

Posudek oponenta bakalářské/diplomové práce

Název studijního programu: Matematická biologie a biomedicína

Jméno autora/ky: Bc. Michal Bubeník

UČO: 474140

Jméno vedoucí/ho práce: Stanislav Mazurenko, PhD

Jméno oponenta/tky: Ing. Miloš Musil, Ph.D.

Název práce: Predikce substrátové specifity haloalkan dehalogenáz s využitím strojového učení

	Kritérium hodnocení	max. bodů	Body
1	Úvodní přehledová kapitola práce (úplnost relevantních zdrojů a zpracování publikovaných poznatků, vyvození závěrů pro navazující práci)	10	8
2	Tvůrčí část práce (správnost použitých postupů – z hlediska cílů práce i z hlediska principů a podmínek jejich použití, srozumitelnost jejich popisu)	10	10
3	Výsledky práce (prezentace experimentálních výsledků z hlediska jejich výpovědní hodnoty, úroveň diskuze nad získanými výsledky, úroveň a správnost závěrů vzhledem k zadání)	10	10
4	Jazyková a stylistická úroveň, gramatická správnost	5	4
5	Grafická úprava práce, obrázky, tabulky a grafy (legendy a popis, srozumitelnost bez zřetele k ostatnímu textu, vysvětlení značek, odkazy v textu)	5	5
Celkem bodů			37

Má práce tvůrčí charakter: ANO

Slovní hodnocení, dotazy a připomínky (možno připojit i samostatný list):

Práce se zaměřuje na predikci substrátové specifity pro rodinu haloalkan dehalogenáz. K tomuto účelu byl použit dataset 32 enzymů a 27 substrátů. Práce má spíše experimentální charakter. Student testuje různé přístupy strojového učení na podmnožinách sestaveného datasetu společně s využitím různých deskriptorů pro trénování predikčního modelu. Kromě samotného cíle, práce představuje problematiku proteinů a enzymů, experimentálního měření proteinové aktivity, dostupné databáze, způsoby vyhledávání v nich a přístupy strojového učení, které byly použité pro natrénování modelů. Všechna tato témata jsou v práci dostatečně rozebrána a doplněna o potřebné reference. Stejně tak praktická část jasně vytyčila své cíle i postup řešení a podala výsledky ve formě srozumitelné k další interpretaci. Samotná práce je psána pěknou angličtinou. Jedinou výhradu bych měl k příliš dlouhým souvětím, v důsledku kterých je text někdy obtížněji čitelný. Tento problém je patrný především v úvodu a první kapitole. Druhou výhradu bych pak měl k rozdělení textu mezi teoretickou a praktickou část. Principy konkrétních metod strojového učení jsou blíže popsány až v kapitole určené pro metody, kde se volně prolínají se samotným pracovním postupem. Teorii ohledně metod strojového učení by bylo vhodnější přesunout do předcházející kapitoly. Čtivost textu tím ale není příliš narušena a text jako celek dává smysl. Všechny

obrázky jsou dobře čitelné, řádně popsány a odkazovány v textu. Závěr práce přehledně shrnuje dosažené výsledky a navrhuje možnosti pro další pokračování projektu.

Otázky:

- Proč jste při výběru parametrů nevyzkoušeli fyzikálně-chemické vlastnosti aminokyselin, které byly použity i v jiných vámi odkazovaných pracích? Bylo by možné tyto parametry do modelu zahrnout?

- Model byl trénován na sadě haloalkan dehalogenáz. Je možné, že se v rámci trénování modelu podařilo strojovému učení odhalit vlastnosti, které se uplatňují i v jiných enzýmech? Bylo by možné model zgeneralizovat i pro další rodiny?

Závěr:

práci doporučuji k obhajobě a navrhuji známku:

Výborně (A). V rámci práce bylo provedeno velké množství experimentů, jejichž výsledky byly náležitě vylíčeny v textové zprávě. Tyto výsledky navíc mohou posloužit jako základ pro další výzkum, případně přípravu časopisecké publikace. Menší výhrady mám pouze k příliš dlouhým větám (jinak napsaným gramaticky správnou angličtinou) a k umístění popisu principů metod strojového učení do metodické části.

Navržená známka:

výborně

A výborně; B velmi dobře; C dobře; D uspokojivě; E vyhovující; F nevyhovující.
40-37 36-33 32-28 27-23 22-16 <16

V Brně dne: 31.05.2022

Ing. Miloš Musil, Ph.D.