

# HODNOCENÍ DIPLOMOVÉ PRÁCE

## Oponent DP

**Jméno diplomanta:** Bc. Ondřej Macek

**Garantující katedra:** KKY

**Název diplomové práce:** Klasifikace dokumentů s částečnou informací od učitele

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Formální a obsahová stránka práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	Vhodnost použitých metod	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Správnost získaných výsledků	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	Vlastní přínos	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Doplnění hodnocení, připomínky, dotazy:

Diplomant se ve své práci zabývá problematikou učení s částečnou informací od učitele v úloze klasifikace textových dokumentů. Vzhledem ke stále rostoucímu množství neanotovaných dat se jedná o mimořádně významnou úlohu umělé inteligence.

V teoretické části se autor poměrně dobře vypořádal s množstvím odborných pojmů, které se v češtině běžně nepoužívají, a shrnul řešenou problematiku. Hlavní nedostatek této části spatřuji v tom, že mnoho odstavců – zejména ve stěžejních kapitolách 3-6 – jeví známky pouhého překladu z originálního anglického textu do češtiny bez vlastní formulace vět.

V experimentální části autor navrhl algoritmus, který implementoval a otestoval s různým nastavením. Výsledky jsou správné a přehledně vykreslené do grafů. Navržený algoritmus je vhodný pro řešenou úlohu i pro praktické využití nastíněné v poslední kapitole.

Autor v této práci jednoznačně prokázal schopnost samostatně aplikovat metody strojového učení na reálná data. I přes popsané nedostatky práci hodnotím jako velmi dobrou, doporučuji ji k obhajobě a do diskuze navrhuji tyto otázky:

1. Autor se rozhodl implementovat vlastní algoritmus. Čím se liší od algoritmů, které jsou již implementovány v použitém nástroji scikit-learn (např. "SelfTrainingClassifier")?
2. Navržený algoritmus předpokládá zhruba stejný počet dokumentů v každé třídě. To však v plánované úloze třídění emailů splněno nebude. Jak by bylo možné Váš algoritmus modifikovat pro úlohu s nevyváženými třídami?

Splnění bodů zadání  úplně  částečně  nesplněno

Doporučení práce k obhajobě  ano  ne

**Celkové hodnocení práce**  výborně  velmi dobře  dobře  nevyhověl

Jméno, příjmení, titul oponenta: Ing. Jan Lehečka, Ph.D.

Pracoviště oponenta: KKY

2.6.2022

Datum

  
Podpis