

Семинарски рад из Статистичког софтвера 3

Лука Перовић

2. јануар, 2026.

Дата је табела са информацијама о свим огласима за продају станова у граду Београду, прикупљених 29.12.2025. године. Примарни циљеви су да се на примеру реалних података анализа цена станова, као и да се на основу анализе текста извуку битне информације.

Подаци су доступни у фајлу `oglas.csv`.

СЕМИНАРСКИ РАД - ЗАДАЦИ

ЗАДАТАК 1 - Одвајање тест скупа [5%]. Одвојити опсервације на тренинг и на тест скуп. При подели је препоручено да се цена стана узима у обзир за репрезентативнију поделу. Обратити пажњу на проценат опсервација који се одваја за тест скуп - узети у обзир укупан обим података.

ЗАДАТАК 2 - Припрема података [20%]. У одређеним обележјима су подаци дати у `json` формату - истражити како се врши трансформација из датог формату у табеларни и извршити је. Постоје редундантна обележја, као и неинтерпретабилне вредности. Избацити непотребне колоне. Обратити пажњу на недостајуће вредности и проценити шта има смисла избацити, а шта треба импутирати.

ЗАДАТАК 3 - Основни модел цене стана [25%]. Направити модел који предвиђа цену стана у односу на остале факторе. Издвојити и приказати станове који значајно одступају од предвиђања модела. Препоручено је да модел буде интерпретабилан и да се изврши анализа утицаја различитих предиктора. Извршити евалуацију модела на тест скупу - размислити о томе које су метрике перформанси најсмисленије за ову примену.

ЗАДАТАК 4 - Бинарна класификација општине у односу на текст [20%]. Циљ овог задатка је да се понашамо као да немамо структуриране податке о локацији стана, већ морамо да пробамо да их извучемо из текста.

Извршити обраду и анализу описа стана - трансформисати у табеларни облик који је даље могуће користити за прављење модела. Направити модел бинарне класификације који предвиђа, само на основу текста, да ли стан припада једној од централних општина (Стари град, Савски венац, Врачар). Колико је прецизан модел? Евалуирати модел стандардним метрикама на тест скупу. При анализи обратити пажњу на то да укупно има мање огласа централних општина него других.

ЗАДАТАК 5 - Модел цене стана припомогнут тесктуалним подацима [15%]. Искористити обраду података из претходног задатка и направити модел који предвиђа цену стана који, осим стандардних предиктора, узима у обзир и обрађен текст. Да ли је могуће извући из текста додатне факторе који значајно утичу на цену? Извршити евалуацију на тест скупу и упоредити са моделом из задатка 3.

ЗАДАТАК 6 - Кластеризација [15%]. Методама невођеног учења извршити поделу станова на кластере. Извршити анализе кластера - Да ли може да се протумачи смисао из њих?

СЕМИНАРСКИ РАД - НАПОМЕНЕ

НАПОМЕНА 1. Семинарски рад се ради самостално и вреди 70 поена, уз то да одбрана рада носи половину тих поена. Одбране се организују уживо - на њима вас испитујем о томе шта сте радили и о градиву курса. Рок за израду је крај школске године. У зависности од рока у којем желите да браните семинарски рад, накнадно ћемо се договарати о термину када пошаљете рад. Радове слати на мејл luka.perovic@matf.bg.ac.rs

НАПОМЕНА 2. Није довољно послати само код, већ је неопходно детаљно исписати шта је урађено. Опције су: (1) RMarkdown документ, у коме би били интегрисани текст и код заједно; (2) LaTeX документ, у коме би било описано шта је рађено, а искоментарисан код да се пошаље посебно као обичан .R фајл; Правите се да пишете извештај послодавцу. Дакле, нек буде шарено и лепо, нико не воли да чита досадан текст. Препоручена је опција 1 због прегледности кода.

НАПОМЕНА 3. Додатни труд при анализи података и оптимизацији модела ће бити награђена.

НАПОМЕНА 4. Дозвољено је коришћење свих средстава, осим преписивања од колега. Дакле гуглање, књиге... Дозвољено је питати и колегу са курса, уколико сте у стању да испродукујете два различита решења. Битно је разумевање урађеног.

НАПОМЕНА 5. Дозвољено је писање семинарског рада и на енглеском.

НАПОМЕНА 6. Путем мејла можете постављати питања у вези са задацима. Уколико прогуглам ваше питање и нађем одговор међу првим резултатима, нећете добити одговор. У свим другим ситуацијама бићу више него рад да помогнем.