



# Posudek oponenta závěrečné práce

**Oponent práce:** doc. Ing. Pavel Kordík, Ph.D.  
**Student:** Bc. Tomáš Detko  
**Název práce:** Clustering a datová analýza Jupyter notebooků na GitHubu  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 30. května 2022

## Hodnotící kritéria

### 1. Splnění zadání

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

Práce se moc neřídí instrukcemi v zadání. Jde vlastním směrem, ale selhává ve vysvětlení motivace odchýlení se od zadání.

### 2. Písemná část práce

40 / 100 (F)

Práce je psána dobře čitelnou angličtinou, obsahuje drobné překlepy. Hlavní problém práce je ale absence příběhu a motivace - proč by se měl čtenář vůbec věnovat například metodám pro predikci časových řad, když pak nejsou v experimentální části použity? Jak souvisí se zadáním? Chaotické je i členění práce, kde je například analýza schovaná v implementaci a rešeršní část predikce časových řad rozdělena na 2 kapitoly proložené popisem akvizice dat. Experimentální část je velmi stručná a chybí metodologie experimentů. Ve finále je hlavním výstupem srovnání CNN a LSTM architektury, ale interpretace výsledků hlavně směrem k otázkám v zadání pokulhává.

### 3. Nepísemná část, přílohy

45 / 100 (F)

Dost práce zřejmě zabralo předzpracování dat, škoda že v textu není více popsána extrakce atributů a související metody. Pokud jde o modely, vedle zmíněného CNN a LSTM jsem v repozitáři ještě objevil jednu RNN (echo state machine), ale není jasné, zda jde o autorovu implementaci (žádné komentáře) a text o ni mlčí.

#### **4. Hodnocení výsledků, jejich využitelnost**

36 /100 (F)

Kdyby bylo lépe popsáno předzpracování dat, dalo by se na práci lépe navázat. Jinak je rozsah experimentů poměrně omezený a nejsem si jistý, zda k něčemu budou.

#### **Celkové hodnocení**

42 /100 (F)

Dlouho jsem se rozhodoval, ale nakonec si myslím, že bude lepší minimálně text přepracovat. Z rešeršní části a dobré angličtiny tuším, že autor má výrazně větší potenciál. Chce to upravit zadání, zapracovat na experimentální části - hlavně na motivaci pro experimenty a interpretaci výsledků.

#### **Otázky k obhajobě**

Implementace Echo state RNN je Vaše? Proč jste ji pro řešení tohoto problému zvolil? Co clustering notebooků, ten jste vůbec nedělal?

## **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.