



# Hodnocení vedoucího závěrečné práce

**Vedoucí práce:** Ing. Magda Friedjungová, Ph.D.  
**Student:** Anna Kapitánová  
**Název práce:** Predikce výkonů studentů  
**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

Zadání bylo splněno bez výhrad.

### 2. Písemná část práce

85 /100 (B)

Písemnou část práce hodnotím pozitivně. K textu mám ale několik mírných výtek, které jsem dle mého měla zdůraznit během konzultací. Práce je psána v českém jazyce, je logicky a přehledně strukturována. Jednotlivé kapitoly jsou informačně vyvážené a studentka využívá dostatek relevantních zdrojů.

Práce obsahuje menší množství překlepů a gramatických chyb. Místy se vyskytují obraty jako "velmi populární" (sekce 2.4.1), "jedna z nejrozšířenějších způsobů" (sekce 2.4.2) apod., které nejsou podloženy žádnou referencí. Samotný text bych uvítala věcnější/ odbornější bez subjektivních hodnocení, nicméně text jako takový je i při svém rozsahu čtivý a přehledný. V tabulce 3.4 a dalších by bylo dobré vždy zvýraznit nejlepší výsledky. Zkratky je vhodné zavádět při prvním použití pojmu. V sekci 4.4.1 chybí zdůvodnění, proč je v datovém skladu tak nízký počet záznamů. Obrázek 4.2 by byl lepší v jednom grafu - úspěšní/neúspěšní studenti a z toho gymnazisté/negymnazisté. V sekci 5.3.1 mi chybí popis nastavení parametrů algoritmu kMeans (počet shluků apod.). S atributy "matematika" a "programování" ta úloha svádí k manuálnímu rozřazení, zajímavější by bylo shlukování bez těchto sloupců. Dále si nejsem jistá výrokem "častým případem je, že se studentům předměty do dalšího semestru zapisují automaticky při předběžných zápisech" - toto by bylo dobré ověřit u rozvrháře. V kapitole 6 by u výsledných/nejlepších modelů bylo dobré uvést i konkrétní kombinaci hyperparametrů.

Výše zmíněné ale nepovažuji za nijak závažné.

### 3. Nepísemná část, přílohy

92 /100 (A)

Studentka v nepísemné části splnila vše, co bylo v zadání vytyčeno. Zvolila přiměřené technologie - Python a Jupyter Notebook, a veškeré experimenty jsou opakovatelné.

### 4. Hodnocení výsledků, jejich využitelnost

92 /100 (A)

V práci se vyskytuje několik analýz na různá témata (vliv pohlaví, národnosti, střední školy apod.), které považuji za cenné. Studentka během své práce také odhalila několik nekonzistencí a chyb v datovém zdroji, což je významná zpětná vazba pro pracovníky datového skladu a budoucí reporting. Odhalování těchto nekonzistencí, které bylo součástí předzpracování dat, je nejen časově náročnou prací, protože k datům není dostupná řádná dokumentace. Navíc poskytnuté záznamy nebyly v některých případech kompletní, takže studentka musela vynaložit větší úsilí, než by bylo nutné při úplných datech.

Přidanou hodnotou této práce je tak snaha vytěžit z dosavadních dat co nejvíce informací. Studentka navrhla a otestovala několik modelů pro predikci výkonu studentů - predikce dokončení studia, predikce dokončení předmětů a predikce postupu do následujícího semestru. Dále musela vyřešit několik krajních situací - ať už se jednalo o různé akreditace a předměty, znovu nastupující studenty nebo distanční výuku v době COVID-19.

Práce je dle mého využitelná a může také posloužit jako vodítko při dalších analýzách. S novou akreditací však bude nutné vytvořit nové modely nebo nějakým způsobem provést transfer stávajících. Jak studentka zmiňuje v kapitole 6.4, budeme muset počkat několik let, než získáme dostatečné množství dat na sestavení nových modelů. Případně bychom mohli vyzkoušet mapování předmětů a metody pro práci s malým množstvím záznamů či transfer learning. Aktuální predikce jsou možné pro studenty od 4. semestru dál (novější studenti jsou již v nové akreditaci), což není moc nápomocné např. pro odhad kapacit předmětů (kapacitně náročné předměty jsou zpočátku studia). Nicméně studentka připravila několik skriptů, které proces automatizují, konkrétně predikce dokončení studia a predikce dokončení 4. semestru. S mírnými úpravami tyto skripty lze využít i při nové akreditaci.

Zlepšit by se dalo snad jen následující: V sekci 4.3 při nekvalitních datech studentka využívá některá data z Bílé knihy (web scraping), zde by bylo vhodné navrhnout způsob (včetně výsledných hodnot), jak konkrétní data v datovém skladu (na endpointech) vyčistit. Takové čištění by pak mohli využít pracovníci datového skladu a data tak vyčistit i pro jiné analýzy.

Na výsledky lze navázat dalšími činnostmi, které studentka zmiňuje v závěru, nebo např.: Sekce 3.4 "COVID a jeho dopad" lze rozšířit i o analýzu SZZ a počtu získaných kreditů. V sekci 6.2.5 "Pátý semestr" by bylo vhodné zkusit nějakou metodu oversamplingu minoritní třídy. Nebo výše zmíněným návrhem čištění dat.

### 5. Aktivita studenta

- [1] výborná aktivita
- [2] velmi dobrá aktivita
- [3] průměrná aktivita

[4] slabší, ale ještě dostatečná aktivita

[5] nedostatečná aktivita

Studentka byla během řešení zadání aktivní, dodržovala termíny a průběžně svoji práci konzultovala.

## 6. Samostatnost studenta

► [1] výborná samostatnost

[2] velmi dobrá samostatnost

[3] průměrná samostatnost

[4] slabší, ale ještě dostatečná samostatnost

[5] nedostatečná samostatnost

Studentka pracovala samostatně.

## Celkové hodnocení

91 /100 (A)

Předloženou práci navrhuji hodnotit klasifikačním stupněm A. Studentka prokázala velké úsilí při zpracování zadaného tématu, práce je také opakovatelná a v mezích dané oblasti znovu využitelná. Na dosažené výsledky lze navázat dalšími činnostmi.

## **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.

### **Aktivita studenta**

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržel dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

### **Samostatnost studenta**

V souvislosti s průběhem a výsledkem práce posudte schopnost studenta samostatně tvůrčí práce.

### **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.