

## МОЛБА ЗА ОДОБРАВАЊЕ ТЕМЕ МАСТЕР РАДА

Молим да ми се одобри израда мастер рада под насловом:

*Предвиђање функција протеина методама бинарне класификације*

### Значај теме и области:

Протеини су макромолекули који играју важну улогу у организму сваког живог бића. То су биолошки најактивнији молекули неопходни за изградњу и функционисање ћелија и имају велики број есенцијалних функција. Структура протеина зависи од распореда аминокиселина и утиче на његову функцију. Примарна структура обухвата секвенцу аминокиселина које изграђују протеин. Она је основни извор информација о протеину и његовој функцији. Иако је познавање секвенце прилично корисно у многим пољима, као што су филогенетика и еволуциона биологија, за разумевање свих процеса који се дешавају у ћелији потребно је знање и о функцији протеина. Најчешћи начин за представљање функције протеина је преко система Gene Ontology. Овај систем дели функције протеина на три онтологије: биолошки процеси, молекулске функције и ћелијске компоненте. Свака онтологија представља усмерени ациклички граф где су чворовима придружени називи функција, и то сваком чвору специфичнија функција од функције његовог родитељског чвора. На тај начин, укупна функција једног протеина представља подграф онтологије који садржи све чворове означене појединачним функцијама које протеин обавља. Познато је да протеини са сличним примарним структурама теже да обављају исте функције.

### Специфични циљ рада:

Сваке године секвенцира се велики број нових генома чиме расте број новооткривених протеина. Функција протеина одређује се експериментално што је скуп и спор процес због чега се улаже труд у развој рачунарских метода које могу да предвиде функцију протеина. Циљ овог рада је развој алата за одређивање функције протеина на основу његове примарне структуре помоћу метода бинарне класификације као што су метод потпорних вектора, случајне шуме и логистичка регресија обучених на скупу протеина са експериментално утврђеним функцијама. Резултати добијени различитим методама ће бити поређени међусобно као и са актуелним алатима за предвиђање функције протеина на одговарајућим benchmark скуповима.

### Остале битне информације:

За израду софтвера биће коришћен програмски језик Python. Биће коришћене имплементације наведених метода из библиотеке sklearn. Онтологија функција преузета је са адресе <http://geneontology.org/docs/download-ontology/> у .obo формату, а база протеина са експериментално утврђеним функцијама са адресе [https://biofunctionprediction.org/cafa-targets/CAFA3\\_training\\_data.tgz](https://biofunctionprediction.org/cafa-targets/CAFA3_training_data.tgz).

Ања Букуров, 1082/2016, информатика

(име и презиме студента, бр. индекса, модул)

Сагласан ментор доц. др Јована Ковачевић

(својеручни потпис студента)

(својеручни потпис ментора)

27.8.2019.

(датум подношења молбе)

Чланови комисије

1. проф. др Гордана Павловић-Лажетић
2. доц. др Младен Николић

Катедра за рачунарство и информатику

је сагласна са предложеном темом.

(шеф катедре)

(датум одобравања молбе)