



# Hodnocení vedoucího závěrečné práce

**Student:** Bc. Jakub Štercl  
**Vedoucí práce:** Ing. Mgr. Jan Romportl, Ph.D.  
**Název práce:** Interpretabilita modelů strojového učení  
**Obor:** Webové a softwarové inženýrství

**Datum vytvoření:** 23. 1. 2020

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – následující škálou 1 až 4:</b>
<b>1. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.	
<b>Komentář:</b> Předložená DP plně dosahuje zadaných cílů. Některé body zadání, např. (3), dokonce značně přesahuje.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>2. Písemná část práce</b>	<b>95 (A)</b>
<b>Popis kritéria:</b> Zhodnotte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnotte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnotte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	
<b>Komentář:</b> Práce je velmi dobře strukturována a je psána čtivě, srozumitelně a přehledně. Jednotlivé části působí vyváženě a čtenář může text bez problémů využít jako velmi dobrý úvod a kompaktní shrnutí přístupů k interpretačním metodám ML modelů. Text popisující vlastní přínos diplomanta - jak v oblasti praktické tvorby kódu knihovny v Pythonu, tak v teoreticko-experimentální oblasti analýzy reálného credit scoringového modelu - je velmi srozumitelný a užitečný. Trochu na škodu je poněkud urychlené ukončení poslední kapitoly (jež je velmi zajímavá a důležitá) - nejspíše z důvodů nedostatku času. Typograficky je dílo zpracované velmi kvalitně. Po formální stránce jsou všechny náležitosti též splněny (včetně důsledného citování). Jazykově a stylisticky je vše v pořádku, gramaticky taktéž - s výjimkou několika málo překlepů. V této rovině lze drobně vytknout používání čárek: někde lehce přebývají, jinde se jich čas od času trošku nedostává. Stylisticky text místy vyklouzne do neformálního vyjadřování, které je však v rámci DP v tomto rozsahu akceptovatelné - ostatně na většinu čtenářů může působit příjemněji než nadměrně přehrávaný a neautentický koženě-akademický styl. Dále mohu konstatovat, že nakolik je mi známo, tak diplomant využívá knihovny v souladu s jejich licencemi.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>3. Nepísemná část, přílohy</b>	<b>100 (A)</b>
<b>Popis kritéria:</b> Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů	

#### Komentář:

Diplomant věnoval obrovskou pečlivost zprovoznění a otestování většiny reálně dostupných interpretačních knihoven a algoritmů. Takto vzniklé závěry jsou velmi užitečné. Nad těmito knihovnamy zároveň vytvořil užitečný nástroj, který zjednoduší další interpretační analýzy ML modelů v praxi. Tento nástroj je napsán kvalitně, zdrojový kód působí kultivovaně a je open source zveřejněn na GitHubu. Dokumentace je v podobném duchu zveřejněna na readthedoc.io. Uděluji plný počet bodů zejména proto, že tato solidně zpracovaná SW část je pak zásadně doplněna kvalitní výzkumnou a experimentální prací. Její větší část je zcela veřejně reprodukovatelná (využívá otevřený dataset Titanic); u zbývajících částí postavené na velmi zajímavé interpretaci reálného proprietárního modelu to diplomant vyřešil rozumnou obfuskací modelu i dat.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

#### 4. Hodnocení výsledků, jejich využitelnost

94 (A)

#### Popis kritéria:

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

#### Komentář:

Diplomant velmi dobře a smysluplně volil experimenty, které s danými daty a modely prováděl. Následná interpretace je také užitečná. Je zřejmé, že diplomant sám má bohatou zkušenost s reálným vývojem, nasazením a problémy ML modelů v praxi. Svým směřováním experimentů a interpretací v DP tak přímo adresoval tyto problémy. Obzvláště zajímavá je poslední kapitola (7) věnující se interpretaci reálně fungujícího credit scoringového modelu vyvinutého pro banky nad telco daty. V ní diplomant testuje interpretační schopnosti jednotlivých algoritmů na strukturně známém modelu logistické regrese s WOE. Tím je schopen analyticky ověřovat vytvořené interpretace a jejich správnost. Dochází k užitečným závěrům vyvracejícím poněkud temné hypotézy a scénáře, v nichž by testované interpretační algoritmy selhávaly už při interpretaci lineárního modelu (tj. logit regrese, čili binomického rozdělení, jehož parametry závisí na datech lineárně) s pečlivě zanesenou nelinearitou pomocí WOE binningu. Naplnění takového temného scénáře by asi bylo špatnou zprávou pro interpretabilitu hlubokých nelineárních modelů (např. neuronových sítí). Škoda jen, že tato kapitola se jeví jako poněkud narychlo ukončená - chybí v ní určitá podrobnější syntéza získaných poznatků a jejich extrapolace na další důležité otázky interpretability ML. Za tento nedostatek odeírám několik bodů.

Hodnotící kritérium:

Způsob hodnocení – následující škálou 1 až 5:

#### 5. Aktivita a samostatnost studenta

5a:

**1=výborná aktivita,**  
2=velmi dobrá aktivita,  
3=průměrná aktivita,  
4=slabší, ale ještě dostatečná aktivita,  
5=nedostatečná aktivita

5b:

**1=výborná samostatnost,**  
2=velmi dobrá samostatnost,  
3=průměrná samostatnost,  
4=slabší, ale ještě dostatečná samostatnost,  
5=nedostatečná samostatnost

#### Popis kritéria:

V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posuďte schopnost studenta samostatně tvůrčí práce (5b).

#### Komentář:

Po celou dobu přípravy DP byl diplomant velmi proaktivní a samostatný. Zároveň byl vždy velmi pečlivě připraven na všechny konzultace. Byla radost s ním spolupracovat a líbilo by se mi, kdyby všichni diplomanti, které jsem kdy vedl (a budu vést), fungovali takto. Nepochybuji o tom, že diplomant dále dokáže svoje dílo samostatně rozvíjet, třeba jako open source projekt.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

#### 6. Celkové hodnocení

96 (A)

#### Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

#### Text hodnocení:

Výborně zpracovaná a čtivá diplomová práce s užitečnými výstupy jak v oblasti praktické (příprava SW), tak i v oblastech výzkumně-experimentální (analýza reálných ML modelů) a rešeršní (přehledné shrnutí dostupných metod a nástrojů). Čtenář v ní nalezne jak zajímavé experimentální informace, které mohou usnadnit jeho další práci v oblasti praktické interpretability ML modelů, tak i užitečný SW nástroj.

Podpis vedoucího práce: