

Posudek oponenta bakalářské práce

Student: Čermák Jiří

Téma: Bioinformatický nástroj pro predikci rozpustnosti proteinů (id 20048)

Oponent: Hon Jiří, Ing., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**

Cílem této práce bylo navrhnout a implementovat nástroj vhodný k predikci rozpustnosti proteinů, tj. vlastnosti klíčové pro výrobu proteinů v laboratorních podmínkách. Vzhledem k nutnosti pochopit poměrně široké téma z oblasti molekulární biologie a nastudovat řadu existujících přístupů hodnotím zadání jako obtížnější.
- 2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**

Podle zadání bych očekával, že bude nejvíce prostoru věnováno návrhu, implementaci a zhodnocení nástroje pro predikci rozpustnosti. Práce se však více zaměřuje na zhodnocení jednotlivých predikčních rysů.
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**

Odhaduji, že rozsah práce se pohybuje spíše na minimální hranici stanovené pro bakalářskou práci. Teoretické i praktické části práce by velmi prospělo, kdyby byly podrobněji rozepsány a rozšířeny o další informace.
- 4. Prezentací úroveň předložené práce** **60 b. (D)**

Ke struktuře textu mám několik výhrad: (i) V kapitole o výběru příznaků postrádám zdůvodnění, proč byly jednotlivé příznaky vybrány a proč by měly mít na rozpustnost proteinů významný vliv. (ii) Velice málo prostoru je věnováno popisu stěžejní datové sady FitSet, která je dále využita jako trénovací i testovací sada. (iii) Není jasné, jak autor získal nebo vypočetl vzdálenosti zdrojových organismu (taxonomickou, fylogenetickou podle rRNA a fylogenetickou podle ribozomů), kterou testuje jako možný rys pro predikci rozpustnosti, protože chybí citace nebo odkaz do zdrojového kódu práce. (iv) Závěr je příliš stručný a interpretaci výsledků se zabývá jen zběžně. Chybí vysvětlení, proč experimenty dopadly právě takto.
- 5. Formální úprava technické zprávy** **65 b. (D)**

V práci se vyskytuje množství jazykových a typografických chyb. Text obsahuje zbytečné anglicismy. Za všechny cituji např. *trials* nebo *thresholdem*. U převážné části grafů jsou anglické popisky, přestože je hlavní text práce psán česky.
- 6. Práce s literaturou** **85 b. (B)**

Práce je založena na 25 referencích, z nichž část se odkazuje na existující nástroje a část na kvalitní publikace z oblasti biologie. Výhradu mám jen k části textu věnované experimentu s nástrojem Tango, kde je uvedeno, že *data jsou upravena pro vizualizaci v grafu pomocí nástroje XY*. Pravděpodobně se jedná o nechtěnou chybu, ale každopádně chybí citace použitého nástroje.
- 7. Realizační výstup** **65 b. (D)**

Výstupem není ucelený nástroj, ale sada skriptů jazyce R, Python a PHP pro práci s datovými sadami eSol a TargetTrack a dále pro extrakci rysů buď přímo ze sekvence nebo nepřímo pomocí existujících nástrojů (např. BiPPred nebo Tango). V samotném kódu chybí podrobnější komentáře nebo návod na spuštění skriptů s popisem vstupů a výstupů. Struktura kódu by také mohla být lepší.
- 8. Využitelnost výsledků**

Vytvořené výstupy navazují na předchozí výsledky vytvořené v rámci projektové praxe (nástroj FitSeq). Výsledky jsou spíše experimentálního charakteru, ale přináší užitečné informace o vztahu některých vlastností proteinů k rozpustnosti.
- 9. Otázky k obhajobě**

V závěru práce naznačujete, že pro dosažení větší přesnosti prediktoru může vést upravení datové sady nebo zavedení nových příznaků. Jaké konkrétní úpravy datové sady nebo nové příznaky by to mohly být?
- 10. Souhrnné hodnocení** **65 b. uspokojivě (D)**

Silnou stránkou předložené práce je velké množství užitečných experimentů, které zkoumají možné predikční rysy vztahující se k rozpustnosti proteinů. Na druhou stranu schází dotažení do uceleného a snadno spustitelného predikčního nástroje. Technická zpráva obsahuje nedostatky jednak v oblasti struktury textu, ale i řadu typografických a jazykových chyb uvedených výše. Z těchto důvodů a s přihlédnutím k vyšší obtížnosti zadání **navrhuji hodnocení stupněm D (uspokojivě)**.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 30. května 2017

.....
podpis