

# Hodnocení vedoucího diplomové práce

Autor/Autorka

Bc. Ivana Gabrišková

Název práce

Modelování a odhadování výsledků hokejových utkání

Studijní obor

Finanční informatika a statistika

Vedoucí práce

Ing. Patrice Marek, Ph.D.

## Splnění cílů práce:

- nadstandardně     velmi dobře     splněny     s výhradami     nebyly splněny

## Odborný přínos práce:

- nové výsledky     netradiční postupy     zpracování výsledků z různých zdrojů     shrnutí výsledků z různých zdrojů     bez přínosu

## Matematická (odborná) úroveň:

- vynikající     velmi dobrá     průměrná     podprůměrná     nevyhovující

## Věcné chyby:

- téměř žádné     vzhledem k rozsahu přiměřený počet     méně podstatné, větší množství     podstatnější, větší množství     závažné

## Grafická, jazyková a formální úroveň:

- vynikající     velmi dobrá     průměrná     podprůměrná     nevyhovující

## Přístup autora k řešení práce, spolupráce s vedoucím práce:

- samostatná práce s výbornou komunikací     pečlivá práce, drobné zásahy vedoucího     pečlivá práce, podstatnější zásahy     horší komunikace     špatný přístup k práci

## Slovní hodnocení a dotazy:

### Shrnutí hodnocení

Předložená práce se zabývá odhadem výsledků hokejových utkání. K tomu jsou využity existující modely a autorkou je navržena a ověřena jedna změna v těchto modelech. Výsledky jsou demonstrovány na fiktivním sázení proti sázkové kanceláři, tedy zprostředkovaně přes kurzy jsou srovnány s modely sázkových kanceláří.

Na práci lze ocenit zajištění velkého množství špatně dostupných dat, jejich kompletaci, kontrolu a úpravu. Naopak, velice problematický byl přístup autorky ke zbytku práce, kdy autorka začala intenzivně pracovat až v dubnu 2016. Modely předkládané na konzultacích byly často chybně zpracovány a bylo nutné práci na nich podstatně korigovat a upozorňovat na zásadní chyby. Díky pozdnímu začátku a problémům při zpracování modelů je na textu práce velice znát, že byl sepsán v rychlosti a obsahuje tedy formální i věcné chyby. Často se stává, že některé zvolené postupy a hodnoty nejsou dostatečně a přehledně zdůvodněny (např. zásadní volba, proč jsou používány jen výsledky po 60 minutách). Práce by zasluhovala i logičtější členění kapitol (např. podkapitola 6.9 by měla být uvedena jako samostatná kapitola).

Dalším problémem práce je používání zdrojů, ty jsou sice obvykle uváděny, ale často jen jako inspirace pro danou kapitolu. Přitom se v některých případech jedná o citaci, případně překlad anglického textu a mělo by to tedy být z textu zřejmé. Konkrétně jde o často užívané zdroje [3] a [12]. Uváděné zdroje [9] a [10] v práci nejsou použity vůbec a v kapitole 2 naopak zdroje chybí. Dle provedené kontroly plagiátorství se texty uvedené v této kapitole shodují s texty na stránce [cs.wikipedia.org/wiki/NHL](http://cs.wikipedia.org/wiki/NHL) (podobně i pro další dvě ligy).

Vlastní model navržený v kapitole 6.9 (str. 39) není formulován v souladu s předchozími modely a hodnoty parametrů výhody domácího prostředí nejsou relativní, jak tomu je u ostatních dvou parametrů (síla útoku a obrany) v originálním modelu. Vlastní modifikace by rovněž měla obsahovat více než jeden pokus o úpravu modelů.

### Otázky

- Jakým způsobem funguje Bonferroniho korekce? Demonstrujte hlavní princip a uveďte, jak mají znít správné závěry v tabulce 5 (str. 10).
- Jakým způsobem by bylo možno z parametrů výhody domácího prostředí ve vlastním modelu vytvořit parametry s relativní hodnotou?

### Hodnocení

I přes uvedené výhrady doporučuji předloženou práci k obhajobě a navrhuji hodnocení stupněm **DOBŘE**.

### Podrobné připomínky k práci

#### Kapitola 2

- *Neuvedené zdroje citovaného textu popisu soutěží.*

#### Kapitola 3

- *Neúplná definice Poissonova rozdělení a přejatý popis ze zdroje [6] je vytržený z kontextu.*
- *Pro Chí-kvadrát test dobré shody uvádí autorka, že očekávané četnosti musí být vyšší než 5 (str. 6), ale při testování toto nedodržuje (např. soubor „CZE\_testování predpokladu.xlsx“, tým Olomouc, Hradec Králové, aj.).*
- *Chybné označení rozsahů ve vzorci (3.7).*
- *Chybně zapsaný počet stupňů volnosti pod vzorcem (3.8).*

#### Kapitola 4

- *Neskloňování názvů měst týmů a popisků (tabulka, obrázek) – platí v celém textu.*
- *Tabulka 3 (str. 9) – není jasné, co znamená  $H_0$ , zde pravděpodobně individuální hypotéza, výše ale definováno jako souhrnná hypotéza.*
- *Tabulka 5 (str. 10) – vzhledem k p-hodnotám chybné vyhodnocení hypotéz (např. Kraków, kde je i chybný zápis p-hodnoty). Toto má zásadní vliv i na další prezentované modely!*

#### Kapitola 5

- *Kapitola 5.2.1 shrnuje výsledky kapitoly 4.1, navíc ale chybně (viz předchozí bod) a tvrdí, že testy proběhly po sezónách, což není pravda.*
- *Bod 5 postupu na str. 15 určuje, že postup proběhne vždy pevně pětkrát. Měla by být ale definována nějaká podmínka konvergence.*
- *Hypotézy na str. 17 jsou nepřesně nedefinovány, není jasné, co přesně je testováno.*

#### Kapitola 6

- *Na str. 19 je chybně uvedeno spojení „spojitá pravděpodobnostní funkce“, má se jednat o „sdruženou pravděpodobnostní funkci“.*
- *Nesmyslné první dvě věty na začátku str. 20.*
- *Kapitola 6.5.1 tvrdí, že dříve uvedené modely používají váhovou funkci. To není pravda, ta je zavedena až dále a jedná se o rozšíření daných modelů.*

- V kapitole 6.6 se objevuje tvrzení, že v prvních kolech jsou odhady nestabilní, ale není to nikde ukázáno.
- Všechny grafy od strany 28 jsou špatně zvolené, jelikož je v nich použito vyhlazení. V této situaci je vždy potřeba mít konkrétní (nevyhlazené) hodnoty parametrů.
- Na konci kapitoly 6.6.2 (str. 29) je tvrzení, že náhodné veličiny jsou závislé. Toto je usouzeno pouze z hodnoty jednoho parametru, ale není proveden žádný test.
- Na str. 29 je konstatováno, že pravděpodobnosti se odhadují do stavu 15:15. Toto není nijak zdůvodněno a ani není popsáno, jak je řešena skutečnost, že součet pravděpodobností není v tom případě roven jedné.
- V tabulce 15 (str. 31) je špatně označena optimální hodnota.
- Z obrázku 13 (str. 35) plyne, že je bez vysvětlení používána podmínka „větší nebo rovno 0.1“ místo v teorii udané podmínky „větší než 0“.
- Vlastní model (str. 39) není formulován v souladu s předchozími modely a hodnoty parametrů výhody domácího prostředí nejsou relativní, jak tomu je u ostatních dvou parametrů (síla útoku a obrany) v originálním modelu. Mělo by proběhnout více pokusů o modifikaci modelů, než jen jeden uvedený. Vzhledem k nedostatku času ale k tomuto nedošlo.

#### Kapitola 7

- Nejsou zde řádně vysvětleny „triviální“ strategie sázení.
- V tabulce 21 (str. 41) je chybné označení „relativní zisk“. Jedná se o „relativní hodnotu majetku“.
- Na str. 43 je tvrzení, že model ani jednou nekončí bankrotem, z obrázku na další straně je ale zřejmé, že k bankrotu (tj. stavu pod 1000 Kč) došlo.
- Tvrzení na konci kapitoly 7.1 nedává smysl. Díky průměrování je naopak zisk pravděpodobně snížen.

**Navrhuji hodnocení známkou:**

**DOBŘE**

25. 8. 2016, Patrice Marek:

