

Увод у вештачку интелигенцију

Поправни II колоквијума: Машинско учење – 7 поена

20.01.2022.

Цене дијаманата

На основу историјских података продаји дијаманата креирати моделе машинског учења којима се може проценити цена дијаманата. Доступна су два сета података. Сет податка `train.csv` намењен је обуци модела и њиховим оценама. На другом сету података треба одредити непознате цене дијаманта.

Потребно је извршити следеће кораке:

1. Учитати сет податка `train.csv`.
2. Проверити да ли неке вредности података недостају. Уколико недостаје део података, попунити просечним вредностима за тај тип података уколико се ради о нумеричким вредностима, односно најчешћом категоријом уколико се ради о категоријским варијаблама.
3. Све категоријске варијабле кодирати методом One-Hot-Encoding.
4. Поделити податке на скуп за обуку и скуп за тестирање у односу 70:30
5. Потребно је креирати 2 модела машинског учења:
 - a. Креирати и обучити модел вишеструке линеарне регресије чији су улази `x`, `y` и `z`, а излаз вредност `depth`.
 - b. Креирати и обучити модел вештачке неуронске мреже којим се одређује цена дијаманата. Све преостале варијабле чине улазе у мрежу. Архитектура мреже садржи 2 скривена слоја са по 7 неурона.
6. Извршити предикције на тестном скупу.
7. Потом одредити метрике оба модела на тестном скупу:
 - a. Модел линеарне регресије оцијенити грешком MAE
 - b. Модел вештачке неуронске мреже оцијенити грешком RMSE
8. Учитати сет података `to_predict.csv`.
9. Недостајуће вредности у колони `depth` попунити одговарајућим предикцијама модела вишеструке линеарне регресије.
10. Извршити потребну припрему сета података да би се применио модел вештачке неуронске мреже.
11. Одредити цене дијаманта. Добијене предикције, заједно с улазним подацима, сачувати у фајл `predictions.csv`.

НАПОМЕНА: Свој рад предати као *python* или *jupyter notebook* фајл и сачувати на адреси

- `~\Desktop\Rad\ImePrezimeBrojIndeksa\`

Време рада – 45min.