



# Posudek oponenta závěrečné práce

Oponent práce:	Ing. Jitka Hrabáková, Ph.D.
Student:	Bc. Martin Rameš
Název práce:	Hledání $t_{BH}+(\tau)$ s optimalizací výkonnosti pro oddělení signálu a pozadí za použití strojového učení a dat z ATLASu
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	1. června 2023

## Hodnotící kritéria

### 1. Splnění zadání

- [1] zadání splněno
- ▶ [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Student nezkoumal možnost optimalizace separace signálu a pozadí na základě systematické nejistoty. Dle vyjádření studenta po dohodě se školitelem. Zadání považuji za mírně obtížnější, neboť student musel nad rámec své specializace alespoň částečně porozumět fyzikálnímu pozadí dat, která zpracovával.

### 2. Písemná část práce

70/100 (C)

Rozsah diplomové práce je dostatečný, všechny části jsou informačně důležité. Práce je napsána vcelku srozumitelně, jednotlivé části na sebe navazují, gramatických chyb není v textu mnoho. Zdroje jsou korektně citovány, stejně jako použitý software. Drobnou výtku mám k uspořádání odkazů v literatuře, nezdá se být abecední podle žádného klíče ani chronologické. Některé části vlastní práce by mohly být popsány méně úsporně. Čtenář by jistě ocenil detailnější popis toho co vidí na obr. 4.5 nebo 4.14 a co to prakticky znamená. V teoretickém úvodu do částicové fyziky se bez znalosti toho oboru čtenář spíše nevyzná. Objevuje se značení bez vysvětlení (například symboly v rozpadovém kanálu), většina symbolů v diagramu na obr. 2.2, nevysvětlené jednoty  $\sigma$  veličiny cross-section, atd. Podkapitola Weights ( v části částicová fyzika) týkající se simulovaných dat čtenáře mírně zaskočí, čekal by jí jinde. Další výhrady mám k popisu vyhodnocovacích metrik. Senzitivita není vůbec vysvětlena a popis ke statistické signifikanci je místy nepřesný, např.  $\hat{\mu}$  je MLE odhad parametru  $\mu$ , ne parametr, rovnice 2.5 (díky přeuspořádání a chybějícímu kontextu ze zdrojového článku) nikterak nepřispívá k pochopení. Další nejasnosti se týkají obr. 3.4 kde graf pro Background - v5 data příliš nekořesponduje s textem a navíc vypadá značně jinak než ostatní grafy.

### 3. Nepísemná část, přílohy

80 /100 (B)

Nepísemným výsledkem jsou zdrojové kódy v jazyce Python umožňující zopakování experimentů. Krátký popis jednotlivých částí a jejich návazností by byl vhodný.

### 4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Provedenou analýzu a její výsledky považuji za dále použitelné pro další práci v oblasti separace signálu a pozadí pomocí ML algoritmů.

## Celkové hodnocení

80 /100 (B)

Student vyzkoušel a optimalizoval 4 různé algoritmy strojového učení pro separaci signálu a pozadí, pro každou ze 7 úloh stanovil nejlepší model, dále se věnoval hodnocení významnosti příznaků a jejich redukci. Svě výsledky porovnává s předchozím výzkumem, což lze jen částečně, neboť pracoval na jiné (pravděpodobně složitější) datové sadě než předchozí práce. Celkové vyznění práce mírně poškozuje výše zmíněné připomínky které místy ubírají na snadné čitelnosti a porozumění textu.

## Otázky k obhajobě

1. Řada obrázků má potenciál být velmi informačně bohatá, pokud dobře vysvětlíte, co prezentujete. Mohl byste detailně popsat co znázorňuje např. obr. 4.5, co to prakticky znamená? Jak byla získána modře šrafovaná oblast 'Uncertainty'?
2. V části 3.1.2 Features provádíte doplnění chybějících hodnot, je to Vámi navržená metoda, nebo metoda doporučená 2ISS1tau skupinou? Stejná otázka se týká i normalizace příznaků. Stanovoval jste Vy, které příznaky a jak škálovat, nebo je to dle doporučení 2ISS1tau skupiny?

## **Instrukce**

### **Splnění zadání**

Posudte, 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, posudte 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.

### **Písemná část práce**

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, 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. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, 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.

### **Nepísemná část, přílohy**

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.

### **Hodnocení výsledků, jejich využitelnost**

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.

### **Celkové hodnocení**

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.