

# Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: doc. RNDr. Iveta Mrázová, CSc.

Jméno a příjmení autora práce: Tomáš Konečný

Název práce: Předpovídání výsledků zápasů v šipkách

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Snadná dostupnost velkého množství dat nejrůznějšího typu a rozvoj technik pro jejich efektivní zpracování umožňují vývoj aplikací i pro dříve méně obvyklé oblasti, např. odhad výsledného skóre závodů ve sportu. Tato oblast je pak zajímavá ať už z pohledu analýzy probíhajících turnajů anebo pro sázkové kanceláře. Předkládaná diplomová práce Tomáše Konečného se věnuje problematice modelování sportovních utkání a predikce jejich výsledků. Student měl nastudovat dostupnou literaturu zabývající se předpovídáním výsledků sportovních zápasů obecně a specificky i metodami používanými pro soutěže v šipkách. Na základě provedené rešerše měl uchazeč navrhnout model pro předpovídání výsledků zápasů v šipkách a porovnat jej s existujícími přístupy pro řešení tohoto problému, zejména se stávajícími ratingovými modely, obecně používanými statistickými metodami a novějšími technikami založenými na principech strojového učení.

Podstatná část předkládané práce je věnována popisu údajů o jednotlivých zápasech a struktuře data setu použitého pro vytváření jednotlivých predikčních modelů. Těžiště práce pak spočívá ve vytipování vhodných technik pro spolehlivý odhad výsledku zápasů v šipkách. Kromě přístupů známých z literatury autor navrhl i původní ratingový model založený na logistické regresi a pravděpodobnostní Markovův-fyzikální model. Z modelů založených na principech strojového učení jsou v práci testovány mj. i rozhodovací stromy (Lightgbm), SVM stroje a (dopředné) neuronové sítě. K realizaci popsáných modelů a jejich vzájemnému porovnání diplomant použil vlastní software implementovaný v jazyce Python. Vytvořený „darts-model“ tvoří přílohu práce.

Práce sama je napsaná česky, je čtivá, přehledně členěná a má hezkou grafickou úpravu (výjimkou je např. 1. ř. na st.13) s celou řadou ilustrativních obrázků. Seznam použité literatury čítá téměř 30 vesměs relevantních položek, ze seznamu doporučené literatury obsaženého v zadání práce je však v práci použita jen jediná publikace (z celkem 4). Některé z uváděných citací jsou neúplné, např. publikace Haugha a Wanga (2020), Sullivana (2015) a Oettinga a kol. (2019). Poněkud diskutabilní jsou uváděné výsledky testovaných modelů, více viz. otázka 1 ze závěrečné části posudku. Podrobná diskuse dosažených výsledků a použitých postupů nicméně dokládá dobrou orientaci autora v řešené problematice.

Předkládaná práce splňuje svůj původní cíl. Diplomant pronikl dostatečně hluboko do problematiky klasifikačních úloh a pravděpodobnostních modelů. Prokázal schopnost samostatně řešit i poměrně náročné úlohy spojené s modelováním sportovních utkání a predikcí jejich výsledků. Vlastní výsledky a zkušenosti dokázal objektivně vyhodnotit. K obhajobě práce mám následující otázky:

1. Tabulka 5.1 podává přehled nejlepších výsledků dosahovaných jednotlivými testovanými metodami. Jak velký byl rozptyl dosahovaných výsledků? Jaké je nejmenší množství vzorků potřebné pro natrénování úspěšných modelů a mohlo by využití stratifikačních technik nároky na počet trénovacích dat významným způsobem snížit?
2. Autorem jinak favorizovaný Markovův-fyzikální model není příliš návratový. Posuďte, jak by bylo možné tento model změnit/rozšířit, aby dosahoval pozitivních hodnot celkové realizované návratnosti (RoI.). Bylo by možné dosáhnout vyšší návratnosti i u jiných testovaných modelů, popřípadě tyto modely navzájem kombinovat?
3. Odhadněte generalizační schopnosti použitého modelu `LGBMClassifier` v závislosti na množství použitých příznaků (`feature_fraction`). Vysvětlíte, jaký vliv by mohlo mít využití zašuměných vzorků při učení, resp. vyhodnocování vlastností tohoto modelu.

## Doporučení k obhajobě:

Z výše uvedených důvodů práci doporučuji k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	NE [x]
---	--------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

V Praze dne: 25. 1. 2023

Podpis:\*\*

\* *nehodící se škrtněte (vymažte)*

\*\* *do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*