



Posudek oponenta závěrečné práce

Student: Filip Kojan
Oponent práce: Ing. Magda Friedjungová
Název práce: Predikce výsledků zápasů v NHL
Obor: Znalostní inženýrství

Datum vytvoření: 9. 6. 2019

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: 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.	
Komentář: Zadání je splněno bez výhrad.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	68 (D)
Popis kritéria: Zhodnoťte 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. Zhodnoťte 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. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	
Komentář: Text je logicky členěn. Teoretická část čtenáře seznamuje s problematikou NHL a dostupnými statistikami. Uvítala bych v ní, kdyby se proces práce dále opíral o nějakou metodiku jako např. CRISP-DM (práce pokrývá všechny kroky, pouze to není zmíněno). Po gramatické stránce se v práci vyskytují drobné chyby, především při použití interpunkce. Dále se vyskytuje nesourodé používání kurzívy. Student dostatečně pracuje se zdroji, nicméně jsou místa, kde zdroje chybí. Např. u statistik Corsi, Fenwick a PDO v sekci 1.2, v sekci 4.1.2 na dané knihovny apod. V sekci 2.1 mi přijde nešťastné použití odkazů na webové stránky v textu. Některé formulace ponechávají zbytečně prostor pro čtenářovu fantazii ("to může napovědět o množství dat" - moc/málo?, "vzhledem k vývoji hokejové hry" - k jakému? apod.). Kapitola 3 Popis dat je dobře zpracovaná. Mezery v názvech skriptů jsou bizarní. U popisu atributů by se hodila např. i distribuce dat a datové typy. Místy jsou argumentace poněkud vágní nebo nesprávné, např. použití více modelů, zdůvodnění nízké přesnosti MLP, přeučení rozhodovacího stromu atd. Student používá algoritmus k-NN, nikde však není zmíněna volba k. V práci chybí zmínka o způsobu vyhodnocení úspěšnosti modelů (tzn. přesnost klasifikátoru, přestože je použita). Argumentace dosažených výsledků mi nepřijde úplně silná - přestože student dosáhl dobrých výsledků, jejich získání při daných argumentech považuji trochu za magii. Nicméně práce je to v jádru dobrá, skýtá prostor pro další rozšíření.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	80 (B)
Popis kritéria: 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: Zvolené technologie a nástroje jsou adekvátní vůči tématu práce.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

80 (B)

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:

Student provedl rozsáhlou rešerši, jak dostupných statistik, tak dat, které následně připravil pro dataset využitý k prediktivnímu modelování. Cílem predikce bylo určení výhry. Student si kladl za cíl dosáhnout přesnosti jako sázkové kanceláře - 60,76 %, což se mu téměř povedlo (59,58 %). Na práci lze navázat dalšími experimenty, v práci byly vyzkoušeny pouze základní metody a moderní XGBT. Stálo by za to srovnat i další ensemble algoritmy nebo neuronové sítě.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

- 1) Proč jste zvolil dělení dat (trénovací:validační:testovací) v poměru 56:19:25?
- 2) Zkoušel jste vybírat parametry modelů i jiným způsobem, než je ParametrGrid, u kterého lze očekávat vyšší časovou náročnost než např. u GridSearchCV nebo RandomizedSearchCV?
- 3) Proč jste pro výběr příznaků použil právě metodu SelectKBest?

Hodnotící kritérium:

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

6. Celkové hodnocení

70 (C)

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:

Práce se zabývá predikcí výsledků zápasů v NHL. Student provedl rozsáhlou rešerši, jak dostupných dat, tak metod. Při přípravě dat student projevil značné úsilí, k prediktivnímu modelování mám několik výhrad. Práce splňuje veškeré náležitosti, nicméně ji doporučuji hodnotit stupněm C jako kompromis mezi dosažením dobrých výsledků za použití adekvátních metod a nedostatečnou argumentací zvoleného řešení.

Podpis oponenta práce: