

Završni zadatak, PR-P-LV2

Upute za rješavanje

1. Riješite dane zadatke koristeći programski jezik Python te pripadajuće biblioteke i metode strojnog učenja koje ste učili u okviru laboratorijskih vježbi. Zadatke rješavajte u skripti prezime_ispit.py.
2. Rješenja zadataka, u obliku skripte prezime_ispit.py, kao i popunjeni list za odgovore u obliku prezime_ispit.pdf, uploadajte na Merlin u jednom .zip folderu naziva prezime_ispit.

Zadatak 0.0.4 Datoteka titanic.csv sadrži podatke o putnicima broda Titanic, koji je potonuo 1912. godine. Upoznajte se s datasetom i dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Za koliko žena postoje podatci u ovom skupu podataka?
- b) Koliki postotak osoba nije preživio potonuće broda?
- c) Pomoću stupčastog dijagrama prikažite postotke preživjelih muškaraca (zelena boja) i žena (žuta boja). Dodajte nazine osi i naziv dijagrama. Komentirajte korelaciju spola i postotka preživljavanja.
- d) Kolika je prosječna dob svih preživjelih žena, a kolika je prosječna dob svih preživjelih muškaraca?
- e) Koliko godina ima najstariji preživjeli muškarac u svakoj od klasa? Komentirajte.

Zadatak 0.0.5 Datoteka titanic.csv sadrži podatke o putnicima broda Titanic, koji je potonuo 1912. godine. Upoznajte se s datasetom. Učitajte dane podatke. Podijelite ih na ulazne podatke X predstavljene stupcima Pclass, Sex, Fare i Embarked i izlazne podatke y predstavljene stupcem Survived. Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 70:30. Izbacite izostale i null vrijednosti. Skalirajte podatke. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Izradite algoritam KNN na skupu podataka za učenje (uz K=5). Vizualizirajte podatkovne primjere i granicu odluke.

- b) Izračunajte točnost klasifikacije na skupu podataka za učenje i skupu podataka za testiranje. Komentirajte dobivene rezultate.
- c) Pomoću unakrsne validacije odredite optimalnu vrijednost hiperparametra K algoritma KNN.
- d) Izračunajte točnost klasifikacije na skupu podataka za učenje i skupu podataka za testiranje za dobiveni K . Usporedite dobivene rezultate s rezultatima kada je $K=5$.

Zadatak 0.0.6 Datoteka `titanic.csv` sadrži podatke o putnicima broda Titanic, koji je potonuo 1912. godine. Upoznajte se s datasetom. Učitajte dane podatke. Podijelite ih na ulazne podatke X predstavljene stupcima `Pclass`, `Sex`, `Fare` i `Embarked` i izlazne podatke y predstavljene stupcem `Survived`. Podijelite podatke na skup za učenje i skup za testiranje modela u omjeru 75:25. Izbacite izostale i null vrijednosti. Skalirajte podatke. Dodajte programski kod u skriptu pomoću kojeg možete odgovoriti na sljedeća pitanja:

- a) Izgradite neuronsku mrežu sa sljedećim karakteristikama:
 - model očekuje ulazne podatke X
 - prvi skriveni sloj ima 12 neurona i koristi `relu` aktivacijsku funkciju
 - drugi skriveni sloj ima 8 neurona i koristi `relu` aktivacijsku funkciju
 - treći skriveni sloj ima 4 neurona i koristi `relu` aktivacijsku funkciju
 - izlazni sloj ima jedan neuron i koristi `sigmoid` aktivacijsku funkciju.Ispišite informacije o mreži u terminal.
- b) Podesite proces treniranja mreže sa sljedećim parametrima:
 - **loss** argument: `binary_crossentropy`
 - **optimizer**: `adam`
 - **metrika**: `accuracy`.
- c) Pokrenite učenje mreže sa proizvoljnim brojem epoha (pokušajte sa 100) i veličinom `batch-a` 5.
- d) Pohranite model na tvrdi disk te preostale zadatke izvršite na temelju učitanog modela.
- e) Izvršite evaluaciju mreže na testnom skupu podataka.
- f) Izvršite predikciju mreže na skupu podataka za testiranje. Prikažite matricu zabune za skup podataka za testiranje. Komentirajte dobivene rezultate i predložite kako biste ih poboljšali, ako je potrebno.