Redukcija dimenzionalnosti

Tim 6_23

- Anastasija Samčović, SW44/2019
- Strahinja Popović, SW51/2019
- Srđan Đurić, SW63/2019

Obrada podataka

Prvi korak u rešavanju datog problema bila je vizualizacija i analiza datog trening skupa podataka po obeležjima. Primetili smo postojanje numeričkih obeležja(masa, godina, starost) i kategoričkih(jastuk, ceoni, pojas, tip, pol).

Uklonili smo obeležje “ishod”, jer je ono ciljna varijabla. Eksperimentalnim putem smo zaključili da obeležje pol ne igra bitnu ulogu u predikciji ishoda. Nakon uklanjanja ovog obeležja zabeležili smo značajno poboljšanje performansi.

Budući da je skup podataka sadržao nedostajuće vrednosti, sve redove koji imaju više od 2 obeležja koja su nedostajuća smo odbacili. Ostatak praznih obeležja smo popunili koristeći “median” strategiju za numerička i “most frequent” strategije za kategorička obeležja.

Za kategorička obeležja smo koristili OneHotEncoding, a na numerička obeležja smo primenili StandardScaler.

Rešenje

Koristili smo Ansambl metod uz korišćenje Ada Boost klasifajera. Testirani su RandomForest i DecisionTree algoritmi. Ustanovili smo da oba pristupa daju slične rezultate, za krajnji pristup smo se odlučili za DecisionTree algoritmom.

Takođe je primenjena implementacija PCA iz biblioteke scikit-learn. Eksperimentalnim putem smo uočili da naš algoritam radi najbolje sa parametrom n_components=4. Isprobali smo i KernelPCA koji je davao znatno lošije rezultate i duže se izvršavao.

Poziv:

```
ada_clf = AdaBoostClassifier( base_estimator=DecisionTreeClassifier(max_depth=10,
min_samples_split=4, random_state=42, class_weight='balanced'),
learning_rate=0.5, n_estimators=150, random_state=42, algorithm="SAMME"
```

U Tabela 1 su prikazani rezultati dobijeni korišćenjem DecisionTree i RandomForest algoritama.

Skup podataka korišćen za testiranje predstavlja 20% datog trening skupa, a skup podataka korišćen za treniranje modela predstavlja 80% datog trening skupa.

algoritam	**rezultat (80/20)**
RandomForest	0.319
DecisionTree	**0.325**

Tabela 1 Prikaz rezultata