

Извештај за задатак 4 – модел ансамбла

Даница Газдић
SV 12/2020

Милош Обрадовић
SV 55/2020

1. Проблем

У овом извештају се решава проблем класификације моделом ансамбла над проблемом идентификовања жанра игрице. Задатак је решен употребом два модела: *Random Forest* и *Gradient Boosting*, комбинованим у *VotingClassifier* ансамбл.

2. Решење

Што се претпроцесирања тиче, недостајуће вредности су предиктоване уз помоћ *KNNImputer*. Категоријска обележја су енкодирана уз помоћ *One-Hot* енкодинга. У првом покушају су модели тренирани са подразумеваним хиперпараметрима и добили смо ф1 макро 0.41863. Затим смо претражили најбољу комбинацију хипер параметара уз помоћ *GridSearchCV*.

```
param_grid = {  
    'rf__n_estimators': [50, 100, 150, 200, 250, 300],           # 6  
    'rf__max_depth': [None, 5, 10, 15, 20, 25, 30],             # 7  
    'rf__min_samples_split': [2, 3, 4, 5, 6, 7, 8, 9, 10],      # 9  
    'gb__n_estimators': [100, 150, 200, 250, 300],             # 5  
    'gb__learning_rate': [0.01, 0.05, 0.1, 0.15, 0.2, 0.3]      # 6  
}  
grid_search = GridSearchCV(voting_clf, param_grid, cv=5, scoring='f1_macro', n_jobs=-1)  
grid_search.fit(X_train, y_train)
```

Најбоље решење (0.4318) добили смо користећи:

```
rf_n_estimators = 50  
rf_max_depth = 20  
rf_min_samples_split = 7  
gb_n_estimators = 100  
gb_learning_rate = 0.3
```