

## Posudek vedoucího diplomové práce

**Student: Bc. Gesa-Maret Neitzert**

**Vedoucí práce: Ing. David Příhoda**

**Název práce: Enzyme optimization using sequence homology and machine learning**

Ve své práci se studentka věnovala implementaci a ohodnocení metod pro optimalizaci enzymů na základě informací z homologních sekvencí.

Její přínos spočívá jmenovitě v:

- Nalezení vhodné sady dat určené k testování implementovaných metod
- Navržení a implementace „baseline“ metody založené na přirozené frekvenci mutací získané pomocí BLAST analýzy a zarovnání sekvencí
- Navržení a implementace tradiční machine learning metody na základě biochemických příznaků
- Aplikace existující moderní metody využívající deep learning pro predikci frekvence mutací. Zde bych rád vyzvedl technickou náročnost, jelikož tyto metody nejsou připravené „ready to use“ jako tradiční bioinformatické nástroje, studentka demonstrovala samostatnost v řešení problémů s python závislostmi a výpočetní náročností metody.
- Navržení a implementace machine learning metody využívající tzv. embeddings, vektorové reprezentace sekvencí vytvořené pomocí deep learning metod
- Shromáždění kódu a výsledků v git repozitáři umožňující reprodukovatelnost výsledků pomocí skriptů a Jupyter notebooků. Jelikož tento kód byl vytvořen v rámci stáže na MSD, nebyl připojen k práci, studentce ale doporučuji umožnit komisi do kódu nahlédnout. Zde bych pouze vytkl chybějící BLAST pipeline, jelikož BLAST analýza byla provedena ručně ve webovém rozhraní.
- Srozumitelné shrnutí výsledků v textu práce a uvedení do kontextu s existující literaturou. Zde bych vytkl drobné nedostatky ve stylistice a prezentace grafů, těmto mělo být věnováno více pozornosti pro lepší porozumění z pohledu čtenáře.

Ačkoli dosažené výsledky navržených a implementovaných metod nebyly velice uspokojivé, jejich návrh a implementace byly korektní. Studentka implementovala spravedlivé srovnání s tradičními metodami, které v machine learning publikacích často chybí. Cíl práce byl velice ambiciózní, výpočetní optimalizace biologické aktivity je velice složitý problém, který stále nemá zavedený způsob řešení a je aktivně zkoumán. Práce tedy splňuje požadavky na diplomovou práci, doporučuji ji k obhajobě a navrhuji udělit 2. nejvyšší hodnocení.