

Istraživanje podataka - primer praktičnog dela ispita

1. Koristeći skup podataka *car.csv* izvršiti klasifikaciju metodom SVM u alatu KNIME. U okviru modela uraditi:

- Primeniti tehniku PCA radi smanjenja dimenzija skupa podataka.
- Izdvojiti izveštaj o klasifikaciji prema preciznosti i prikazati matricu konfuzije za trening i test skup.
- Prikazati grafički test podatke pomoću grafika sa razbacanim elementima (eng. scatter). Koristiti 2 atributa koja nose najviše informacija o promenljivosti u skupu.

U komentarima:

- Na koji broj atributa je smanjen skup korišćenjem PCA tehnike i zašto?
- Opisati dobijeni model.

Radni tok eksportovati i rezultat imenovati u formatu **KNIME_klasifikacija_vasBrojIndeksa**.
Odgovore pišite u datoteci sa nazivom **KNIME_klasifikacija_vasBrojIndeksa_odgovori**.

2. Skup podataka *oscar.csv* sadrži podatke o tvitovima koji se odnose na dodelu Oskara 2017. Koristeći alat KNIME izdvojiti pravila pridruživanja o rečima iz tvitova. Prilikom pravljenja modela zadati sledeće uslove:

- najmanja podrška je 1%.
- najmanja pouzdanost je 50%.

Odgovoriti na pitanja:

- Koliko je čestih skupova stavki u modelu?
- Kolika je najmanja, a kolika najveća podrška čestih skupova stavki?
- Koji je najveći čest skup stavki?
- Koliko je pravila pridruživanja u modelu?
- Kolika je najmanja, a kolika najveća pouzdanost pravila?
- Koje pravilo je najzanimljivije po Lift meri? Objasniti zašto.

U okviru modela uraditi:

- Napraviti tabelu koja sadrži pravila koja su nabolja prema podršci ili Lift meri. Za svako pravilo izdvojiti: podršku, pouzdanost, Lift meru, telo pravila i glavu pravila.
- Napraviti tabelu koja sadrži pravila koja ne sadrže reč *oscar*.
- U datoteku *pravila.csv* izdvojiti sva dobijena pravila.

Radni tok eksportovati i rezultat imenovati u formatu **KNIME_pravila_vasBrojIndeksa**.
Odgovore pišite u datoteku sa nazivom **KNIME_pravila_vasBrojIndeksa_odgovori**.

3. Skup podataka *diamonds.csv* sadrži podatke o dijamantima. Koristeći skup i alat IBM SPSS Modeler izvršiti klasterovanje nad skupom primenom algoritama K-sredina i Kohonen za 4 klastera.

U komentarima odgovoriti na pitanja:

- Koliki je kvalitet dobijenih modela?
- Kako ste postavili parametre? Zašto?
- Koji atributi su najznačajniji za pravljenje modela?
- Uporediti najveće klustere iz dobijenih modela.

Radni tok eksportovati i dodeliti mu ime u formatu **SPSS_klasterovanje_vasBrojIndeksa**.
Odgovore pišite u datoteci sa nazivom **SPSS_klasterovanje_vasBrojIndeksa_odgovori**.

Uputstvo za čuvanje rada: Na Desktopu napravite direktorijum sa nazivom u formatu **ip.xxx.2017.ime.prezime.brojIndeksa** gde umesto ime, prezime i broj indeksa stavite vaše podatke. Npr, **ip.xxx.2017.petar.petrovic.543_2014**. U tom direktorijumu čuvajte rešenja zadataka i datoteke sa odgovorima.