

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	<b>Automatická predikce funkce bílkovin jako hierarchická multi-label klasifikace.</b>
<b>Jméno autora:</b>	<b>Martin Miadok</b>
<b>Typ práce:</b>	<b>bakalářská</b>
<b>Fakulta/ústav:</b>	<b>Fakulta elektrotechnická (FEL)</b>
<b>Katedra/ústav:</b>	<b>Katedra kybernetiky.</b>
<b>Vedoucí práce:</b>	<b>Doc. Ing. Jiří Kléma, PhD.</b>
<b>Pracoviště vedoucího práce:</b>	<b>Katedra počítačů FEL ČVUT</b>

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>průměrně náročné</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem zadání je anotovat bílkoviny popisnými termíny na základě znalosti jejich primární sekvence. Úlohu v širším kontextu pokládám za velmi náročnou. Vyžaduje základní znalost molekulární biologie a velmi dobrou znalost strojového učení. Vzhledem k tomu, že jde o bakalářskou práci, jsme od počátku po dohodě vycházeli z existujících prediktorů a hledali jejich vhodné kombinace a modifikace. Po těchto zjednodušeních práci souhrnně hodnotím jako průměrně náročnou.	

<b>Splnění zadání</b>	<b>splněno s menšími výhradami</b>
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Práce splňuje všechny psané požadavky zadání. Student se seznámil s problematikou anotace bílkovin (AFP) a nastudoval základní nehierarchické a hierarchické metody k této anotaci vhodné. Vypracoval rešerši, implementoval vlastní modifikaci hierarchického algoritmu a otestoval ji. Ke splnění každého z bodů mám výhrady (viz níže), přihlédnou-li k tomu, že jde o první větší práci autora, shrnuji zde tyto výhrady jako menší.	

<b>Aktivita a samostatnost při zpracování práce</b>	<b>D - uspokojivě</b>
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Martin Miadok na bakalářské práci aktivně pracoval především v posledním měsíci před odevzdáním. V průběhu naší spolupráce byla delší období, kdy se bakalářské práci nevěnoval, či pokroky byly velmi malé. Jde zejména o období po odevzdání projektu za zimní semestr. Postup prací tak byl spíše nárazový. Konzultace přitom byly k pokroku v řešení nutné a student byl schopen samostatné práce jen v kratších intervalech a na předem daných podúlohách. Pokud jde o vlastní návrhy řešení, držel se převážně mých pokynů.	

<b>Odborná úroveň</b>	<b>D - uspokojivě</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Student si musel řadu odborných znalostí nutných k řešení práce dostudovat. Návrh algoritmu, jeho implementace a popis vznikaly těsně před odevzdáním a je to znát na stručnosti závěrečného textu, srozumitelnosti práce i menší míře odchylky navrženého algoritmu od referenčního Valentiniho true-path rule (TPR) algoritmu. Rozdíl mezi výkonem navrženého algoritmu a BLAST-KNN je na hraně statistické signifikance, z výsledků díky zjednodušenému hodnocení bez intervalů spolehlivosti chyb o významu nelze rozhodnout. Dělení dat na trénovací, testovací a validační je správné, experimenty ale nebyly z důvodu jejich časové náročnosti opakovány.	

**Formální a jazyková úroveň, rozsah práce**

**D - uspokojivě**

*Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.*

Jak už jsem uvedl, práce je velmi stručná, informace jsou prezentovány spíše útržkovitě a pro čtenáře neseznámeného s problematikou AFP bude text pravděpodobně jen obtížně srozumitelný bez nahlédnutí do primárních zdrojů. Typograficky práce vychází z FEL ČVUT šablony, jazykově také není bez chyb.

**Výběr zdrojů, korektnost citací**

**A - výborně**

*Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

K rozsahu literatury a jejímu výběru nemám výhrady. Student odpovídajícím způsobem doporučenou literaturu zapracoval do své práce, byl schopen sám dohledat i sekundární zdroje nad rámec doporučené literatury. Student na zdroje ve své práci odpovídajícím způsobem odkazuje. Formálně by bibtex soubory mohly být jednodušší, na bakalářskou práci jde ale o drobné nedostatky.

**Další komentáře a hodnocení**

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

Vložte komentář (nepovinné hodnocení).

**III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.*

*Student splnil všechny hlavní požadavky zadání. Navrhl vlastní úpravu TPR algoritmu vhodnou pro soubory proteinových sekvencí a otestoval ji. Výsledky naznačují, že úprava patrně prakticky ani statisticky významně nepřekoná BLAST-kNN algoritmus, výkon tohoto referenčního algoritmu je ale z literatury překvapivě dobrý a vylepšení není snadné. Za nedostatky snižující hodnocení tedy považuji spíše délku a kvalitu textu práce a také formální nedostatky spojené s popisem a vyhodnocením algoritmu. Ty byly bezesporu odstranitelné při lepším rozvržení práce.*

Předloženou závěrečnou práci hodnotím klasifikačním stupněm D - uspokojivě.

Datum: 31/05/2022

Podpis: