

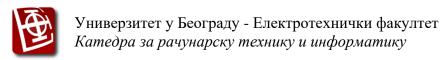
Трећи пројектни задатак из предмета Интелигентни системи школске 2022/23. године

Основне информације

Циљ пројектног задатка је реализација модела линеарне регресије за предвиђање CO2 емисије аутомобила и модела за класификацију колача користећи алгоритам к-најближих суседа.

Решење сваког од алгоритама треба да прати следеће кораке:

- Учитавање скупа података и приказ првих пет редова у табели.
- Приказ концизних информација о садржају табеле и статистичких информација о свим атрибутима скупа података.
- Елиминисање примерака са недостајућим вредностима атрибута или попуњавање недостајућих вредности на основу вредности атрибута осталих примерака.
- Графички приказ зависности континуалних атрибута коришћењем корелационе матрице.
- Графички приказ зависности излазног атрибута од сваког улазног континуалног атрибута расејавајући тачке по Декартовом координатном систему.
- Графички приказ зависности излазног атрибута од сваког улазног категоричког атрибута користећи одговарајући тип графика.
- Одабир атрибута који учествују у тренирању модела.
- Извршавање додатних трансформација над одабраним атрибутима.
- Формирање тренинг и тест скупова података.
- Релизација и тренирање модела користећи све наведене приступе.
- Приказ добијених параметара модела, вредности функције грешке и прецизности модела за све реализоване приступе.
- Приказати додатна интересантна запажања уочена током реализација модела.



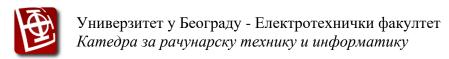
1. *Линеарна регресија* [5 поена]

Неопходно је реализовати алгоритам линеарне регресије са више атрибута. Скуп података који се користи ($co2_emissions.csv$) је дат на предметном сајту. Атрибути скупа података возила су следећи:

- *MODELYEAR* година производње.
- МАКЕ назив произвођача.
- МОДЕ ознака модела.
- VEHICLECLASS врста каросерије.
- *ENGINESIZE* запремина мотора.
- *CYLINDERS* број цилиндара мотора.
- TRANSMISSION ознака мењача.
- *FUELTYPE* тип горива.
- FUELCONSUMPTION_CITY потрошња горива у граду.
- FUELCONSUMPTION_HWY потрошња горива на отвореном путу.
- FUELCONSUMPTION_COMB комбинована потрошња горива.
- *FUELCONSUMPTION_COMB_MPG* комбинована потрошња горива (друга мерна јединица).
- *CO2EMISSIONS* емисија CO2.

Потребно је реализовати следеће приступе:

- Сопствени алгоритам линеарне регресије са градијентним спустом за одређивање оптималних параметара модела (без коришћења уграђених модела).
- Уграђени модел алгоритма линеарне регресије.



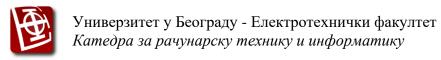
2. Кпп (К-најближих суседа) алгоритам [5 поена]

Неопходно је реализовати модел алгоритма к-најближих суседа са више атрибута. Скуп података који се користи (cakes.csv) је дат на предметном сајту. Атрибути скупа података колача су следећи:

- *flour* количина брашна коришћена за колаче изражена у грамима.
- eggs број јаја коришћена за колаче (сматрати да једно јаје има масу од 63 грама).
- sugar количина шећера коришћена за колаче изражена у грамима.
- *milk* количина млека коришћена за колаче изражена у грамима.
- *butter* количина путера коришћена за колаче изражена у грамима.
- baking_powder количина прашка за пециво коришћена за колаче изражена у грамима.
- *type* тип колача који може бити *muffin* или *cupcake*.

Потребно је реализовати следеће приступе:

- Сопствени алгоритам к-најближих суседа (без коришћења уграђених модела).
- Уграђени модел алгоритма к-најближих суседа.



Напомене

Електронску верзију решења овог домаћег задатка послати најкасније до 29.1.2023. године до 15:00, као ZIP архиву на следећем <u>линку</u>. Одбрана домаћег задатка планирана је 30.1.2023. године.

Домаћи задатак из предмета Интелигентни системи се ради самостално и није обавезан за полагање испита (на испиту се може заменити са теоријским питањима из целокупног градива). Овај домаћи задатак се може бранити само у за то предвиђеном термину. Домаћи задатак вреди максимално 10 поена (уз могућност додатних поена за радове који се посебно истакну).

Решење домаћег задатка реализовати у програмском језику Пајтон уз дозвољено коришћење стандардних библиотека и структура података, *sklearn*, *nympy*, *pandas*, *matplotlib* и *seaborn* библиотека.

На усменој обрани кандидат мора самостално да покрене своје решење. Кандидат мора да поседује потребан ниво знања о задатку, мора да буде свестан недостатака приложеног решења и могућности да те недостатке реши. Кандидат мора тачно да одговори и на одређен број питања која се баве тематиком домаћег задатка.

Евентуална питања послати асистентима на мејл, али као једну поруку, а не две одвојене (другог асистента обавезно ставити у копију - СС поруке).

aki@etf.rs
jocke@etf.rs