Kaggle 발표

개요

타이타닉 승객 데이터를 활용, 데이터 분석을 통하여 머신 러닝 모델을 생성 후 승객의 생존유무 예측

Kaggle data, Feature engineering.

코드를 필사.

```
[8] df_train.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S

```
[72] df_test.head(3)
```

	Pclass	Sex	FamilySize	Age_cat	Fare_cat	Embarked_0	Embarked_1	Embarked_2	Initial_0	Initial_1	Initial_2	Initial_3	Initi
0	3	1	1	2	0	0	1	0	0	0	1	0	
1	3	0	2	2	0	0	0	1	0	0	0	1	
2	2	1	1	3	0	0	1	0	0	0	1	0	

```
time: 33.4 ms
```

사용모델.

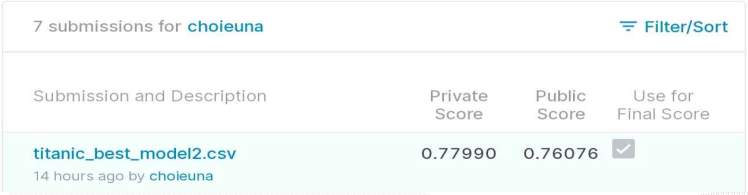
XGBoost(파라미터 변경 X)

1. XGBoost

```
83] %%time

    import xgboost as xgb

    model_xgb = xgb.XGBClassifier(max_depth=9, learning_rate=0.01, n_estimator
                                  colsample_bytree=0.9, subsample=0.9, n_jobs=5
    model_xgb.fit(X_tr, y_tr, eval_set=[(X_vld, y_vld)], verbose=False, early_
    pred_xgb = model_xgb.predict(X_vld)
    score_xgb = metrics.accuracy_score(pred_xgb, y_vld)
    print("XGBoost Test score: ", score_xgb)

    XGBoost Test score:  0.8694029850746269
```

결과 스코어.

◆ private / public



소감

◆감사합니다.