



Posudek oponenta závěrečné práce

Student: Jan Garček
Oponent práce: Ing. Mgr. Pavla Vozárová, Ph.D., M.A.
Název práce: Predikce hodnoty hráčů fotbalu
Obor: Znalostní inženýrství

Datum vytvoření: 16. 6. 2020

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Zadání bylo splněno, pokud "hodnotu hráče", tedy termín použitý v názvu i zadání práce, chápeme jako přestupovou částku a nikoli tržní hodnotu (je pochopitelné, že nepozorovanou tržní hodnotu zkoumat nemůžeme).	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	80 (B)
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Práce je psaná čtivě, je na vysoké stylistické i jazykové úrovni. Celkové členění práce by mně osobně více vyhovovalo jiné, rešerši existujících studií bych předřadila kapitole o zpracování dat, nicméně toto není na úkor pochopení textu. Místy v práci ruší drobné logické skoky, kdy text zcela plynule nenavazuje, nicméně v kontextu bakalářské práce technického směru se jedná vcelku o nadprůměrně zdařilé dílo. Jediná věc, kterou bych autorovi v oblasti stylistiky vytkla, je přílišné využití trpného rodu ("data byla zpracována" apod.) a formulací typu "práce testuje model" - místy opravdu není jasné, co udělal autor práce a co je převzaté odjinud. Bibliografické citace jsou v pořádku a je jich dostatečné množství. Jako mírně problematickou spatřuji přílišnou úspornost textu v některých oblastech, zejména mi chybí detailnější popis výsledného datového souboru, který je v práci používán, a to základní textová i statistická deskripce. Z textu například vyplývá, že bylo třeba doplnit chybějící pozorování, zde mi chybí zásadní informace, jaká část pozorování pro jednotlivé sloupce chyběla (a byla doplněna). Také není úplně jasné, jaké proměnné datový soubor obsahuje a hlavně proč byly modely odhadovány nejprve na menším datasetu (kapitola 4.1) a poté na rozšířeném datasetu (4.2). Také bych ocenila detailnější rozbor výsledků jednotlivých metod, k tomu však více v části o využitelnosti práce.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	80 (B)
<i>Popis kritéria:</i> 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:

Na práci rozhodně oceňuji shromáždění bohatého datového souboru z více zdrojů, které si jistě vyžádalo pečlivé a časově náročné čištění a zpracování dat. Také se mi líbí, že autor využil a porovnal několik různých predikčních metod a zamyslel se i nad problematikou výběru proměnných v modelu použitých (byť zde své úvahy z mého pohledu v práci dostatečně nevysvětlil). Mám však několik výtek k technickým a statistickým aspektům empirické části práce.

- 1) V kapitole 1.3 autor zmiňuje, že v jednom z použitých zdrojů jsou více zastoupeni brankáři a záložníci než útočníci. Není ale jasné, zda se tato vlastnost po pročištění dat přenáší do výsledného datasetu (z tohoto důvodu mi v textu chybí základní statistický popis použitých charakteristik, viz hodnocení písemné části). Pokud je to pravda, pak se nejedná o náhodný výběr a výsledky mohou být zkreslené.
- 2) Z některých jasně kvalitativních proměnných (např. pozice hráče) jsou vytvářeny proměnné kvantitativní (ordinální), to ale znamená, že je modelu vnučována předem daná struktura (zde to, že útočník je n-krát lepší než brankář, kde n je pevně zvolené), což zkresluje výsledky.
- 3) Vyplnění chybějících pozorování bez detailnější diskuse mě poněkud "znervózňuje" - může to mít, v závislosti na použité metodě, výrazný vliv na výsledky. Bylo by vhodné doplnit práci o nějaký test robustnosti, tedy například prozkoumat, jak by predikční metody fungovaly, pokud by se pozorování s chybějícími záznamy vůbec nepoužily, případně pokud by k doplnění záznamů byla použita jiná metoda.
- 4) Je mi jasné, že ve třetí kapitole se jedná převážně o vizualizaci dat, nicméně i tak by bylo vhodné některá tvrzení o tom, zda daná charakteristika snižuje či zvyšuje přestupovou částku, doplnit standardním statistickým testem, který by ukázal, zda je toto snížení či zvýšení statisticky významné. Na straně 35 autor dokonce mluví o pravdivosti jisté hypotézy - zde by čtenář nějaký formální test opravdu čekal.

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:

Práce je celkově zajímavá a až na jisté výše uvedené výhrady plní predikční metody v ní použité svůj účel, nicméně z pohledu fotbalového klubu by mě nazajímalo ani tak to, nakolik je odhad přestupové částky přesný, jako to, které proměnné jsou pro tuto částku nejdůležitější (a tedy na jaké hráče se zaměřit, případně na jakých charakteristikách klubu "zpracovat", aby byl splněn ekonomický cíl - tedy pokud možno co nejlevnější nákup a co nejdražší prodej hráče). Jisté vodítko by nám mohla poskytnout kapitola 4.3, ve které autor používá metody umožňující vybrat 15 nejdůležitějších charakteristik, nicméně ani v jednom případě nediskutuje, jaké charakteristiky to jsou a jestli jsou v souladu s pozorováními uvedenými v třetí kapitole.

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 na obrázku 3.5 použil 4 různé typy grafů pro znázornění 4 velmi podobných problémů?
- 2) Pokud by váš výsledný datový soubor nebyl náhodným výběrem (například co se týče pozice hráče), jak by to ovlivňovalo vaše výsledky?
- 3) Jaké jiné charakteristiky než RMSE a doba běhu by mohly určit, kterou metodu byste pro predikci přestupové částky nejraději použil?
- 4) Na str. 21 označujete jeden z modelů zmiňovaných v rešerši jako "robustní". Jak chápete tento termín?

Hodnotící kritérium:

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

6. Celkové hodnocení

80 (B)

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 je čtivě napsaná a zajímavá a jsem ráda, že jsem měla možnost ji oponovat. Po metodologické stránce k ní mám pár výše uvedených výhrad a některé aspekty měly být lépe vysvětlené, nicméně autor prokázal, že umí definovat výzkumnou otázku, najít data nezbytná pro její zodpovězení, tato data zpracovat a aplikovat na ně vybrané metody predikce. Drobné nedostatky, které jsem v práci našla, jsou vyváženy celkovou kvalitou textu a jistě značnou náročností práce s daty, a proto navrhuji celkové hodnocení B.

Podpis oponenta práce: