

## **Oponentský posudek bakalářské práce**

Název práce: Hodnocení krajinného pokryvu Vojenského újezdu Libavá s použitím klasifikátoru Random Forest

Autor: Vít Žďánský

Oponent: Mgr. Josef Laštovička

### **HODNOCENÍ PRÁCE – OBSAHOVÁ STRÁNKA**

#### **název**

- (x) dobře vystihuje obsah práce, je věcný a stručný
- ( ) vystihuje obsah práce, avšak ne zcela jednoznačně
- ( ) nevystihuje obsah práce

#### **cíle a hypotézy**

- (x) jasně formulované
- ( ) nejasně formulované
- ( ) chybí

#### **úvod a literární přehled**

- (x) jasný a přehledný, vztahuje se k zadané problematice a obsahuje dostatečné množství informací, které shrnuje na přiměřené úrovni poznání v dané oblasti; kriticky hodnotí použité informace
- ( ) nepřehledný – nelogicky členěný, bez kritické analýzy, avšak s dostatečným množstvím informací
- ( ) příliš stručný (s nedostatečným množstvím podkladů) nebo nevýstižný a ke zpracovávané problematice se vztahuje pouze částečně

#### **použité informační zdroje založené zejména na**

- (x) původní zahraniční a domácí literatura
- ( ) učebnice, slovníky a monografie
- ( ) „šedá literatura“

#### **použitá literatura**

- (x) v odpovídajícím rozsahu
- ( ) v nedostatečném rozsahu

#### **data a metody**

- (x) jasné, přehledné a srozumitelné, nechybí nic podstatného, množství materiálu (pozorování, opakování v prostoru a čase) a použité metody jsou takové, že pomohou splnit vytčené cíle
- ( ) jasné a srozumitelné, nechybí nic podstatného, ale množství materiálu je nedostačující
- ( ) nesrozumitelné (není možno posoudit adekvátnost použitých metod a materiálu) nebo nedostatečně popsané
- ( ) nevhodně zvolené metody, nemohou dát odpověď na vytýčené cíle

#### **výsledky**

- (x) vhodně prezentované, odpovídají použité metodice, k vyhodnocení použity vhodné metody,
- ( ) zbytečně se opakující výsledky (např. dvojí prezentace v tabulkách i grafech),
- ( ) k vyhodnocení nebyly použity vhodné metody
- ( ) prezentace nedostatečná

#### **interpretace dat (diskuse)**

- (x) odpovídající, autor prokázal dobrou znalost studované problematiky, vhodně cituje dostatečné množství literárních zdrojů

- diskuse dat je nedostatečná (diskuse neodpovídá úrovni a rozsahu uvedených dat)
- data nepodložená, svým rozsahem neodpovídá zpracovaným datům a údajům, spekulace převládají nad fakty

#### **závěry**

- práce má jasné a jednoznačné závěry, které jsou podloženy a odpovídají na cíle a hypotézy práce
- závěry jsou sice přesné a podložené, ale úplně neodpovídají cílům práce, nebo některé cíle a hypotézy nejsou zmíněny
- závěry nejsou podložené či nevycházejí z předkládané práce

#### **HODNOCENÍ PRÁCE – FORMÁLNÍ STRÁNKA**

##### **obrázky a tabulky**

- přehledné a obsahují dostatečné množství informací
- nepřehledné, ale obsahují dostatečné množství informací
- nejsou součástí textu
- nevyhovující

##### **text**

- formálně dokonalý
- bez větších formálních nedostatků
- po formální stránce nevyhovující (nemá doporučené členění)

##### **jazyk**

- odpovídá gramatické správnosti, možnostem a zvyklostem používání v odborném textu
- částečně odpovídá gramatické správnosti, možnostem a zvyklostem používání v odborném textu
- neodpovídá gramatické správnosti, možnostem a zvyklostem používání v odborném textu

##### **literatura**

- citována bez chyb, jednotně, citace použité v textu odpovídají seznamu literatury a způsob citací odpovídá mezinárodním nebo českým normám
- citována s chybami, nejednotně, citace v textu neodpovídají seznamu literatury (chybějící nebo přebývající citace)

##### **obsažené informace jsou**

- pro obor nové nebo rozšiřující poznání a v dostatečném rozsahu, mohou tvořit základ publikace v odborném vědeckém časopise
- jsou cenným potvrzením (aplikací) známých skutečností
- jsou jen opakováním již známých skutečností bez nového přínosu

**Práce splňuje požadavky kladené na bakalářské práce předkládané na Katedře aplikované geoinformatiky a kartografie, PřF UK. Práci doporučuji k obhajobě a hodnotím klasifikačním stupněm: Dobře.**

Datum: 27/8/2019

Podpis oponenta:

## Doplňující komentáře:

Bakalářská práce pana Víta Žďánského (s názvem *Hodnocení krajinného pokryvu Vojenského újezdu Libavá s použitím klasifikátoru Random Forest*) se v teoretické rovině zabývá oblastmi vojenských újezdů a možnostmi klasifikací těchto oblastí pomocí RF klasifikátoru. V minulosti bylo napsáno mnoho podobných prací na tematiku vojenských újezdů a na porovnání klasifikačních algoritmů. V této rovině nepřináší práce nic zlomového. Domnívám se, že literární rešerše a úvod do problematiky v kapitole 2.1 a 2.2 mohl autor zcela vypustit a odkázat se pouze na předchozí práce. Naopak větší úsilí mohl věnovat využití RF klasifikátoru, který je dnes používán například i v oblastech časových řad nebo i jiných odvětvích Dálkového průzkumu Země. Zájmové území bylo vyčleněno správně. Avšak mnoho parametrů z části Data a metodika by šlo upřesnit (viz Doplnující otázky). Práce mnohdy obsahuje i překlepy a chyby. Výsledky, diskuze, závěr a přílohy představují standardní výstup pro BP. Diskuzi by však bylo dobré více rozepsat, vzhledem k množství předchozích prací s obdobným zaměřením. Práce přes všechny nedostatky splňuje požadavky kladené na bakalářské práce předkládané na Katedře aplikované geoinformatiky a kartografie, PřF UK. Proto ji doporučuji k obhajobě a hodnotím klasifikačním stupněm dobře.

Autor v textu nejdříve používá „dálkový průzkum země“ a následně „Dálkový průzkum Země“.

Při definici cílů bylo uvedeno: „Hlavním cílem této bakalářské práce je klasifikace land cover (krajinného pokrytí) Vojenského újezdu Libavá pomocí volně dostupných družicových dat Sentinel 2 z roku 2019. Za tímto účelem bude použit klasifikátor Random Forest (RF) v procesu klasifikace využitím programu SNAP. Výsledky krajinného pokryvu budou porovnány s klasifikátorem Maximum Likelihood. Nedílnou součástí této klasifikace bude navržení vhodné legendy land cover. V závěru bude vyhodnocení přesnosti obou klasifikátorů.“ Nebylo by spíše vhodné uvést, že cílem práce je porovnání možností přesnosti v současnosti nejužívanějších typů algoritmů pro klasifikace obrazových snímků? Domnívám se, že jde o špatnou interpretaci cílů.

Str. 29 – chyba v tabulce 1 („infrered“).

Str. 18 – chyba v názvu klasifikace („Likelyhoody“).

Str. 19 – tato věta u SVM nedává smysl: „Ideálně umístěná nadrovina je taková, kde nejmenší vzdálenost nejbližších bodů od roviny je co největší.“

Str. 19 – překlep: „Způsobů, jakým je přístupováno k segmentaci je několik, například top-down (od největších objektů k menším), bottom-up (od nejmenších objektu k větším), boundary seeking, object seeking atd.“

Str. 20 – opět chyby: „Klasifikační model používaný v objektových klasifikacích může být založen na tzv. tvrdých (např metoda rozhodovacího stromu) nebo měkkých klasifikátorech (Fuzzy model využívající příkazových operátorů a matematických operací) (GISAT 20??). Mezi další užívané klasifikátory patří SVM, neuronové sítě, rozhodovací stromy, Maximum Likelihood a další.“

Str. 21 – chybí citace pro Ho. Na str. 22 citace je, avšak není uvedena v seznamu literatury.

Str. 35 – překlep v metodice: „Jako referenční data byla využito Ortofoto ČR, ZABAGED a LPIS. Orná půda se vyskytuje většinou na okraji území.“ Správně „využita“.

V metodice je zaměněn pojem převzorkování s transformací.

Řešením pro lepší klasifikaci by dle mého mínění bylo zkusit vyselektovat zástavbu, která u výsledků tvořila největší rozdíly obou klasifikací.

K řešení klasifikace byla užita CORINE databáze, v souhrnu použitých dat a databází však není zahrnuta.

Na str. 42 dole je chyba u popisu Kappa indexu.

Špatně pojem InSitu – správně In Situ nebo In-Situ.

Pro vytvoření klasifikace bylo užito polygonů. Domnívám se, že lepší by bylo užít pixely (tedy body, pomocí kterých by byly vybrány jednotlivé pixely).

Není přesně popsáno, jak byly automaticky vytvořeny kontrolní body, zejména jejich rozmístění po území, bylo by dobré ukázat rozmístění kontrolních bodů a ploch polygonů pro klasifikaci.

U Přílohy 3 a 4 nebyla použita stejná legenda pro lepší možnost porovnání.

### **Doplňující otázky:**

Proč byl vybrán snímek z období června? Nebyl by lepší srpen/září předchozího roku?

Není jasné, proč bylo vyloučeno pásmo 8a. K čemu se toto pásmo používá? Z jaké literatury bylo v tomto kroku čerpáno?

V případě užití softwaru SNAP pro klasifikaci RF bylo dříve zapotřebí transformovat snímek do WGS84, proběhlo tomu tak i zde? Proběhlo tedy další převzorkování? Nebo je již možné klasifikovat v novějších verzích SNAP bez transformace obrazu z jednoho koordinačního systému do druhého?

Proč byly využity dva různé programy? Nebylo lepší využít programu R, případně extenze v ENVI pomocí IDL?

Proč nebylo vytvořeno In-Situ měření? Jak byla ověřena přesnost leteckých snímků? Kdy byly pořízeny použité letecké snímky pro porovnání s družicovými daty? K jakým změnám krajiny zde během rozdílu těchto období mohlo dojít?

Není zcela jisté, k čemu přesně byly využity databáze CORINE, ZABAGED, LPIS. Mohl by prosím student objasnit využití těchto databází v rámci procesu klasifikace. Z jaké doby pocházela aktualizace daných databází pro porovnání s Vaším užitým snímkem? Jaká je polohová přesnost všech zmíněných databází? Informace není uvedena v sekci Data, tento typ informací je vždy nutné uvádět. Také by student mohl pohovořit, jak tyto databáze vznikají a s jakou přesností.

V diskuzi/závěru je uvedeno, že pro sběr trénovacích ploch byla využita barevná syntéza v nepravých barvách, není však uvedeno, jaká kombinace spektrálních pásem byla využita, případně proč a na základě jaké literatury bylo takto postupováno. Tato poznámka by měla být naopak uvedena v metodice (nikoliv v závěru/diskuzi). Dokáže student pohovořit jaká pásma byla pro syntézu využita a proč? Využit byl také index NDVI, proč nebylo použito místo tohoto indexu například NDMI nebo TC transformace?