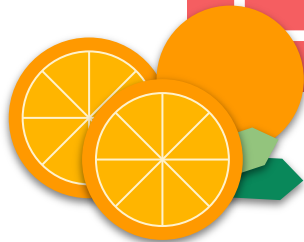


감귤 착과량 예측 AI 경진대회

팀명: 감귤없이잘살아



목차

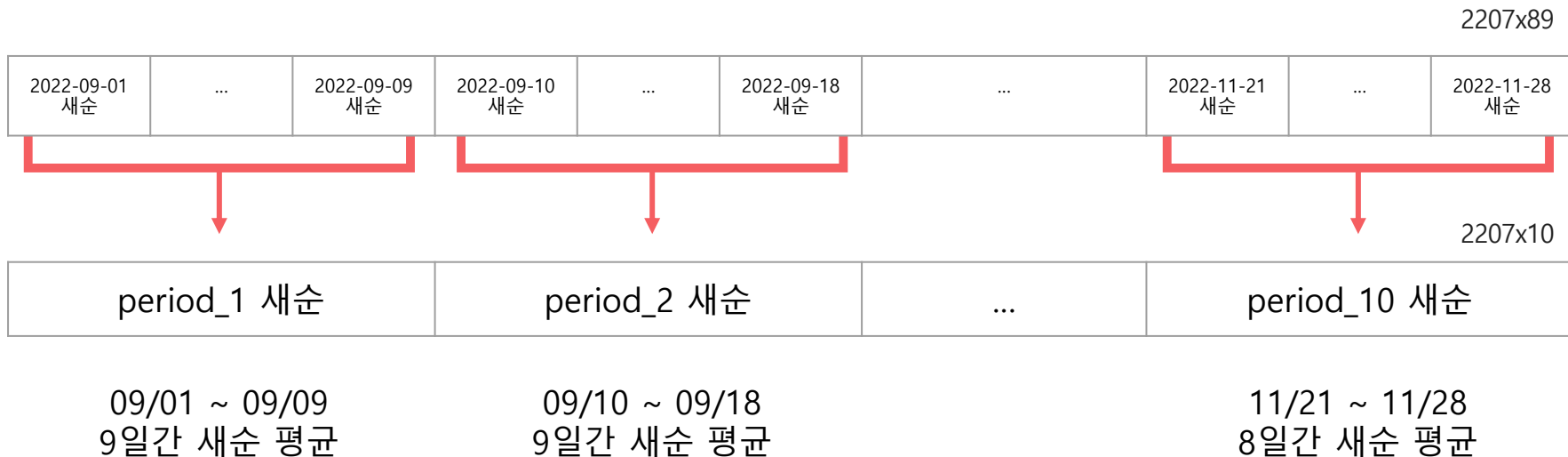
1. 전처리 및 파생 변수 생성
2. 모델 학습 및 검증
3. 결과
4. 시도한 모델

[1] 전처리 및 파생 변수 생성

새순

9일간 새순 평균

2022-09-01 ~ 2022-11-28 89일 간의 새순 feature 89개를 Period 10개로 변환



※ 마지막 period_10만 8일 평균

[1] 전처리 및 파생 변수 생성

새순

최댓값, 최솟값, 범위

새순 최댓값, 새순 최솟값, 새순 범위 파생 변수 생성

| | |
|-------------|---------------|
| 새순max (최댓값) | 2022-09-01 새순 |
| 새순min (최솟값) | 2022-11-28 새순 |
| 새순diff (범위) | 새순max-새순min |

※ 9/1 새순이 대체로 ID별 새순 최댓값이며, 11/28 새순이 대체로 ID별 새순 최솟값으로 시간에 따라 새순 변수가 점차 감소함을 반영

[1] 전처리 및 파생 변수 생성

나무 생육 상태

부피

수고와 수관폭을 곱하여 나무 부피와 관련된 파생 변수 생성

나무 부피

수고(m) * 수관폭평균(max와 min의 평균)

※ 나무의 수고율(수고/수관폭평균) 변수도 시도했으나
나무 부피 변수를 사용한 모델 성능이 더 좋았음

[2] 모델 학습 및 검증

RandomForest

```
param_grid = {  
    'max_depth': [3, 4, 5, 6, 7],  
    'bootstrap': [True, False],  
    'max_features': ['auto', 'sqrt', 'log2'],  
    'criterion': ['squared_error', 'absolute_error', 'friedman_mse', 'poisson']  
}  
  
ss=ShuffleSplit(test_size=0.3,random_state=0)  
grid_rf_cv=GridSearchCV(rf,  
    return_train_score=True,  
    param_grid=param_grid,  
    cv=ss,  
    verbose=0,  
    scoring='neg_mean_absolute_error',  
    n_jobs=-1)
```

최적의 매개변수 조합: {'bootstrap': False,
'criterion': 'friedman_mse', 'max_depth': 6, 'max_features': 'sqrt'}

최고의 교차 검증 점수: -30.18017600138905
NMAE : 0.06610512938333356



XGboost

```
param_grid = {  
    'learning_rate':[0.001,0.01,0.1,0.2],  
    'max_depth': [3, 4, 5, 6, 7]  
}  
  
ss=ShuffleSplit(test_size=0.3,random_state=0)  
grid_xgb_cv=GridSearchCV(xgbr,return_train_score=True,  
    param_grid=param_grid,  
    cv=ss,  
    verbose=0,  
    scoring='neg_mean_absolute_error',  
    n_jobs=-1)
```

최적의 매개변수 조합: {'learning_rate': 0.1, 'max_depth': 3}

최고의 교차 검증 점수: -30.6408947532533
NMAE : 0.06664120807503442

[2] 모델 학습 및 검증

| 모델 | Public Score (제출 모델 내 순위) | Private Score |
|-----------------------------|---------------------------|---------------|
| Xgboost 단일모델 | 0.0730279607 (4) | |
| RandomForest 단일모델 | 0.0730401155 (5) | |
| 0.6*Xgboost + 0.4*RF | 0.0728737434 (2) | |
| 0.4*Xgboost + 0.6*RF | 0.0728737434 (2) | |
| 0.5*Xgboost + 0.5*RF | 0.0728735117 (1) | 0.7286 |

0.5* Pred_xgb + 0.5*Pred_rf 를 최종 제출파일로 선택

[3] 결과

| 사용한 피쳐 | |
|--------------------|--|
| period 1~ 10 새순 | |
| 새순 max | |
| 새순 min | |
| 새순 diff (max-min) | |
| 나무 부피 (수고 x 수관폭평균) | |

| 사용한 모델 | |
|-------------------------------------------------------------------------------------------------------|--|
| RandomForestRegressor | |
| param : 'bootstrap': False, 'criterion': 'friedman_mse', 'max_depth': 6, 'max_features': 'sqrt' | |
| XGBoostRegressor | |
| param : {'learning_rate': 0.1, 'max_depth': 3} | |

| 최종 제출 | |
|-----------------------------|------------------------|
| 0.5* Pred_xgb + 0.5*Pred_rf | Private Score : 0.7286 |

[4] 시도한 모델

모든 변수 포함

- 나무 생육 상태 변수 (5개)
- 새순 변수 (89개)
- 엽록소 변수(89개)

모든 변수 + 수고율

- 나무 생육 상태 변수 (5개)
- 새순 변수 (89개)
- 엽록소 변수(89개)
- 수고율 (수관폭평균/수고)

나무 생육 상태 변수 + period 변수

- 나무 생육 상태 변수 (5개)
- period_1 ~ period_10 새순 변수 (10개)
- period_1 ~ period_10 엽록소 변수 (10개)

엽록소 제거

- 나무 생육 상태 변수 (5개)
- 새순 분산 (1개)

감사합니다

팀명: 감귤없이잘살아
(사실 못살아)

