí výraznými nedostatky, co se týče otázek zobecnitelnosti a přenositelnosti zjištěných výsledků, jelikož využívají data z jednoho či pouze několika málo různých kurzů. Ostatně v celé oblasti analytiky učení a data miningu ve vzdělávání se téma zobecnitelnosti a přenositelnosti dostává do popředí teprve v posledních letech. Jak uvádějí Gašević, Dawson a Siemens (2015), přestože byla věnována jistá pozornost identifikaci hlavních indikátorů úspěšnosti studentů, existuje zcela zásadní nedostatek studií, které by usilovaly o zhodnocení dopadu a přenositelnosti výsledků této počáteční práce napříč různými doménami a kontexty. A to se netýká pouze blended learningových kurzů, ale problematiky predikce úspěšnosti studentů v online kurzech obecně. Nedostatek studií tohoto typu přitom přetrvává dodnes, a to i přesto, že je toto téma v kontextu predikce úspěšnosti obecně považováno za jedno z nejdůležitějších (srov. Romero & Ventura, 2019). Navíc i studie, které se výslovně zaměřují na otázky zobecnitelnosti prediktivních modelů, využívají data z poměrně malého množství kurzů. Např. aktuální studie Gitinabardové et al. (2019) věnující se zobecnitelnosti přímo v kontextu blended learningu využívá při analýzách data pouze ze dvou běhů dvou různých kurzů.

Přitom zvláště v kontextu blended learningu se zdají být otázky zobecnitelnosti velmi palčivé. Např. již zmiňovaní Jo, Park, Kim a Song (2014), kteří analyzovali data ze dvou typově odlišných blended learningových kurzů, zjistili zásadní rozdíly v tom, které proměnné fungují jako prediktory úspěšnosti. Podobně také Estaciová a Raga (2017), jež analyzovali čtyři různé kurzy, našli značnou variabilitu mezi jednotlivými kurzy, co se týče potenciálních prediktorů úspěšnosti studentů. Uzávívají tak, že právě povaha či design kurzů je zřejmě oním hlavním faktorem, který způsobuje variabilitu v účinnosti jednotlivých prediktorů. Joksimović et al. (2015) dochází na základě výsledků své studie k ještě o něco silnějšímu tvrzení a uvádí, že rozdíly mezi kurzy jsou dokonce v kontextu predikce úspěšnosti studentů důležitější než rozdíly mezi jednotlivými studenty. Jako potenciálně důležité faktory na úrovni kurzů pak uvádí design kurzu, styl či způsob výuky učitele, využívané formy hodnocení a obor, resp. disciplínu, které se kurz týká. Přestože se studie nevěnuje přímo temporálním charakteristikám návštěvnosti jakožto prediktorům úspěšnosti, její výsledky jsou extrémně důležité, jelikož jde o jednu z mála studií, která pracuje s relativně velkým počtem kurzů. Za velmi důležitou pak lze považovat i studii Gaševicé et al. (2016), která rovněž velmi silně zdůrazňuje, že vztah mezi aktivitou studentů v LMS a jejich následnými výsledky v kurzech je výrazně moderován různým designem kurzů. Autoři jsou přitom přesvědčeni, že tyto

rozdíly v designu jednotlivých kurzů musí být brány v úvahu při snahách o tvorbu obecných prediktivních modelů. Nezohlednění těchto odlišností na úrovni kurzů podle nich povede k nadhodnocování, či naopak k podceňování vlivu jednotlivých indikátorů při predikci úspěšnosti studentů.

7.2 Řešený problém a otázky

S ohledem na výše uvedená východiska a prezentované dosavadní výzkumy usiluje tato analýza o zodpovězení následující hlavní výzkumné otázky (HVO):

- *HVO: Jaké jsou možnosti predikce úspěšného absolvování blended learningových kurzů na základě temporálních aspektů návštěvnosti jejich online opor v LMS?*

Výše uvedená hlavní výzkumná otázka se přitom v rámci této analýzy rozpadá do tří specifických výzkumných otázek (SVO). První specifická výzkumná otázka se primárně věnuje možnostem predikce úspěšnosti studentů a jejím cílem je zjistit, jak se možnosti predikce mění při použití různých prediktivních technik, resp. které techniky jsou v rámci predikce úspěšné a do jaké míry. Jelikož je v analyzovaných datech úspěšnost studentů v kurzech vyjádřena v podobě binární proměnné (tj. úspěch versus neúspěch), jde v kontextu data miningu a strojového učení o typ úlohy označovaný jako klasifikace⁵². V tomto případě jde o binární klasifikaci, kdy je třeba na základě vstupních dat (tj. prediktorů – zde temporálních aspektů návštěvnosti) správně zařadit studenta do jedné ze dvou kategorií: úspěšné splnění kurzu, či neúspěšné splnění kurzu.

Pro samotnou klasifikaci pak lze použít širokou škálu různých klasifikačních algoritmů; od relativně dlouho známé, a i mimo oblast data miningu široce používané, logistické regrese až po v dnešní době velmi populární umělé neuronové sítě. V kontextu analytiky učení a data miningu ve vzdělávání je predikce úspěšnosti studentů v kurzech za využití metody klasifikace jednou z nejvíce akcentovaných výzkumných oblastí (viz kapitolu 6.2 či Papamitsiou & Economides, 2014; Peña-Ayala, 2014b). Mezi nejčastěji používané analytické techniky a klasifikační algoritmy pak patří logistická regrese a algoritmy založené na technice rozhodovacích stromů (Peña-Ayala, 2014b). Při zjišťování možností predikce úspěšnosti studentů za využití různých analytických metod jsou tak jako klasifikační algoritmy využity primárně právě logistická regrese a rozhodovací stromy. Mimo to je rovněž použita technika náhodného lesa, která také vychází z rozhodovacích stromů a která obvykle dosahuje lepších výsledků. Jako poslední klasifikační algoritmus je využita technika podpurných vektorů, jež zatím v kontextu analytiky učení a data miningu

⁵² V terminologii strojového učení bývá klasifikace rovněž považována za jednu z metod tzv. učení s učitelem.

ve vzdělávání není příliš často využívána⁵³. První specifickou výzkumnou otázku tak lze formulovat následovně:

- *SVO1: Jaké jsou predikční schopnosti vybraných klasifikačních algoritmů (logistická regrese, rozhodovací strom, náhodný les a algoritmus podpůrných vektorů) při predikci úspěšnosti studentů v blended learningových kurzech?*

V rámci druhé specifické výzkumné otázky je věnována pozornost analyzovaným temporálním aspektům návštěvnosti e-learningových opor blended learningových kurzů v LMS. Cílem je přitom zjistit, které proměnné vyjadřující určité temporální aspekty návštěvnosti jsou pro predikci nejdůležitější (tj. nejpřínosnější z hlediska predikce úspěšnosti). Jak již bylo naznačeno v předchozí části, přestože existuje několik studií věnujících se temporálním aspektům návštěvnosti (např. Jo, Kim, & Yoon, 2015; Kim, Park, Yoon, & Jo, 2016; Kim, Yoon, Jo, & Branch, 2018; Miyamoto et al., 2015), zatím není dostatečně zřejmé, jaké konkrétní proměnné týkající se temporálních aspektů návštěvnosti jsou nejhodnějšími prediktory úspěšnosti studentů. Zvláště doposud nebyla věnována dostatečná výzkumná pozornost temporálním aspektům návštěvnosti v kontextu blended learningových kurzů, což se snaží tato analýza kompenzovat. Navíc autoři doposud realizovaných studií nepracovali s příliš velkými vzorky, které by obsahovaly větší množství různých kurzů. Druhou specifickou výzkumnou otázku této analýzy lze proto formulovat následovně:

- *SVO2: Jaké proměnné týkající se temporálních aspektů návštěvnosti e-learningových opor blended learningových kurzů v LMS Moodle jsou nejdůležitější při predikci úspěšnosti studentů?*

Třetí specifická výzkumná otázka se zaměřuje na ověření toho, nakolik jsou testované klasifikační algoritmy schopny predikovat úspěšnost studentů napříč různými blended learningovými kurzy, tzn. zda a případně nakolik je možné natrénovat klasifikační algoritmy na datech z určité množiny kurzů a následně predikovat úspěšnost studentů v jiné (zcela nové) množině kurzů. Jedno z důležitých specifík vzdělávacího kontextu je totiž to, že zkoumané fenomény na individuální úrovni bývají relativně často významně ovlivňovány vyššími úrovněmi, jako je např. úroveň třídy (či úroveň kurzu v LMS) nebo úroveň školy. Tyto vyšší úrovně jsou přitom potenciálním zdrojem další variability (typicky např. rozdíly mezi různými školami), jež je odlišná od variability na individuální úrovni. Ignorování víceúrovňové povahy zkoumaného fenoménu (resp. dat, která máme k dispozici) tak může vést k nesprávným závěrům. Specificky v souvislosti s predikcí úspěšnosti studentů v blended learningových kurzech přitom dosavadní výzkumy naznačují, že na úrovni kurzů se skutečně nachází velké množství variability, která je dána především různou povahou a designem jednotlivých kurzů.

⁵³ Přesto je nutné zmínit, že v několika studiích tato technika využita je (např. Rivera, Goel, & Kincaid; 2016; Şen, Uçar, & Denle, 2012; Zafra, Romero, & Ventura, 2011).

V kontextu českého pedagogického výzkumu se zatím zohledňování víceúrovňové povahy zkoumaného problému objevuje zejména v souvislosti s velkými mezinárodními výzkumy jako je PISA, ICLIS, TIMSS či PIRLS, které již z charakteru svého designu vyžadují víceúrovňovou analýzu dat (viz Juhaňák, Zounek, Záleská, Bárta, & Vlčková, 2018, 2019; Soukup, 2016; Straková, 2016). Přesto, jak upozorňuje např. Soukup (2006), důležitost zohledňování různých úrovní a využití odpovídajícího způsobu analýzy se týká i jiných zkoumaných oblastí, a to nejen v rámci pedagogického výzkumu. V kontextu výzkumu v oblasti analytiky učení a data miningu ve vzdělávání se prozatím většina studií ve svých analýzách zaměřuje spíše na studenty jednoho či několika málo kurzů, a nepracuje tak s daty víceúrovňového charakteru (např. Estacio & Raga, 2017; Jo, Park, Kim, & Song, 2014; Lu, Huang, Lin, Ogata, & Yang, 2018; Sheshadri, Gitinabard, Lynch, Barnes, & Heckman, 2018; Yu & Jo, 2014). To však znamená jednak to, že zobecnitelnost a aplikovatelnost takových zjištění může být značně omezená (v některých případech mohou výsledná zjištění platit skutečně pouze pro daný zkoumaný kurz), jednak není doposud dostatečně známo, nakolik se víceúrovňová povaha dat (tj. data týkající se studentů v různých kurzech) promítá do predikčních schopností klasifikačních algoritmů běžně používaných v oblasti LA a EDM. Třetí specifickou otázkou pokládanou v této analýze lze tudíž formulovat následovně:

- *SVO3: Jak se mění míra úspěšné predikce vybraných klasifikačních algoritmů (logistická regrese, rozhodovací stromy, náhodný les a podpůrné vektory) při jejím testování externí validací a při zohlednění dvouúrovňového charakteru dat?*

7.3 Metodologie

Charakter této analýzy lze vzhledem ke stanoveným cílům považovat především za evaluační. A to v tom smyslu, že primárně usiluje o zhodnocení účinnosti jednotlivých prediktivních metod a použitých prediktorů týkajících se temporálních aspektů návštěvnosti blended learningových kurzů v LMS Moodle. V následujících částech této metodologické kapitoly bude nejprve představena problematika klasifikace jakožto obecný metodologický přístup použitý k predikci úspěšnosti studentů v této analýze. Poté budou podrobně popsána analyzovaná data a představeny použité proměnné. V závěru kapitoly je pak podrobně nastíněn proces analýzy a evaluace jednotlivých klasifikačních algoritmů.

7.3.1 Metodologický přístup – klasifikace

V kontextu data miningu a strojového učení jsou dílčí metodologické oblasti či směry, kterým výzkumníci věnují pozornost, obvykle rozdělovány podle základního

problému či úkolu, na jehož řešení se primárně zaměřují. Jedním z takových problémů a jednou z hlavních oblastí data miningu a strojového učení je právě problém klasifikace. V případě klasifikace přitom jde o takový druh problému či úkolu, v jehož rámci je cílem správné zařazení nového (doposud neznámého) případu do odpovídající z možných kategorií, a to na základě dostupných (trénovacích) dat obsahujících jiné případy, u kterých je správné zařazení známo (Gorunescu, 2011).

Samotný proces klasifikace je možné popsat na základě jeho následujících součástí (srov. Dean, 2014; Gorunescu, 2011; Han & Kamber, 2006; Witten, Frank, & Hall, 2011):

- Třída (*class*) není nic jiného než závisle proměnná kategorického charakteru, se kterou při klasifikaci pracujeme a kterou se snažíme predikovat. V případě této analýzy je tedy třídou závisle proměnná indukující, zda student úspěšně ukončil kurz, či nikoli (tj. jde o binární klasifikaci, jelikož je třída tvořena pouze dvěma kategoriemi).
- Prediktory. Vedle třídy musíme mít sadu dalších proměnných, na jejichž základě pak bude docházet k odhadu správné kategorie v rámci třídy. Tyto další proměnné se označují jako prediktory (někdy též vlastnosti či atributy aj.) a odpovídají tomu, co v „tradičním“ statistickém výzkumu bývá označováno jako nezávisle proměnné.
- Trénovací data. Při klasifikaci rozlišujeme dva druhy dat: trénovací a testovací data. Trénovací data jsou taková data, která obsahují obě výše uvedené složky (tj. třídu a sadu prediktorů) a která jsou použita pro trénování klasifikačního modelu pomocí určitého klasifikačního algoritmu.
- Klasifikační algoritmus lze chápat jako specifický postup, pomocí kterého dochází k tvorbě klasifikačního modelu na základě vstupních (trénovacích) dat. Cílem klasifikačního algoritmu je tedy na základě dostupných dat nalézt co možná nejlepší (obvykle nejpřesnější) klasifikační model. Způsob hledání a tvorby klasifikačního modelu se pak liší v závislosti na konkrétním klasifikačním algoritmu. V této analýze jsou přitom využity následující klasifikační algoritmy: logistická regrese, rozhodovací strom, náhodný les a algoritmus podpůrných vektorů.
- Klasifikační model je výstupem klasifikačního algoritmu aplikovaného na trénovacích datech. Jde o model (z matematického hlediska funkci), který na základě sady prediktorů přiřazuje každý případ do odpovídající kategorie v rámci třídy. Klasifikační model je tedy tou komponentou, která následně zajišťuje predikci pro nové případy, u nichž již neznáme správnou kategorii v rámci dané třídy.
- Testovací data jsou data, která se používají k evaluaci vytvořeného klasifikačního modelu. Strukturně jsou stejná jako trénovací data (tj. opět obsahují prediktory a třídu), všem nejsou využita při trénování klasifikačního

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech

modelu⁵⁴. Data v testovacím souboru se na základě dostupných prediktorů nechají klasifikovat pomocí vytvořeného klasifikačního modelu. Srovnáním skutečných hodnot v rámci třídy s těmi predikovanými pak dochází ke zhodnocení výkonu modelu.

Konkrétnější informace o způsobu využití klasifikace v rámci této analýzy jsou prezentovány v části 7.3.4 Postup analýzy dat.

7.3.2 Analyzovaná data, sběr a předzpracování dat

Data použitá v následujících analýzách tvoří celkem 2 321 studentů studujících v 35 různých blended learningových kurzech s různou mírou využití online podpory v LMS. Všechny kurzy probíhaly v LMS Moodle na Filozofické fakultě Masarykovy univerzity a byly vyučovány v průběhu jednoho semestru. Konkrétně šlo o jarní semestr 2014. Kritérium pro výběr kurzů bylo dvojí:

1. Hlavním kritériem bylo to, zda byla v rámci kurzů dostupná informace o minimálních požadavcích na studenty pro splnění kurzu⁵⁵. Zároveň musela být v rámci kurzu využita nějaká forma sledování plnění stanovených požadavků⁵⁶, jež umožňovala rozlišit studenty, kteří splnili stanovené požadavky (a tím úspěšně absolvovali kurz) a kteří nikoli.
2. Doplnujícím kritériem pak byla velikost kurzů ve smyslu počtu studentů, kteří daný kurz studují. Vybrány byly jen takové kurzy, ve kterých studovalo více jak 25 studentů. Hlavním důvodem pro zaměření pozornosti spíše na větší kurzy bylo to, aby v rámci jednotlivých kurzů bylo dostatek jednotek pro trénování použitých klasifikačních algoritmů. Celkově se počet studentů ve vybraných kurzech pohyboval od 26 do 334, přičemž střední hodnota (Me) byla 47 studentů v kurzu.

Po extrakci dat z databáze systému Moodle bylo hlavním úkolem jejich transformování do takové podoby, v níž by mohly vstupovat do plánovaných analýz. Východí data o návštěvnosti kurzů jsou totiž v systému LMS ukládána pomocí tzv. logů (viz tabulka 5). Každý záznam o aktivitě studenta (tj. log) v LMS pak obsahuje

54 Běžný postup při klasifikaci je tudíž takový, že se dostupná data nejprve rozdělí na trénovací a testovací, přičemž testovací data nejsou využívána, dokud nedojde k vytvoření klasifikačního modelu a nepřistoupí se k jeho hodnocení.

55 Tj. v kurzu byla vždy stanovena jednoznačná hranice, co musí student splnit, aby úspěšně absolvoval kurz. V zásadě šlo o stanovení bodové hranice, kterou musí student získat, aby úspěšně absolvoval kurz.

56 Tj. pro každého studenta bylo sledováno (s)plnění stanovených aktivit v online podpoře kurzu. V zásadě šlo o využití funkce pro zadávání známek, resp. bodového hodnocení k jednotlivým aktivitám studentů. To znamená, že vyučující v daném kurzu tuto funkcionalitu musel využívat a aktivity studentů hodnotit. Pro kontext je možné doplnit, že v LMS Moodle na FF MU není neobvyklá situace, kdy vyučující ve svých kurzech tuto funkci vůbec nevyužívají.

mimo jiné unikátní identifikátor (*log.id*), informaci o času, kdy k aktivitě došlo (*log.time*), identifikátor uživatele, který aktivitu provedl (*log.userid*), identifikátor kurzu, ve kterém byla aktivita provedena (*log.course.id*), informaci o tom, v jakém nástroji či v jakém kontextu aktivita proběhla (*log.module*), identifikátor konkrétního nástroje či kontextu (*log.cmid*) a informaci o tom, jaký typ aktivity byl proveden (*log.action*).

Tabulka 5: Příklad tabulky logů v LMS Moodle⁵⁷ (identifikátory byly anonymizovány)

log.id	log.time	log.userid	log.courseid	log.module	log.cmid	log.action
100	12.06.2014 23:58	111	123	course	123456	view
200	13.06.2014 1:03	222	123	course	123456	view
300	13.06.2014 1:26	222	123	course	123456	view
400	19.06.2014 14:13	333	123	resource	232323	view
500	19.06.2014 14:15	333	123	resource	232323	view
600	19.06.2014 14:20	333	123	resource	232323	view
700	02.03.2014 11:38	444	321	url	445544	view
800	02.03.2014 11:38	444	321	url	445544	view
900	16.06.2014 14:36	555	321	quiz	363636	attempt
1000	16.06.2014 14:39	555	321	quiz	363636	attempt
1100	16.06.2014 14:39	555	321	quiz	363636	attempt

Cílem transformace dat bylo převedení výchozích dat v podobě logů do podoby souhrnných dat, která by na každém řádku obsahovala určitého studenta určitého kurzu a v rámci jednotlivých sloupců by poskytovala agregované informace o návštěvnosti daného studenta v daném kurzu (za využití jednotlivých ukazatelů návštěvnosti použitých v této analýze), a nadto informaci o úspěšnosti, či neúspěšnosti daného studenta při absolvování daného kurzu. Teprve po této transformaci dat bylo možné přistoupit k samotné analýze a aplikaci klasifikačních algoritmů.

7.3.3 Použité proměnné

Jako závisle proměnná, resp. třída, figuroval v datech binární indikátor (Ano/Ne, resp. 1/0) toho, zda student splnil minimální požadavky pro ukončení kurzu, a tedy zda úspěšně splnil kurz. V podobě nezávisle proměnných, resp. prediktorů, pak do realizovaného modelování vstupovalo několik kvantitativních

⁵⁷ Nutno doplnit, že v současnosti již systém Moodle pracuje s bohatší tabulkou logů, která obsahuje více údajů.

temporálních charakteristik návštěvnosti blended learningových kurzů, resp. on-line studijních opor blended learningových kurzů. Výběr nezávisle proměnných, které byly v této analýze využity, vycházel ze studia dosavadní literatury věnující se problematice temporálních aspektů návštěvnosti blended learningových kurzů (viz kapitolu 7.1.2). Celkem bylo v realizovaných analýzách využito devět různých prediktorů. Pro účely jejich popisu lze přitom hovořit o pěti základních skupinách proměnných.

V první skupině tvořené třemi proměnnými se vychází z počtu záznamů o aktivitě studenta v kurzu (tj. z celkového počtu logů). Pro zjednodušení lze tuto míru chápat jako počet kliknutí myši v rámci kurzu⁵⁸. Celkový počet záznamů o aktivitě byl přitom spočítán zvlášť za tři různá období v průběhu semestru: období výuky, zkouškové období a prodloužené zkouškové období (jak bylo řečeno, šlo o jarní semestr). Přesné vymezení jednotlivých tří období se řídilo oficiálním harmonogramem Filozofické fakulty Masarykovy univerzity pro daný semestr. První skupinu proměnných tedy tvoří následující (v závorce je uveden zkrácený název používaný dále v textu a v tabulkách):

- Celkový počet záznamů o aktivitě studenta v průběhu výuky (*Počet záznamů – výuka*).
- Celkový počet záznamů o aktivitě studenta v průběhu zkouškového období (*Počet záznamů – zkouškové*).
- Celkový počet záznamů o aktivitě studenta v průběhu prodlouženého zkouškového období (*Počet záznamů – p. zkouškové*).

Druhou skupinu proměnných tvoří počty návštěv studentů v kurzech. Konceptuálně se návštěvou v tomto kontextu chápe situace, kde student vstoupí do kurzu v LMS, nějakou dobu se zde pohybuje či pracuje a následně kurz a práci v něm opouští. Když se student po nějaké době (např. za týden) do kurzu opět vrací, jde již o další návštěvu. Technicky však nelze jednotlivé návštěvy zcela přesně odlišit, jelikož konec návštěvy obvykle není v systému nijak explicitně zaznamenáván (tj. student jednoduše přestane v systému pracovat a zavře okno prohlížeče, aniž by se ze systému explicitně odhlásil kliknutím na odpovídající tlačítko). Proto bývá rozlišení jednotlivých návštěv obvykle (viz např. Kovanović et al., 2015b) řešeno tak, že je stanovena určitá prahová hodnota neaktivity, která indikuje konec návštěvy. V případě této analýzy byla zvolena prahová hodnota neaktivity 30 minut. Zjednodušeně to lze chápat tak, že pokud student v kurzu průběžně kliká myší a následně alespoň třicet minut systém nezaznamená v kurzu od daného studenta žádné kliknutí, je to vyhodnoceno tak, že studentova návštěva v kurzu skončila.⁵⁹

58 Tato paralela slouží pouze jako přiblížení. Ve skutečnosti se totiž nezaznamenává úplně každé kliknutí myši. Zároveň mohou naopak některé aktivity studenta v kurzu vygenerovat více jak jeden záznam v logovacím systému za jedno kliknutí myši.

59 Že využití tohoto postupu nemusí být vždy zcela přesné lze v kontextu využití LMS ilustrovat např. situací, kdy má student otevřen (rozsáhlý) studijní materiál, který čte déle, než je stanovená prahová

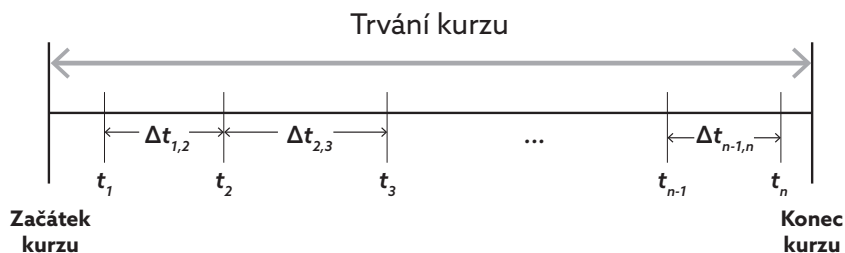
I v případě počtu návštěv přitom byla věnována pozornost výše uvedeným třem obdobím v průběhu semestru, celkem jsem tedy pracoval s následujícími třemi proměnnými:

- Celkový počet návštěv studenta v kurzu v průběhu výuky (*Počet návštěv – výuka*).
- Celkový počet návštěv studenta v kurzu v průběhu zkouškového období (*Počet návštěv – zkouškové*).
- Celkový počet návštěv studenta v kurzu v průběhu prodlouženého zkouškového období (*Počet návštěv – p. zkouškové*).

V souvislosti s návštěvností byly pro každého studenta a každý kurz vypočítány další dvě proměnné, které byly následně použity ve všech analýzách. První proměnná odpovídá době od začátku kurzu do prvního přihlášení studenta do e-learningové podpory kurzu (První návštěva kurzu), druhá pak vyjadřuje pravidelnost (resp. nepravidelnost), s jakou student do kurzu přistupuje (Pravidelnost návštěv). V případě proměnné První návštěva kurzu je použitou jednotkou počet dní od začátku kurzu do prvního přihlášení studentem do kurzu. V případě pravidelnosti návštěv pak jde o směrodatnou odchylku časových intervalů mezi jednotlivými návštěvami kurzu (srov. Jo, Kim, & Yoon, 2015; Kim, Park, Yoon, & Jo, 2016; Kim, Yoon, Jo, & Branch, 2018). Pro lepší pochopení výpočtu této proměnné lze využít obrázek níže (viz obrázek 7) znázorňující celkové trvání kurzu pomocí úsečky, jejímiž krajními body jsou začátek a konec kurzu. V průběhu trvání kurzu pak jednotliví studenti kurz v určitých časových okamžicích navštěvují, což znázorňují body t_1 až t_n (zcela přesně řečeno jde o časy začátku jednotlivých návštěv). Vzdálenost mezi body t_1 a t_2 (tj. $\Delta t_{1,2}$) pak odpovídá časovému intervalu mezi první a druhou návštěvou studenta, který je vypočítán tak, že se od času t_2 odečte čas v bodu t_1 . Obdobně je tomu i v případě intervalu $\Delta t_{2,3}$ a ve všech dalších časových intervalech až po poslední interval $\Delta t_{n-1,n}$. Z těchto hodnot je následně vypočítána směrodatná odchylka, která vypovídá o tom, nakolik se od sebe liší délka časových intervalů mezi návštěvami. Čím je hodnota směrodatné odchylky menší, tím pravidelněji student kurz navštěvuje.

hodnota (např. 30 minut). V tomto okamžiku student v kurzu reálně pracuje, ovšem jeho aktivita (tj. čtení studijního materiálu) se neodráží v klikání myši, a tudíž je toto vyhodnoceno jako konec návštěvy v kurzu. Podobnou situaci může být např. sledování delšího naučného videa v kurzu.

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech



Obrázek 7: Konceptualizace časových intervalů mezi jednotlivými návštěvami kurzu (podle Jo, Kim, & Yoon, 2015)

Poslední použitou proměnnou je pak celkový čas strávený studentem v e-learningové podpoře kurzu v průběhu celého semestru (*Celkový čas v kurzu*), přičemž jednotkou je zde počet minut. V rámci této proměnné se opět vychází z prahové hodnoty 30 minut neaktivity, která signalizuje konec návštěvy studenta v kurzu. Jako délka trvání návštěvy je přitom počítán čas od prvního záznamu aktivity po předchozí 30minutové neaktivitě (příp. od úplně prvního záznamu aktivity studenta v kurzu) až po nejbližší nadcházející záznam aktivity, který předchází další 30minutové neaktivitě⁶⁰. K této hodnotě je přitom za každou návštěvu připočítána tzv. kompenzace času stráveného na poslední aktivitě. Jak již bylo zmíněno, studenti obvykle nekončí svoji práci tím, že by klikli na tlačítko pro odhlášení se ze systému, nýbrž jednoduše zavřou okno ve svém prohlížeči. To vede k tomu, že není zaznamenán čas strávený v kurzu na poslední aktivitě před skutečným odchodem z kurzu⁶¹. Proto, že není možné zjistit skutečnou délku trvání této poslední aktivity, je alespoň jako určitá aproximace použita právě kompenzace času na poslední aktivitě. Ta je konkrétně vypočítána jako průměr časových intervalů mezi jednotlivými záznamy o aktivitě studenta v průběhu návštěv v kurzu⁶².

Z celkového počtu 2 321 studentů jich 1 462 (63 %) splnilo minimální požadavky pro úspěšné ukončení kurzu. Zbývajících 859 (37 %) neukončilo studovaný kurz úspěšně. Základní deskriptivní údaje o nezávisle proměnných (prediktorech) jsou pak k dispozici v následující tabulce (viz tabulka 6).

60 Pro snadnější pochopení je relativně přesným přiblížením to, že se v tomto kroku jednoduše spočítá doba mezi prvním a posledním kliknutím studenta v kurzu v rámci jedné návštěvy.

61 Příkladem je následující situace: student nějakou dobu pracuje v kurzu a na závěr své návštěvy si ještě kliknutím otevře studijní materiál, který následně nějakou dobu čte. Poté zavře okno prohlížeče a vypne počítač. Poslední záznam o aktivitě v rámci návštěvy je tak okamžik, kdy si otevřel studijní materiál. Ve skutečnosti však ještě trávil nějaký čas čtením otevřeného studijního materiálu, a proto je žádoucí tento čas (resp. jeho odhad) k naměřenému trvání návštěvy připočítat.

62 Můžeme si to představit tak, že v rámci svých návštěv v kurzu dělá student různé aktivity (plnění testu, komunikace ve fóru apod.), které mu zaberou různé množství času. Kompenzace času na poslední aktivitě je pak průměrná doba trvání těchto aktivit.

Tabulka 6: Deskriptivní popis nezávisle proměnných (tj. prediktorů)

	Min	Max	Průměr	Medián	SD
Počet záznamů – výuka	0	2106	185,2	42	339,3
Počet záznamů – zkouškové	0	1604	51,9	1	125,5
Počet záznamů – p. zkouškové	0	370	1,1	0	11,1
Počet návštěv – výuka	0	274	14,5	8	23,6
Počet návštěv – zkouškové	0	131	3,3	1	7,3
Počet návštěv – p. zkouškové	0	59	0,2	0	2,01
První přihlášení	0,005	196,6	21,3	9,4	27,3
Pravidelnost návštěv	0,026	63,9	10	6,7	9,7
Celkový čas v kurzu	0	4111,1	233,4	85	363,1

7.3.4 Postup analýzy dat

Za účelem zodpovězení stanovených výzkumných otázek bylo provedeno srovnání predikčních schopností čtyř různých klasifikačních algoritmů (logistická regrese, rozhodovací stromy, náhodný les a podpůrné vektory), přičemž byly využity tři různé způsoby predikčního testování. V prvním kroku byla použita kompletní data a výchozí metoda testování modelu v rámci daného klasifikačního algoritmu⁶³. V druhém kroku byly všechny testované algoritmy podrobeny testování tzv. externí validací, v jejímž rámci jsou data rozdělena na trénovací a testovací část. Jak naznačuje pojmenování obou částí, trénovací část je využita k natrénování klasifikačního algoritmu, zatímco testovací část slouží k otestování úspěšnosti predikce na zcela nových datech. Rozdělení souboru na trénovací a testovací část bylo provedeno v poměru dvě ku jedné, tzn. dvě třetiny dat byly použity pro trénování klasifikátoru a jedna třetina pro jeho testování. Zároveň, aby nedošlo ke zkreslení způsobenému konkrétním výběrem trénovacích a testovacích dat, byla využita varianta externí validace s r opakováními (Holčík & Komenda, 2015). Při této variantě nedochází k rozdělení na trénovací a testovací část dat pouze jednou, nýbrž se soubor náhodně rozdělí r -krát a získaných r výsledků je poté zprůměrováno. Hlavní výhodou tohoto přístupu je přesnější odhad úspěšnosti klasifikace. V rámci této analýzy bylo při externí validaci využito 100 opakování.

Ve třetím kroku byla opět aplikována externí validace se 100 opakováními při poměru dvě třetiny na trénink a třetina na testování, ovšem navíc byla při rozdělování dat na trénovací a testovací množinu zohledňována úroveň jednotlivých kurzů. To znamená, že bylo zajištěno, aby všichni studenti ze stejného kurzu byli buď

63 Bude popsáno v sekci výsledků přímo u jednotlivých klasifikačních algoritmů.

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech

v trénovací, nebo testovací množině. V trénovací množině tedy byla data týkající se jiných kurzů než v testovací množině. Taková forma externí validace byla zvolena proto, aby bylo možné ověřit, nakolik jsou klasifikační algoritmy schopny úspěšně predikce v případě, kdy jsou natrénované na určité množině kurzů a následně testované na zcela nových kurzech.

Při aplikaci klasifikačních algoritmů a vlastním modelování byl přítom ve všech fázích predikčního testování následován běžný postup popisovaný řadou odborných publikací věnujících se data miningu či strojovému učení (viz např. Gorunescu, 2011; Kelleher, Mac Namee, & D'Arcy, 2015; Witten, Frank, & Hall, 2011). Konkrétně lze tento postup popsat následovně:

1. Tvorba a vyhodnocení výchozího modelu. Nejprve proběhlo předzpracování dat z vybraných kurzů, jak bylo popsáno v kapitole 7.3.2, a byla vytvořena finální datová sada obsahující devět nezávisle proměnných, jednu závisle proměnnou a jednu proměnnou sloužící jako identifikátor kurzu⁶⁴. Všech devět nezávisle proměnných pak bylo využito při tvorbě výchozího modelu predikujícího binární třídu. V rámci vyhodnocování úspěšnosti výchozího modelu byla vždy vytvořena 2×2 kontingenční tabulka predikovaných a skutečných hodnot a zobrazena tzv. ROC (*Receiver Operating Characteristic*) křivka, běžně využívaná při evaluaci a optimalizaci binárního klasifikátoru. S využitím vytvořené kontingenční tabulky a ROC křivky pak byly vypočítány čtyři základní indikátory, na jejichž základě byla posuzována kvalita modelu. Konkrétně šlo o přesnost, senzitivitu, specificitu a plochu pod ROC křivkou (běžně označováno anglickou zkratkou *AUC* – *Area Under the Curve*)⁶⁵.
2. Hledání a výběr parametrů modelu (tzv. *parameter tuning*). Po tvorbě výchozího modelu probíhalo hledání a výběr optimálních parametrů klasifikačního algoritmu. Tato fáze se samozřejmě lišila podle jednotlivých klasifikačních algoritmů a dle toho, jaké parametry tyto algoritmy vyžadují specifikovat. Ve všech případech byly parametry vybírány tak, aby byla co možná nejvíce zvýšena schopnost modelu správně predikovat úspěšnost studentů v blended learningových kurzech. V případě logistické regrese docházelo k hledání optimálního cut-off skóre pomocí tzv. Youdenova

64 Identifikátor kurzu byl využit pouze při rozdělování dat na trénovací a testovací množinu. Při samotném modelování tato proměnná využita nebyla.

65 Doplňme, že přesnost odpovídá poměru mezi množstvím skutečně pozitivních a skutečně negativních případů k celkovému množství případů (tj. kolik procent případů bylo klasifikováno správně). Senzitivita odpovídá poměru mezi počtem pravdivě pozitivních případů a součtu pravdivě pozitivních případů s falešně negativními případy (tj. kolik procent ze skutečně pozitivních případů bylo klasifikováno správně, resp. nakolik dochází ke správné klasifikaci úspěšných studentů). Specificita pak odpovídá poměru mezi počtem pravdivě negativních případů a součtu falešně pozitivních případů s pravdivě negativními případy (tj. kolik procent ze skutečně negativních případů bylo klasifikováno správně, resp. nakolik dochází ke správné klasifikaci neúspěšných studentů).

indexu (Youden, 1950). U rozhodovacích stromů probíhalo hledání optimální hodnoty parametru složitosti (*complexity parameter*), který určuje složitost, resp. velikost rozhodovacího stromu a umožňuje vybrat optimálně veliký/složitý rozhodovací strom (Therneau & Atkinson, 2018). V případě náhodného lesa byla vybírána optimální hodnota parametru *mtry*, která specifikuje počet proměnných náhodně vybraných pro účely dělení stromu na každém uzlu (Liaw & Wiener, 2002). Přestože v odborných publikacích panuje diskuze ohledně vlivu tohoto parametru na úspěšnost klasifikace, většina zdrojů jej zmiňuje a doporučuje mu věnovat pozornost. V případě náhodného lesa byl rovněž specifikován počet stromů (tj. parametr *ntree*), a to tak, aby byla zvolená hodnota dostatečně vysoká. Při použití algoritmu podpurných vektorů byla pro hledání vhodných hodnot parametrů využita technika vyhledávání v mřížce (*grid search*), přičemž byly hledány optimální hodnoty parametrů *epsilon* a *cost* (Kuhn, 2018).

3. Tvorba a vyhodnocení výsledného modelu. Na základě předchozího kroku byl vytvořen výsledný model, ve kterém již byly specifikovány nalezené vhodné parametry modelu. Výsledný model byl opět vyhodnocen stejným způsobem jako výchozí model, tj. pomocí kontingenční tabulky, ROC křivky, přesnosti, senzitivity, specificity a hodnoty AUC.

Vzhledem k relativní komplikovanosti analytického postupu jsou některé dílčí analytické postupy či pojmy přesně popsány či vysvětleny až v sekci výsledků přímo v souvislosti s konkrétním krokem analýzy vedoucím k prezentovaným výsledkům.

7.4 Výsledky

Základní popisné charakteristiky analyzovaných dat uvádí tabulka výše (viz tabulka 6). Jak je z tabulky patrné, proměnné nemají normální rozdělení (spíše mají blízko k exponenciálnímu rozdělení). Vzhledem k tomu, že logistická regrese ani ostatní klasifikační algoritmy nemají jako svůj předpoklad normální rozdělení, vstupovaly všechny proměnné do modelů v původní podobě bez jakýchkoli dalších transformací.

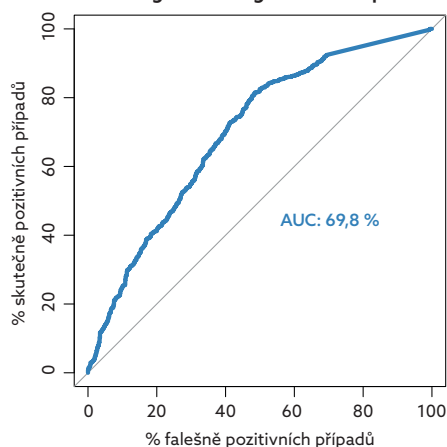
V následujících částech proběhne nejprve základní posouzení predikčních schopností analyzovaných klasifikačních algoritmů za použití kompletní datové sady a výchozí metody validace. Poté jsou prezentovány výsledky hodnocení důležitosti použitých prediktorů týkajících se temporálních aspektů návštěvnosti blended learningových kurzů. Zároveň zde provádím srovnání důležitosti jednotlivých prediktorů při použití různých klasifikačních algoritmů. Na závěr prezentuji výsledky predikčního testování použitých algoritmů pomocí externí validace, a to jak bez zohlednění úrovně kurzů, tak i při jejím zohlednění.

7.4.1 Základní posouzení predikčních schopností analyzovaných klasifikátorů

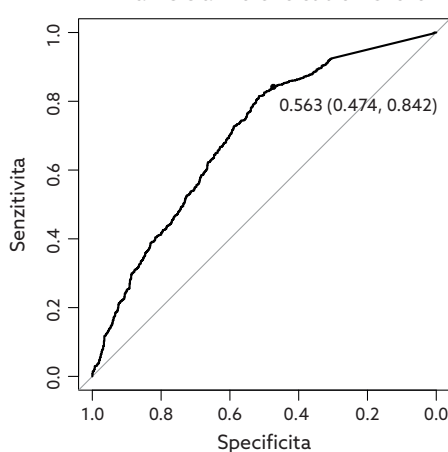
Než jsem přistoupil k samotné aplikaci klasifikačních algoritmů, byla nejprve vypočtena základní úroveň přesnosti či tzv. *baseline model*, se kterým jsou všechny následné modely srovnávány. Pro výpočet základní úrovně přesnosti byla využita technika volby nejčastější kategorie. Baseline model si tedy lze představit jako takový, který za všech okolností predikuje tu kategorii, jež je v datech zastoupena nejčastěji. V našem případě tvoří závisle proměnnou pouze dvě kategorie: úspěch v kurzu a neúspěch v kurzu. Přitom z celkového počtu 2 321 studentů byla větší část (tj. 1 462 studentů) úspěšná. Baseline model bude tedy u všech studentů predikovat úspěšnost, což povede ke správné predikci v 62,99 % případů. Za základní úroveň přesnosti je tak možné považovat 63 %.

Při použití logistické regrese jakožto klasifikačního algoritmu a kompletních dat pro účely trénování bylo dosaženo přesnosti 70,6 %, čili o něco málo více jak 70 % případů mělo správně predikovanou úspěšnost studenta na základě použitých proměnných týkajících se časových charakteristik návštěvnosti kurzů. Došlo tedy ke zlepšení predikce o 7,6 %. Jako cut-off skóre byla zvolena hodnota 0,563, která byla určena pomocí vážené varianty Youdenova indexu (Youden, 1950). Youdenův index funguje v základní variantě tak, že jako optimální cut-off skóre volí takovou hodnotu na křivce ROC, jež maximalizuje vzdálenost od diagonální čáry (viz obrázek 8). Vážená varianta Youdenova indexu zohledňovala poměr obou kategorií v závisle proměnné a byla zvolena proto, že přinášela mírně větší přesnost predikce než nevážená varianta.

Křivka ROC: logistická regrese – kompletní data



Křivka ROC a zvolené cut-off skóre



Obrázek 8: Křivka ROC při použití logistické regrese jako klasifikačního algoritmu a kompletních dat pro účely trénování (vlevo) včetně vyznačeného cut-off skóre (vpravo)

Na základě zvoleného cut-off skóre byla v tomto případě výsledná hodnota senzitivity 84,2 % a specificity 47,4%. Z toho lze usoudit, že tento logistický model je relativně úspěšný v predikci úspěšných studentů (cca 4 z 5 úspěšných studentů predikuje správně), ale relativně slabý v predikci neúspěšných studentů (každého druhého neúspěšného studenta predikuje nesprávně). Při volbě jiného cut-off skóre by bylo možné vylepšit správnost predikce neúspěšných studentů, ovšem pouze při snížení celkové úspěšnosti predikce (tj. snížila by se schopnost predikovat úspěšné studenty)⁶⁶.

Při použití algoritmu typu rozhodovací strom byla opět použita kompletní data pro účely trénování, ovšem zároveň bylo využito tzv. prořezávání (*pruning*) a 10násobná křížová validace pro výběr finálního stromu. Konkrétně byl nejprve vytvořen tzv. úplný strom (*complete tree*), při jehož tvorbě nebyl parametr složitosti nijak omezen (tj. byl roven nule). Tento strom byl samozřejmě příliš velký (celkem 86 větví), takže jsem přistoupil k jeho prořezávání, tzn. omezování počtu větví, a tím pádem zjednodušení stromu. V rámci prořezávání stromu je za použití 10násobné křížové validace vypočítávána míra tzv. klasifikační chyby (*cross-validated classification error*). Následně je vybrán strom s co nejméně větvemi a s co nejnižší klasifikační chybou⁶⁷. Při použití kompletních dat vedl tento postup k výběru rozhodovacího stromu s celkem 11 větvemi, který je zobrazen na obrázku níže (viz obrázek 9).

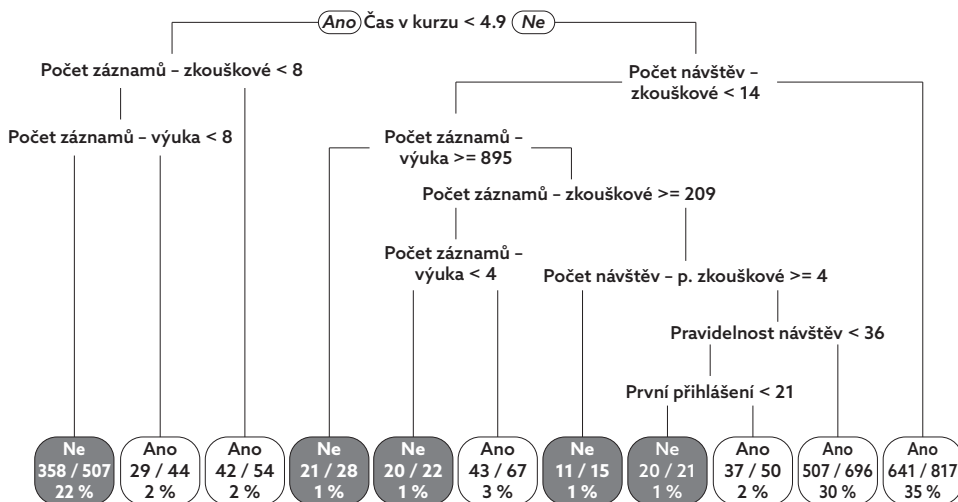
Výhodou rozhodovacích stromů je jejich velmi jednoduchá interpretovatelnost. Na základě rozhodovacího stromu na obrázku 9 tak můžeme velmi jednoduše vidět, podle jakých kritérií probíhá rozhodování o tom, jestli bude určitý student na základě predikce označen za úspěšného či neúspěšného. Zároveň je prezentována informace o poměru správně klasifikovaných případů k celkovému počtu případů v dané větvi a o celkovém procentu případů tvořících danou větev v rámci celého stromu. Z obrázku tak vidíme, že první rozhodovací kritérium v rámci celého stromu se týká celkového času stráveného v kurzu, přičemž je stanoveno kritérium „menší než 4,9 minut“. Pokud se jako příklad vezmou studenti, kteří v e-learningové podpoře kurzu strávili méně než 4,9 minut, pak je druhým rozhodovacím kritériem to, zda má student v průběhu zkuškového období méně než 8 záznamů o aktivitě v kurzu. Pokud nikoli (tj. má ve zkuškovém období více jak 8 záznamů), pak je predikován jako úspěšný. Úspěšnost klasifikace je přitom v této větvi 42/54 (tj. 77,8 %). Do této větve však spadají pouze 2% celkového počtu studentů v analyzovaných datech. Zároveň se nabízí poměrně přímočará interpretace toho, o jaké studenty (s ohledem na jejich studium v blended learningovém kurzu) se jed-

66 Doplníme, že z praktického či aplikačního hlediska může být skutečně vhodné zvýšit specificitu i za cenu celkově horší přesnosti. Pokud by totiž účelem použitím tohoto modelu byla identifikace rizikových studentů, kterým hrozí neúspěch v kurzu (viz kapitolu 6.2), pak by pro nás bylo důležitější správně identifikovat co nejvíce těchto studentů a pomoci jim kurz dokončit, a to i za cenu toho, že by častěji došlo k označení nerizikového studenta za rizikového.

67 Pro podrobný a přesný popis, jak je vypočítávána klasifikační chyba na základě 10násobné křížové validace viz Therneau & Atkinson (2018).

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech

Vizualizace rozhodovacího stromu – kompletní data



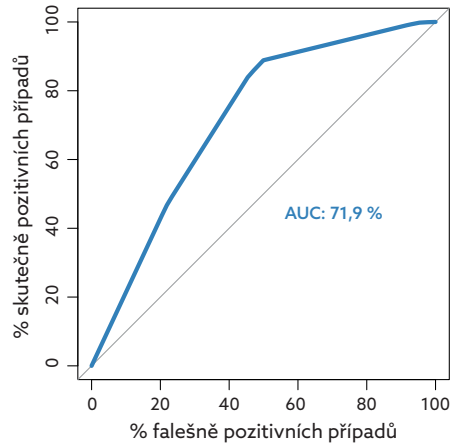
Obrázek 9: Vytvořený rozhodovací strom při použití kompletních dat

ná. Jsou to studenti, kteří v online podpoře strávili v průběhu semestru minimální množství času, ale zároveň ve zkouškovém období projevili alespoň nějakou aktivitu v kurzu (např. odevzdání závěrečného úkolu). Obdobným způsobem pak lze analyzovat i ostatní větve rozhodovacího stromu.

Co se týče predikčních schopností výše znázorněného rozhodovacího stromu při použití kompletních dat pro účely trénování, pohybuje se přesnost predikce na úrovni 74,5%, což je zlepšení predikce o 11,5% oproti výchozímu modelu a o 3,9% oproti logistickému modelu. Senzitivita rozhodovacího stromu je pak na úrovni 88,9% a specificita na úrovni 50,1%. V obou případech je zde patrné zlepšení oproti logistickému modelu, ale stejně jako logistický model není ani rozhodovací strom schopen příliš přesně predikovat neúspěšné studenty (ovšem ty úspěšné předikuje téměř s devadesátiprocentní úspěšností). Obrázek 10 pro srovnání zobrazuje křivku ROC daného rozhodovacího stromu a zároveň hodnotu plochy pod křivkou (AUC).

Algoritmus náhodný les funguje tak, že vytváří větší množství rozhodovacích stromů, přičemž jednotlivé vytvořené stromy nejsou trénovány na celém datovém souboru, nýbrž jen na určité množině náhodně vybraných případů. Přestože tedy v případě náhodného lesa byla pro účely trénování použita kompletní data, ve skutečnosti technika opakovaného náhodného výběru pro tvorbu každého jednotlivého stromu dělá v podstatě to, že je kompletní datová sada rozdělena na řadu menších trénovacích a testovacích množin. Při predikci úspěchu či neúspěchu studentů v blended learningových kurzech pomocí náhodného lesa bylo určeno, aby

Křivka ROC: rozhodovací strom – kompletní data



Obrazek 10: Křivka ROC při použití rozhodovacího stromu jako klasifikačního algoritmu a kompletních dat pro účely trénování algoritmu

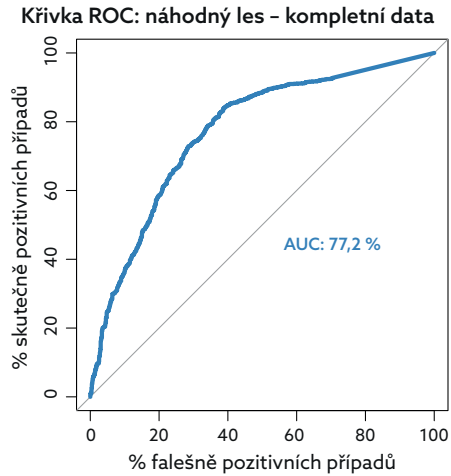
bylo vytvořeno celkem 500 dílčích rozhodovacích stromů. Důvodem bylo to, aby vytvořených stromů bylo dostatečné množství, což také bylo ověřeno⁶⁸. Následně byla hledána optimální hodnota parametru *mtry*, který specifikuje počet proměnných náhodně vybraných pro účely větvení v rámci jednotlivých rozhodovacích stromů. Optimální hodnota parametru *mtry* byla hledána tak, že byly zkoušeny různé hodnoty a sledována tzv. OOB chyba (*out-of-bag error*), již lze chápat jako testovací chybu pro celý model. Jako finální hodnota parametru *mtry* byla vybrána hodnota 3, jelikož vedla k nejnižší hodnotě OOB chyby (konkrétně 0,245).

Predikční schopnosti finálního modelu vytvořeného pomocí náhodného lesa při použití kompletních dat byly následně posouzeny stejným způsobem jako ostatní algoritmy. Přesnost predikce se pohybuje na úrovni 75,5 %, což odpovídá jen mírnému zlepšení (1 %) oproti rozhodovacímu stromu. Senzitivita se pak v případě náhodného lesa pohybuje na úrovni 85,4 % a specificita na úrovni 58,7 %. Ve srovnání s rozhodovacím stromem tak využití náhodného lesa vedlo sice k o něco nižší senzitivitě (o 3,5 %), ovšem zároveň se poměrně výrazně zvýšila specificita (o 8,6 %). Rovněž je třeba doplnit, že při použití náhodného lesa vždy do výpočtu vstupuje prvek náhody, což ve finále vede k tomu, že při každém novém výpočtu jsou výsledné hodnoty přesnosti, senzitivity, specificity apod. mírně odlišné.

Posledním algoritmem, který byl v rámci srovnávání použit, byly tzv. podpůrné vektory (SVM). V případě podpůrných vektorů byla hledána optimální hodnota

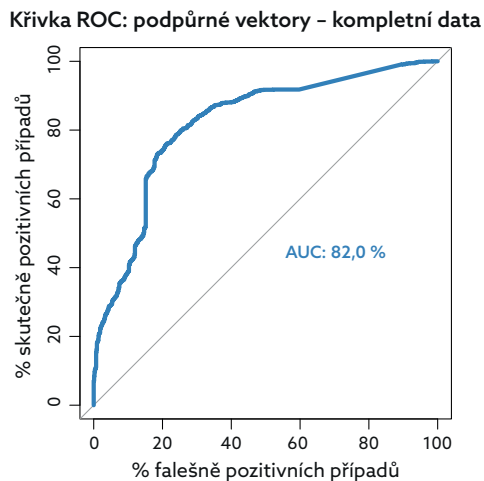
⁶⁸ Za dostatečné množství stromů bylo považováno takové, kdy přidávání dalších stromů již nevedlo k lepšímu výsledku (tj. přesnější predikci).

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech



Obrázek 11: Křivka ROC při použití náhodného lesa jako klasifikačního algoritmu a kompletních dat pro účely trénování

parametrů *cost* a *epsilon*, přičemž pro hledání jejich optimální hodnoty jsem využil techniku vyhledávání v mřížce. Tato technika funguje tak, že vytvoří „mřížku“ z možných hodnot obou parametrů a následně vytváří model pro každou z možností na mřížce. Z vytvořených modelů je pak vybrán ten, který vede k nejlepším výsledkům. Jak je z tohoto popisu zřejmě patrné, technika prohledávání mřížky nemusí vždy vést k optimálnímu řešení, jelikož prohledávaná mřížka může být



Obrázek 12: Křivka ROC při použití podpůrných vektorů jako klasifikačního algoritmu a kompletních dat pro účely trénování algoritmu

příliš „hrubá“ a optimální bod tak na ní nebude ležet. Z toho důvodu bylo při použití podpurných vektorů pracováno s dostatečně „jemnou“ mířížkou. Při testování modelů v průběhu hledání a výběru optimálních parametrů byla využita metoda jednoduchého rozdělení datové sady na trénovací a validační množinu v poměru dvě třetiny na trénink a třetina na validaci.

Co se týče predikčních schopností podpurných vektorů při použití kompletních dat, vedl výše popsany postup k nejlepším výsledkům. V případě přesnosti bylo dosaženo hodnoty 78,5 %, což je nejvyšší hodnota ze všech srovnávaných algoritmů. Oproti náhodnému lesu došlo ke zlepšení o 3 %, oproti baseline modelu pak dokonce o 15,5%. Senzitivita byla v případě podpurných vektorů druhá nejvyšší, konkrétně 87,8 %, zároveň však byla dosažena nejvyšší hodnota specifity (62,7 %). Podpurné vektory si tak ve srovnání predikčních schopností vedly nejen celkově nejlépe, ale zároveň v případě správnosti predikce neúspěšných studentů dosahovaly nejlepších výsledků.

Na závěr této části je pro přehlednost uvedena tabulka (viz tabulka 7) uvádějící výsledky přesnosti, senzitivity, specifity a plochy pod křivkou u všech srovnávaných klasifikátorů. Jako jeden ze základních či obecných indikátorů kvality modelu se obvykle uvádí plocha pod křivkou. Podle tabulky níže je tak patrné, že logistická regrese vedla ze všech klasifikátorů k celkově nejhoršímu modelu, zatímco podpurné vektory k modelu nejlepšímu. Stejně je tomu i v případě přesnosti a specifity. Jiná situace je pouze u senzitivity, kde dosáhl nejlepšího výsledku klasifikátor typu rozhodovací strom.

Tabulka 7: Srovnání přesnosti, senzitivity, specifity a plochy pod křivkou u analyzovaných klasifikačních algoritmů při použití kompletních dat a výchozí metody validace

	Logistická regrese	Rozhodovací strom	Náhodný les	Podpurné vektory
Přesnost	70,6	74,5	75,5	78,5
Senzitivita	84,2	88,9	85,4	87,8
Specifita	47,4	50,1	58,7	62,7
Plocha pod křivkou (AUC)	69,8	71,9	77,2	82

7.4.2 Hodnocení přínosnosti použitých prediktorů

V této části budu věnovat pozornost druhé specifické výzkumné otázce, zaměřující se na to, jaké proměnné týkající se temporálních aspektů návštěvnosti blended learningových kurzů jsou nejdůležitější při predikci úspěšnosti studentů, resp. které

prediktory nejvíce přispívají ke správné predikci úspěšnosti studentů. Problematika identifikace a výběru důležitých proměnných (*variable importance* či *feature selection*) je v kontextu data miningu a strojového učení řešena poměrně intenzivně již delší dobu a je k dispozici řada různých způsobů identifikace důležitých proměnných (Guyon & Elisseeff, 2003). Hlavním důvodem pro identifikaci a výběr důležitých proměnných je obvykle nutnost vybírat ze stovek či tisíců potenciálních proměnných jen určité množství prediktorů (mimo jiné z důvodu šetření časových či výpočetních nákladů). Cílem je samozřejmě vybrat ty proměnné, které mají největší prediktivní sílu. V některých případech může vybírání jen určitých proměnných sloužit rovněž jako prostředek pro zlepšování predikční výkonnosti použitých algoritmů. V kontextu této analýzy je hlavním cílem pro zjišťování přínosnosti jednotlivých prediktorů lepší porozumění tomu, jaké temporální charakteristiky návštěvnosti blended learningových kurzů jsou z hlediska úspěšnosti studentů v kurzech ty nejdůležitější.

V návaznosti na předchozí část a celkové zaměření analýzy byla přítomnost jednotlivých prediktorů zjišťována v souvislosti se všemi čtyřmi srovnávanými klasifikačními algoritmy. Vzhledem k tomu, že každý ze srovnávaných klasifikačních algoritmů funguje na jiném principu, byla i pro kvantifikaci důležitosti jednotlivých proměnných využita u každého z klasifikátorů jiná technika. V případě logistické regrese byla jako indikátor důležitosti jednotlivých proměnných využita absolutní hodnota t-statistiky pro každý z použitých prediktorů. U rozhodovacího stromu je použita specifická metrika důležitosti proměnných, která vychází z toho, nakolik se proměnná v rozhodovacím stromu objevuje v rámci jednotlivých větvení (tj. jak často proměnná v rámci stromu slouží jako kritérium rozdělení) a nakolik svou přítomností v rozhodovacím stromu přispívá k celkovému zlepšení modelu. V případě náhodného lesa je pro výpočet důležitosti proměnných použit tzv. Giniho koeficient, konkrétně je jako indikátor důležitosti proměnné použit průměrný pokles Giniho koeficientu (*mean decrease in Gini coefficient*), který lze chápat jako míru toho, nakolik každá proměnná přispívá k homogenitě větví a listů ve výsledném náhodném lese (tj. v průměru napříč všemi vytvořenými stromy). V případě podpůrných vektorů zatím neexistuje široce používaný a v analytických softwarech běžně implementovaný způsob výpočtu důležitosti proměnných. Proto byl využit výpočet zmiňovaný Guyonovou a Elisseeffem (2003), jenž pracuje s váhou jednotlivých podpůrných vektorů.

Tabulka 8 uvádí vypočtené míry důležitosti pro jednotlivé proměnné v rámci jednotlivých klasifikačních algoritmů. Přitom se vychází z modelů, které byly prezentovány v předchozí části, tj. z modelů trénovaných na kompletních datech. Vedle absolutních hodnot jednotlivých metrik důležitosti proměnných je vždy v závorce uvedena i relativní hodnota důležitosti dané proměnné. Relativní hodnotu důležitosti proměnné lze tedy chápat jako míru toho, nakolik daná proměnná v porovnání s ostatními přispívá v rámci daného klasifikátoru ke správné predikci úspěšnosti studentů v blended learningových kurzech. Jednotlivé řádky v tabulce jsou přítom

seřazeny podle důležitosti proměnných v rámci prvního klasifikátoru, tj. logistické regrese.

Z tabulky lze vidět, že při využití logistické regrese jsou nejdůležitějšími proměnnými první návštěva v kurzu a pravidelnost návštěv. Dále pak počet záznamů o aktivitě studenta v průběhu výuky a v průběhu zkouškového období a celkový čas strávený v kurzu. Zjednodušeně to můžeme interpretovat tak, že při využití logistické regrese je nejdůležitějším prediktorem úspěšnosti v kurzu to, jak brzo po začátku kurzu se student přihlásí do e-learningové podpory kurzu a jak pravidelně se do e-learningové podpory přihlašuje. Dále je pro predikci úspěšnosti studentů relativně důležité to, jak moc je student v kurzu aktivní a kolik celkového času v kurzu stráví. Naopak počty návštěv se při použití logistické regrese nejeví jako významné prediktory. Důležitost proměnných při použití rozhodovacího stromu a náhodného lesa se od logistické regrese mírně liší. Především zde má mnohem větší důležitost celkový čas strávený studentem v kurzu (v obou případech jde o nejdůležitější proměnnou) a také počet návštěv studenta v kurzu v průběhu výuky. Výrazně jiná situace je pak při použití podpůrných vektorů jakožto klasifikátoru úspěšnosti studentů. Zde patří mezi nejdůležitější proměnné počty návštěv studenta v kurzu a počty záznamů o aktivitě studenta v kurzu, a to jak v průběhu výuky, tak i v průběhu zkouškového období.

Tabulka 8: Důležitost jednotlivých proměnných v modelu u analyzovaných klasifikačních algoritmů při použití kompletních dat⁶⁹

	Logistická regrese	Rozhodovací strom	Náhodný les	Podpůrné vektory
První návštěva kurzu	5,1 (21)	95,7 (16)	148,3 (17)	0,3 (0)
Pravidelnost návštěv	5 (21)	123 (20)	140,9 (16)	77 (7)
Počet záznamů – výuka	3,8 (16)	117,1 (19)	153,5 (17)	146 (13)
Počet záznamů – zkouškové	3,4 (14)	27,4 (5)	103,4 (12)	239,8 (22)
Celkový čas v kurzu	2,8 (12)	127,9 (21)	153,8 (18)	79,5 (7)
Počet záznamů – p. zkouškové	1,4 (6)	3,1 (1)	12,6 (1)	5,7 (1)
Počet návštěv – výuka	1 (4)	98,4 (16)	96,8 (11)	348,6 (32)
Počet návštěv – zkouškové	0,8 (3)	8,5 (1)	61,7 (7)	164,6 (15)
Počet návštěv – p. zkouškové	0,6 (3)	5,8 (1)	10,3 (1)	24,5 (2)

Z tabulky výše je patrné, že důležitost jednotlivých prediktorů se poměrně zásadně liší v závislosti na tom, jaký klasifikační algoritmus je použit. Zdá se tedy, že

⁶⁹ Ve všech případech platí, čím větší číslo, tím větší důležitost. V závorkách jsou uvedeny relativní hodnoty důležitosti

otázku po důležitosti jednotlivých prediktorů nelze smysluplně zodpovědět, aniž by byl zároveň zohledněn fakt, že různé klasifikační algoritmy mohou pro účely predikce preferovat odlišené prediktory.

7.4.3 Predikční testování externí validací

Poslední část sekce výsledků se zaměřuje na třetí specifickou výzkumnou otázku a na testování predikčních schopností vybraných klasifikačních algoritmů za využití externí validace. Jak již bylo nastíněno v rámci postupu analýzy dat (viz kapitolu 7.3.4), predikční testování externí validací bylo provedeno dvakrát, přičemž poprvé probíhalo rozdělení dat na trénovací a testovací množinu náhodně⁷⁰, zatímco podruhé byla při náhodném výběru navíc reflektována úroveň kurzu. Všichni studenti z jednoho kurzu tak museli být buď v trénovací, nebo v testovací množině, ale nikoli v obou (tj. nemohlo se stát, aby část studentů jednoho kurzu byla v trénovací množině a část v testovací množině). Tento způsob testování predikčních schopností vybraných klasifikátorů pomocí externí validace mi v prvním kroku umožnil ověřit, nakolik jsou predikční schopnosti klasifikátorů prezentované v části 7.4.1 platné i při testování modelů na zcela nových případech. V druhém kroku pak testování externí validací umožní zjistit, nakolik jsou analyzované klasifikátory schopny natrénovat se na datech z určité sady kurzů a následně správně klasifikovat úspěšnost studentů ve zcela jiných kurzech.

Tabulka 9 prezentuje výsledky predikčního testování externí validací za všechny čtyři analyzované algoritmy. Připomeňme, že šlo o externí validaci se 100 opakováními, tedy, že celkem stokrát byla kompletní datová sada rozdělena na trénovací a testovací část v poměru dvě třetiny na trénink a třetina na testování, a následně bylo výše popsaným postupem vytvořeno 100 různých modelů, které byly vyhodnoceny skrze přesnost, sensitivitu, specificitu a plochu pod křivkou. Z této stovky modelů pak byl v rámci každého klasifikačního algoritmu spočítán průměr a směrodatná odchylka pro každou z hodnotících metrik, jak je patrné z tabulky níže. Při srovnání s tabulkou 8 je zřejmé, že predikční schopnosti logistické regrese a náhodného lesa se výrazně nezměnily. V případě logistické regrese došlo ke snížení přesnosti o 0,5 % (AUC rovněž o 0,5 %) a v případě náhodného lesa byla snížena přesnost o 1,4 % (AUC o 1,4 %). U rozhodovacího stromu je již rozdíl výraznější, jelikož přesnost klesla o 2,9 % a AUC o 4,1 %. Snížení predikčních schopností u podpůrných vektorů již lze považovat za zásadní. Přesnost je zde nižší o 6,5 % a AUC o 9,3 %. Z toho lze usoudit, že při využití podpůrných vektorů pro vytvoření modelu prezentovaného v části 7.4.1 došlo k výraznému nadhodnocení predikčních schopností. I přesto si

70 Ovšem při zachování proporčního zastoupení úspěšných a neúspěšných studentů (tj. závisle proměnné) v trénovací i testovací množině.

ale podpůrné vektory (a ostatně i zbývající klasifikátory) udržely relativně dobré predikční schopnosti při srovnání s výchozím modelem.

Tabulka 9: Srovnání přesnosti, senzitivity, specificity a plochy pod křivkou u analyzovaných klasifikačních algoritmů při použití externí validace se 100 opakováními (bez zohlednění úrovně kurzů)

	Logistická regrese		Rozhodovací strom		Náhodný les		Podpůrné vektory	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Přesnost	70,1	1,4	71,6	1,4	74,1	1,3	72	1,5
Senzitivita	83,1	2,1	85,8	2,7	84,5	1,5	83,8	1,9
Specifická	48	3,6	47,2	5,1	56,5	2,8	51,8	4,3
Plocha pod křivkou (AUC)	69,3	1,8	67,8	2,5	75,8	1,3	72,7	1,6

Tabulka 10 prezentuje výsledky predikčního testování externí validací při zohlednění úrovně kurzů. Zde je již na první pohled zřejmé, že situace je výrazně jiná a že všechny klasifikátory mají výrazně horší predikční schopnosti. V případě rozhodovacího stromu, náhodného lesa a podpůrných vektorů je pak přesnost dokonce horší než u baseline modelu, což v praxi znamená naprostou nepoužitelnost takových modelů. Logistická regrese si sice udržela vyšší přesnost než baseline model, ale pouze o 3,1 %, což nelze považovat za výrazné zlepšení. Z těchto výsledků tak lze usoudit, že s výjimkou logistické regrese mají analyzované klasifikační algoritmy výrazný problém predikovat úspěšnost studentů ve zcela nových kurzech (tj. kurzech, které nebyly součástí trénovací množiny).

Výrazně větší směrodatné odchylky oproti předchozí tabulce zároveň znamenají, že při zohlednění úrovně kurzů dochází k tvorbě výrazně různorodějších

Tabulka 10: Srovnání přesnosti, senzitivity, specificity a plochy pod křivkou u analyzovaných klasifikačních algoritmů při použití externí validace se 100 opakováními zohledňující úroveň kurzů

	Logistická regrese		Rozhodovací strom		Náhodný les		Podpůrné vektory	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Přesnost	66,1	5,4	61,6	6,2	59,2	5	61,3	5,7
Senzitivita	81,7	7,1	72,2	11,7	67,2	9,2	71,3	9,4
Specifická	40,1	9,4	46,2	9	48	7,5	46,8	9,2
Plocha pod křivkou (AUC)	60,4	7,6	56,1	6,4	57,9	5,5	57,6	6,1

7 Analýza 1: Predikce úspěšnosti studentů v blended learningových kurzech

modelů, co se týče jejich predikčních schopností. To znamená, že z vytvořené stovky modelů sice některé dosahovaly relativně dobrých výsledků, ale některé (většina) naopak měly výrazně špatné výsledky. To naznačuje, že z hlediska dvouúrovňového charakteru dat dochází k tzv. přeučování modelů (*overfitting*), kdy jsou klasifikační algoritmy příliš přizpůsobeny kurzům v trénovací sadě dat, a nejsou tak schopny dostatečně zobecňovat a predikovat úspěšnost studentů v kurzech v testovací množině. Relativní úspěšnost logistické regrese při predikčním testování externí validací zohledňující úroveň kurzů je pak zřejmě dána tím, že jde o poměrně jednoduchý klasifikátor, který je tím pádem odolný vůči přeučení.

7.5 Diskuze a závěr

Cílem této analýzy bylo zhodnotit, jaké jsou možnosti využití vybraných klasifikačních algoritmů pro účely predikce úspěšnosti studentů blended learningových kurzů, a to pouze na základě vybraných temporálních ukazatelů návštěvnosti studentů v LMS. Prozaicky řečeno: Nakolik lze vůbec predikovat úspěšnost v kontextu blended learningu, při kterém se pouze část výuky odehrává v LMS? Nakolik lze predikovat úspěšnost studentů pouze na základě toho, jak navštěvují kurz v LMS? A jak si při predikci vedou různé klasifikační algoritmy?

Oproti dosavadním výzkumům v této oblasti přitom bylo pro účely analýz využito výrazně větší množství různých blended learningových kurzů, a tudíž bylo možné věnovat pozornost rovněž otázce zobecnitelnosti a přenositelnosti prediktivních modelů v kontextu blended learningu. Výsledky analýzy tak přináší důležité poznatky k současným možnostem predikce úspěšnosti studentů v blended learningu za využití obecných prediktivních modelů.

7.5.1 Shrnutí odpovědí na stanovené otázky

První specifická výzkumná otázka se zaměřovala na základní posouzení predikčních schopností zvolených klasifikačních algoritmů (logistická regrese, rozhodovací strom, náhodný les, podpůrné vektory), přičemž toto posouzení proběhlo za využití dat ze všech analyzovaných kurzů. Nejlepšího výsledku dosáhl algoritmus podpůrných vektorů, který byl schopen v 78,5% případů správně predikovat úspěšnost či neúspěšnost studenta v blended learningovém kurzu na základě temporálních charakteristik návštěvnosti. To je o 15,5% více než v případě tzv. baseline modelu, což lze považovat za výrazné zlepšení.⁷¹

⁷¹ Je třeba si uvědomit, že v této analýze je záměrně omezen počet prediktorů. V případě praktické aplikace by samozřejmě došlo k využití většího množství prediktorů, a tudíž by se s největší pravděpodobností dospělo k ještě vyšší přesnosti v predikci úspěšnosti studentů.

Nejhoršího výsledku bylo dosaženo za využití logistické regrese (přesnost 70,6 %). U všech klasifikačních algoritmů však docházelo k tomu, že byly mnohem lepší v predikci úspěšných studentů než v predikci neúspěšných studentů. Například algoritmus podpůrných vektorů byl schopen správně predikovat 87,8% úspěšných studentů, ale pouze 62,7% neúspěšných studentů. V případě logistické regrese pak dokonce docházelo ke správné predikci neúspěšných studentů pouze v 47,4% případech. To naznačuje, že v případě praktické aplikace specificky pro účely identifikace rizikových studentů je třeba počítat s výrazně nižší mírou úspěšné predikce.

Druhá specifická výzkumná otázka se věnovala tomu, které z použitých prediktorů založených na temporálních aspektech návštěvnosti blended learningových kurzů nejvíce přispívají k predikci úspěšnosti studentů. Výsledky analýzy ukazují, že míra přínosnosti jednotlivých prediktorů se výrazně liší v závislosti na použitých klasifikačních algoritmech. V případě logistické regrese patřilo mezi nejdůležitější faktory to, jak brzo po začátku kurzu se student poprvé přihlásil do kurzu a jak pravidelně poté kurz navštěvoval. Mezi relativně silné prediktory pak spadal rovněž celkový počet záznamů o aktivitě studenta v kurzu v průběhu výuky a v průběhu zkuškového období a také celkový čas strávený prací v kurzu. Naopak počty návštěv v případě logistické regrese nesehrávaly významnější roli. Tak tomu ale nebylo u ostatních klasifikačních algoritmů, kde byl naopak počet návštěv v průběhu výuky relativně důležitým (v případě rozhodovacího stromu a náhodného lesa), či dokonce jednoznačně nejdůležitějším prediktorem (v případě podpůrných vektorů). Algoritmus podpůrných vektorů pak dává celkově velmi odlišný obraz oproti logistické regresi. První přihlášení do kurzu a pravidelnost v navštěvování kurzu sehrává u tohoto algoritmu jen velmi malou roli. Do značné míry jediným výsledkem, na kterém se v tomto ohledu shodují všechny klasifikační algoritmy, je nepřínosnost počtu návštěv a počtu záznamů o aktivitě v průběhu prodlouženého zkuškového období jakožto prediktoru úspěšnosti. Zřejmě tedy aktivita studenta v průběhu prodlouženého zkuškového období obvykle na výsledném hodnocení už nic nemění, anebo je takových případů příliš malé množství na to, aby sehrály při predikci úspěšnosti významnější roli.

Nutno zároveň zdůraznit, že výše uvedené výsledky analýzy související s druhou výzkumnou otázkou otevírají nové a, jak se domnívám, poměrně zásadní téma. Dosavadní studie, které v souvislosti s predikcí úspěšnosti studentů věnovaly pozornost srovnávání různých klasifikačních algoritmů (např. Romero, Espejo, Zafra, Romero, & Ventura, 2013; Romero, López, Luna, & Ventura, 2013; Sheshadri, Gitinabard, Lynch, Barnes, & Heckman, 2018; Zafra, Romero, & Ventura, 2011), se zaměřovaly primárně na srovnávání klasifikačních algoritmů z hlediska jejich predikčních schopností, resp. z hlediska míry úspěšnosti predikce, které jsou schopny dosáhnout. Zatím však nebyl akcentován fakt, že se jednotlivé klasifikační algoritmy neliší pouze tím, nakolik jsou schopny správně predikovat úspěšnost studentů, ale také tím, jaké k tomu využívají proměnné. To je však zcela zásadní

a věcně možná i důležitější než případné dílčí odlišnosti v prediktivních schopnostech. Znamená to totiž, že otázka po tom, které faktory související s výukou a učením v LMS jsou důležité z hlediska predikce úspěšnosti studentů, se výrazně komplikuje. Klasifikační algoritmy různých typů totiž mohou upřednostňovat zcela odlišné prediktory, a to až do takové míry, že jeden a týž prediktor může být při použití určitého klasifikačního algoritmu vyhodnocen jako zcela nepodstatný, zatímco při využití jiného algoritmu by byl tím nejdůležitějším prediktorem (jako tomu bylo u proměnné První návštěva kurzu a algoritmu podpurných vektorů ve srovnání s logistickou regresí).

V rámci třetí výzkumné otázky jsem přesunul pozornost k problematice přenositelnosti či zobecnitelnosti prediktivních modelů, resp. k tomu, nakolik je možné pro účely predikce úspěšnosti studentů v blended learningových kurzech vytvořit obecný prediktivní model, jenž by byl schopen predikce napříč různými kurzy. Pro zodpovězení této otázky bylo využito predikční testování externí validací a bylo hodnoceno, nakolik se tím změnila míra úspěšné predikce. Nejprve jsem přistoupil k predikčnímu testování externí validací bez zohledňování úrovně kurzů. U většiny algoritmů došlo v tomto kroku spíše k nepatrnému či malému zhoršení predikčních schopností. Ovšem v případě algoritmu podpurných vektorů již došlo k relativně výraznému zhoršení (o 6,5 % nižší přesnost, o 9,3 % nižší AUC), což znamená, že při základním posuzování predikčních vlastností v rámci první výzkumné otázky došlo u podpurných vektorů k nadhodnocení predikčních schopností.

V druhém kroku jsem pak přistoupil k predikčnímu testování externí validací při zohlednění úrovně kurzů, takže klasifikační algoritmy byly trénovány na jiné množině kurzů, než na které byly následně testovány predikční schopnosti výsledných klasifikačních modelů. To dovoluje zhodnotit, nakolik jsou jednotlivé klasifikační algoritmy schopné dospět k dostatečně obecnému modelu, který by fungoval napříč různými blended learningovými kurzy. Výsledky tohoto predikčního testování přitom odhalily neschopnost většiny klasifikačních algoritmů dospět k dostatečně obecnému modelu. Algoritmy rozhodovacího stromu, náhodného lesa i podpurných vektorů dospěly v přesnosti predikce k horším výsledkům, než tzv. baseline model. Jen logistická regrese dospěla k vyšší přesnosti než baseline model, ovšem pouze o 3,1 %, což zřejmě bude dáno relativní jednoduchostí logistické regrese v porovnání s ostatními algoritmy. Jak uvádí Baker a Inventado (2014), konzervativní algoritmy jako logistická regrese mohou ve vzdělávacím kontextu dosahovat lepších výsledků díky poměrně velké míře šumu v analyzovaných datech. V rámci predikčního testování externí validací zohledňující úroveň kurzů tak zřejmě došlo k tomu, že pokročilejší algoritmy nebyly schopny v trénovacích datech dostatečně odlišit signál od šumu, resp. nebyly schopny odlišit obecné charakteristiky související s úspěšností studentů v kurzech, a tudíž nebyly schopny generovat dostatečně obecný model, který by byl funkční i na jiných kurzech obsažených v testovací množině.

7.5.2 Důsledky pro praxi a další výzkum

Z výše uvedených výsledků plyne několik zásadních důsledků jak pro praxi, tak pro další výzkum v oblasti analytiky učení a data miningu ve vzdělávání. Předně se ukazuje, že je na místě určitá skepse, co se týče dosavadních prediktivních řešení a obecně možností predikce úspěšnosti studentů v blended learningových kurzech. Přestože je oblast predikce úspěšnosti a identifikace rizikových studentů jednou z nejvíce akcentovaných výzkumných oblastí data miningu ve vzdělávání i analytiky učení, výsledky této analýzy naznačují, že přinejmenším v kontextu blended learningových kurzů probíhajících v systémech typu LMS jsme zatím stále poměrně daleko od prakticky využitelného obecného prediktivního systému, který by byl schopen s dostatečnou mírou přesnosti predikovat úspěšnost studentů či včas identifikovat studenty s rizikem nedokončení napříč různými kurzy. Je otázkou, nakolik lze souhlasit s názvem studie Gaševiče et al. (2016), který naznačuje, že snahy o obecný prediktivní model možná nemusí být tím nejvhodnějším a nejprínosnějším cílem. Rozhodně lze však souhlasit s výzvou k opatrnosti, kterou uvedení autoři směřují ke vzdělávacím institucím. Investice do komerčních řešení nabízejících fungující a zároveň obecný systém pro predikci úspěšnosti v LMS zřejmě prozatím nebude dobrou volbou. Nadále také platí výzva, o které hovořili již Gaševič, Dawson a Siemens (2015). Ti považovali za naprosto nezbytné zaměřit se na realizaci evaluačních studií, které by využívaly data z širokého spektra různých kurzů a dovolovaly tak získat lepší představu o zobecnitelnosti a přenositelnosti jednotlivých řešení a modelů.

Z dosavadních výsledků výzkumu i z výsledků této analýzy je zároveň patrných několik možností rozvoje, kterým stojí za to věnovat pozornost v nadcházejícím výzkumu. Základním problémem či otázkou je to, jakým způsobem při modelování zohlednit výrazné odlišnosti mezi různými kurzy. Jednou z možných cest by bylo neusilovat o obecný prediktivní model, nýbrž vytvářet specifické modely pro každý kurz zvlášť. Tedy usilovat pouze o predikci úspěšnosti studentů v určitém kurzu na základě dat z dřívějších běhů toho stejného kurzu. Ani zde však zatím není zřejmé, nakolik by byly takové dílčí modely úspěšné. Navíc by to samozřejmě znemožňovalo využít prediktivní systém pro zcela nový kurz (tj. pro takový, pro který ještě neexistují data z předchozích běhů). Je také otázkou, jakým způsobem by do takového řešení vstupovaly průběžné úpravy kurzu či změny ve způsobu jeho vedení. Lze totiž předpokládat, že jakákoli výraznější změna kurzu by vedla ke snížení prediktivních schopností modelu, jenž byl vytvořen na datech z předchozích běhů kurzu.

Druhou z možných cest je pokračovat ve snahách o obecný prediktivní model. V takovém případě pak musí výzkumné snahy směřovat k tomu, jak při modelování zachytit a zužitkovat informace na úrovni kurzů. Prvním krokem by mohlo být využití proměnných, které by byly agregované na úrovni kurzů (tj. např. průměrný

počet návštěv studentů v kurzu, průměrný čas trávený studenty v daném kurzu apod.). Případně využít jiné typy prediktorů na úrovni kurzů a jiné prediktory naopak na úrovni jednotlivých studentů. Např. použít četnosti využívání různých typů nástrojů v LMS jakožto prediktory na úrovni kurzů (což by mohlo zachytit odlišnosti mezi různými typy kurzů), zatímco na úrovni studentů pracovat s temporálními charakteristikami návštěvnosti. Jiným přístupem, který by mohl vést k lepším predikčním schopnostem, by byla kombinace shlukování a klasifikace ve smyslu tzv. *cluster-then-predict* přístupu. V takovém případě je na datech nejprve aplikováno shlukování, jehož výsledky se stanou součástí dat, na kterých je následně realizována klasifikace. Takový přístup by mohl napomoci k identifikaci různých shluků či skupin vzájemně podobných kurzů, v jejichž rámci by tak mohly platit obdobné vztahy mezi použitými prediktory a úspěšností studentů. V neposlední řadě se nabízí využití ještě pokročilejších klasifikačních algoritmů, jako jsou např. umělé neuronové sítě (viz např. Monllaó Olivé, Huynh, Reynolds, Dougiamas, & Wiese, 2019), případně rozvíjení existujících klasifikačních algoritmů takovým způsobem, aby byly schopny lépe pracovat s daty víceúrovňové povahy (viz např. Hajjem, Bellavance, & Larocque, 2014; Lin & Luo, 2019; Speiser et al., 2018).