ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	**Zber a analýza dát získaných pomocou mobilných platforiem pre zberbehaviorálnych dát**
Jméno autora:	**Ondrej Sakači**
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Ing. Michal Huptych, Ph.D.
Pracoviště oponenta práce:	CIIRC, ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání **průměrně náročné**

Hodnocení náročnosti zadání závěrečné práce.

Zadání práce je

Splnění zadání **splněno**

Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Zvolený postup řešení **správný**

Posuďte, zda student zvolil správný postup nebo metody řešení.

Zvolený postup je správný, avšak analýza výsledků sledování by měla mít širší rozsah.

Odborná úroveň **C - dobře**

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Autor podává ucelený přehled k oblasti behaviorálních dat a popisu předpokladů a metodiky. Autor v druhé kapitole také definuje tři hypotézy, které mají být prací ověřeny. K popisu nasazené platformy a implementaci nemám připomínek. U kapitoly popisu dat mám připomínku k popisu dotazníku. V kapitole 3.2.3 v části Vstup o studie autor zmiňuje, že byla použita upravená forma dotazníku M.I.N.I. Avšak není jasné, zda se tento dotazník váže k dotazníku popsaném v kapitole 6.1.2. Bylo by vhodné kdyby dotazník byl součástí příloh. Největší připomínku mám k části výsledků. Autor uvádí přehledové tabulky výsledků pro různá data (aktidata, GPS, mix) a různé kombinace příznaků. V textu je jednou uvedeno., že výsledkem je Accuracy. Avšak tento fakt není nikde deklarován explicitně. Dovedu si jej představit v popiscích tabulek či přímo v jejich hlavičce. Bylo by také vhodné uvést výsledky klasifikace podrobněji, např. použitím senzitivity a specificity, na kterých by se nejspíše lépe projevily problémy s nepříznivou situací, ve které studie probíhala (COVID-19). Autor sice uvádí stručný obecný popis klasifikátorů, ale neuvádí nikde zda bylo provedeno nějaké jejich specifické nastavení. Autor v kapitole 6.2. uvádí, že algoritmus byl přepočítáván sto-krát pro každý subjekt při randomizací dnů s příznaky. Tento postup mi však z uvedeného krátkého popisu není jasný a není jasná ani jeho návaznost na zmiňovanou cross-validaci. U té není uvedeno jaké nastavení cross-validace bylo zvoleno. Není tak zřejmé kolik dat bylo použito na trénování klasifikátorů a kolik na testování. Není zřejmý ani důvod pro velké rozdíly ve výsledcích klasifikace mezi průměrnými nejhoršími a průměrnými nejlepšími výsledky. V práci není žádná vizualizace dat. Bylo by vhodné uvést příklad aktidat v různých třídách (třeba v ideálním případě). U tohoto typu dat bych takovouto vizualizaci očekával.
V diskuzi pak ne zcela rozumím zavržení třetí hypotézy definované autorem. Není k ní totiž v práci žádná specificky přiřazená analýza, která by tento závěr prokazovala.

Formální a jazyková úroveň, rozsah práce **B - velmi dobře**

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Práce je logicky rozdělena do kapitol a napsána srozumitelným technickým jazykem. V práci se objevují několikrát chyby v interpunkci a obsahuje překlepy.

Výběr zdrojů, korektnost citací **B - velmi dobře**

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Zvolené citace jsou relevantní a správně použité. Jedinou připomínku mám k umístění seznamu literatury do dodatků (Dodatek 3), což není standardní způsob jak seznam literatury do práce uvést.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Vložte komentář (nepovinné hodnocení).

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Bakalářská práce Petra Schimperka se věnuje sběru a analýze získaných dat. Práce splňuje zadání a zvolený postup hodnotím jako správný. Připomínky mám hlavně ke způsobu popisu tvorby datasetů pro klasifikaci, který není dostatečně jasný. Ve výsledcích analýzy není explicitně dáno, že mírou pro porovnání výsledků je zvolena Accuracy. Bylo by vhodné ji doplnit nejméně o senzitivitu a specificitu klasifikátorů, které by nejspíš umožnili přesnější interpretaci výsledků klasifikace. Analýza je tak základní a např. vliv rozdělení příznaků do různých skupin není ve výsledcích ani v diskuzi práce rozebrán. Práce definuje tři hypotézy jejichž potvrzení, resp. vyvrácení je v práci provedeno zjednodušeně a není více diskutováno (týká se hlavně třetí zamítnuté hypotézy, viz Odborná úroveň). I když je práce po stránce formální a korektnosti mnou hodnocena velmi dobře, považuji tyto připomínky za převažující v celkovém hodnocení.

K práci mám následující otázky:

- Jak byly vytvářena randomizace při přepočítávání algoritmů a jak nastavena cross-validace?
- Co nejvíce ovlivňuje rozdíly ve výsledcích klasifikace?
- Na základě čeho byla třetí hypotéza definovaná v práci “Rozlíšenie pracovných a voľných dní bude mať najlepšie výsledky pri kombinácií aktigrafických dát s dátami z mobilného zariadenia“ brána jako zamítnutá?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře.**

Datum: 26.1.2022 Podpis: