

결과 보고서

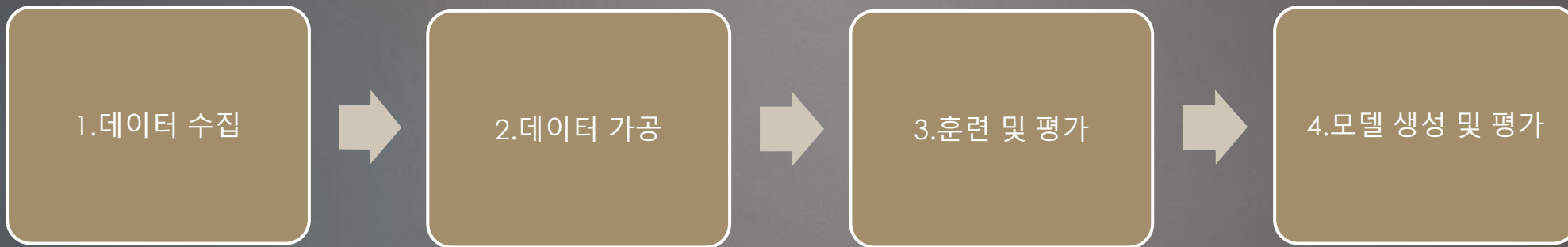
박승원

1.목적

1.프로젝트 목적

1) 머신 러닝을 사용해 최저가 예측 진행

2.순서



2.순서-데이터 수집 (2021.08.23~2021.11.05)

pid	year	week_year	day_week	car_year	holiday_length	duration	howearly	i1	i2	model	agency	price
1	2021	0	4	2020	3	1	0	1	0	NM_HD_PALISADE7_A_D	NM_CENTRALCHEONGJU	210000
2	2021	0	4	2018	3	1	0	1	0	NM_KI_SORENT5_A_D_AN	NM_CENTRALCHEONGJU	110000
3	2021	0	4	2017	3	1	0	1	0	NM_KI_K3_A_G_AN	NM_CENTRALCHEONGJU	70000
4	2021	0	4	2018	3	1	0	1	0	NM_KI_K3_A_G_AN	NM_CENTRALCHEONGJU	70000
5	2021	0	4	2017	3	1	0	1	0	NM_HD_AVANTE_A_L_MD	NM_CENTRALCHEONGJU	70000
6	2021	0	4	2017	3	1	0	1	0	NM_KI_K5_A_L_AN	NM_CENTRALCHEONGJU	80000
7	2021	0	4	2018	3	1	0	1	0	NM_KI_K5_A_L_AN	NM_CENTRALCHEONGJU	80000
8	2021	0	4	2020	3	1	0	1	0	NM_KI_CARNI9_A_D_TN	NM_CENTRALCHEONGJU	210000
9	2021	0	4	2018	3	1	0	1	0	NM_KI_CARNI_A_D_AN1	NM_CENTRALCHEONGJU	210000
10	2021	0	4	2019	3	1	0	1	0	NM_HD_SANTAF5_A_D_TM	NM_CENTRALCHEONGJU	160000
11	2021	0	4	2019	3	1	0	1	0	NM_KI_K7_A_L_PM	NM_CENTRALCHEONGJU	130000
12	2021	0	4	2019	3	1	0	1	0	NM_HD_GRANDEUR_A_L_I	NM_CENTRALCHEONGJU	130000
13	2021	0	4	2017	3	1	0	1	0	NM_KI_MORNING_A_G_AN	NM_CENTRALCHEONGJU	60000
14	2021	0	4	2017	3	1	0	1	0	NM_KI_K7_A_L_AN	NM_CENTRALCHEONGJU	110000

1.데이터 수집

1)찜 카 API를 통해 확보한 렌터카 대여일.구매일.에이전시.모델.가격.보험 정보 등을 수집한 정보를 데이터 베이스에 저장(가공 전 데이터)

2. 데이터 가공 자동화 프로그램 구축

1)서버 와 데이터 가공 자동화 프로그램을 직접 제작 해서 프로젝트 진행

2.순서-데이터 가공(2021.08.23~2021.11.05)

pid	year	week_year	day_week	car_year	holiday_length	duration	howearly	i1	i2	class	engine_cc	seat_cnt	fuel_G	fuel_L	fuel_D	fuel_E	fuel_H	total_car	price
1	2021	0	4	2017	3	1	0	1	0	4	1999	5	0	1	0	0	0	0	80000
2	2021	0	4	2018	3	1	0	1	0	4	1999	5	0	1	0	0	0	0	80000
3	2021	0	4	2017	3	1	0	1	0	1	998	5	1	0	0	0	0	0	60000
4	2021	0	4	2017	3	1	0	1	0	5	3000	5	0	1	0	0	0	0	110000
5	2021	0	4	2017	3	1	0	1	0	4	1999	5	0	1	0	0	0	0	80000
6	2021	0	4	2019	3	1	0	1	0	3	1598	5	1	0	0	0	0	0	75000
7	2021	0	4	2020	3	1	0	1	0	3	1598	5	1	0	0	0	0	0	75000
8	2021	0	4	2019	3	1	0	1	0	1	998	5	1	0	0	0	0	0	65000
9	2021	0	4	2017	3	1	0	1	0	3	1591	5	1	0	0	0	0	0	55000
10	2021	0	4	2017	3	1	0	0	1	3	1591	5	1	0	0	0	0	0	65000
11	2021	0	4	2018	3	1	0	1	0	3	1591	5	1	0	0	0	0	0	55000
12	2021	0	4	2018	3	1	0	0	1	3	1591	5	1	0	0	0	0	0	65000
13	2021	0	4	2017	3	1	0	1	0	4	1999	5	0	1	0	0	0	0	70000
14	2021	0	4	2017	3	1	0	0	1	4	1999	5	0	1	0	0	0	0	80000

중간 가공 데이터

1.수집한 데이터 에 에이전시, 모델 정보를 숫자로 변환하여 가공

2.숫자 변환

1)머신 러닝은 문자를 인식 하지 않는다->숫자로 변환 필요

2.순서-데이터 가공(2021.08.23~2021.11.05)

year	week_year	day_week	car_year	holiday_length	duration	howearly	i1	i2	class	engine_cc	seat_cnt	fuel_G	fuel_L	fuel_D	fuel_E	fuel_H	total_car	low
2021	26	3	2017	0	1	0	1	0	4	1999	5	0	1	0	0	0	0	55000
2021	26	3	2018	0	1	0	1	0	4	1999	5	0	1	0	0	0	0	65000
2021	26	3	2017	0	1	0	1	0	1	998	5	1	0	0	0	0	0	45000
2021	26	3	2017	0	1	0	1	0	5	3000	5	0	1	0	0	0	0	100000
2021	26	3	2019	0	1	0	1	0	3	1598	5	1	0	0	0	0	0	75000
2021	26	3	2020	0	1	0	1	0	3	1598	5	1	0	0	0	0	0	75000
2021	26	3	2019	0	1	0	1	0	1	998	5	1	0	0	0	0	0	55000
2021	26	3	2017	0	1	0	1	0	3	1591	5	1	0	0	0	0	0	50000
2021	26	3	2017	0	1	0	0	1	3	1591	5	1	0	0	0	0	0	60000
2021	26	3	2018	0	1	0	1	0	3	1591	5	1	0	0	0	0	0	50000
2021	26	3	2018	0	1	0	0	1	3	1591	5	1	0	0	0	0	0	60000
2021	26	3	2018	0	1	0	0	1	4	1999	5	0	1	0	0	0	0	75000
2021	26	3	2017	0	1	0	0	1	4	1999	5	0	1	0	0	0	0	65000

1. 훈련 데이터

1)하나의 행에 18개의 공통적인 컬럼 동일한 데이터 값 묶기 -> 최저가(low)가 구하기

2. 18개의 컬럼

Year, week_Year, day_week, car_Year, holiday_length, duration, howearly, i1, i2, class, engine_cc, seat_cnt, fuel_G, fuel_L, fuel_D, fuel_E, fuel_H, total_car

2.순서-데이터 가공(2021.08.23~2021.11.05)

1.데이터 재 가공

- 1)전기차 의 engine_cc 의 0값 수정 후 데이터 베이스에 저장
- 2)렌터카 업체 차량보유 대수 데이터 수정 후 데이터 베이스에 저장
- 3)수정한 정보를 사용해 데이터 재가공 작업 진행

2.데이터

- 훈련 데이터:2021.01~2021.05(5개월)
- 테스트 데이터:2021.06(1개월)

2.순서- 훈련 및 평가(2021.11.06~2021.12.06)

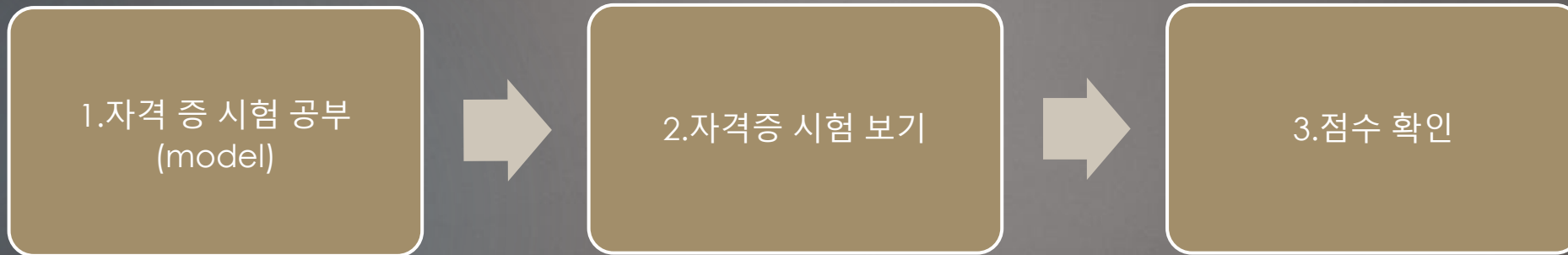
1.훈련 모델(model)정의

시험 공부를 하는 행위를 훈련 모델 이라고 한다

2.훈련 자동화 프로그램 구축

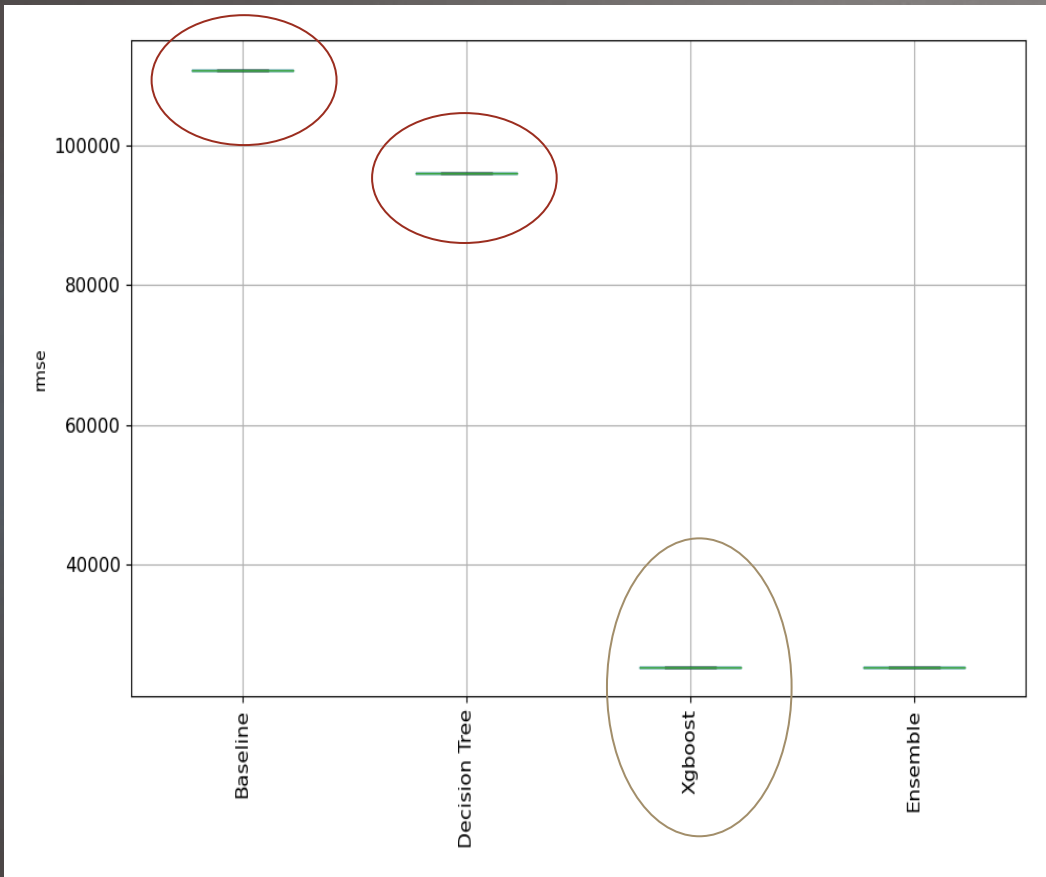
1) 사용 모델:Lightgbm, xgboost, GradientBoostingRegressor, mljar(automl)

4.모델 학습 과정



2.순서- 훈련 및 평가(2021.11.06~2021.12.06)

모델 종류	훈련 데이터 점수 (MAPE)
automl(mljar) 모델:XGBOOST	점수:12.48370113094797 해석:100-12=88점

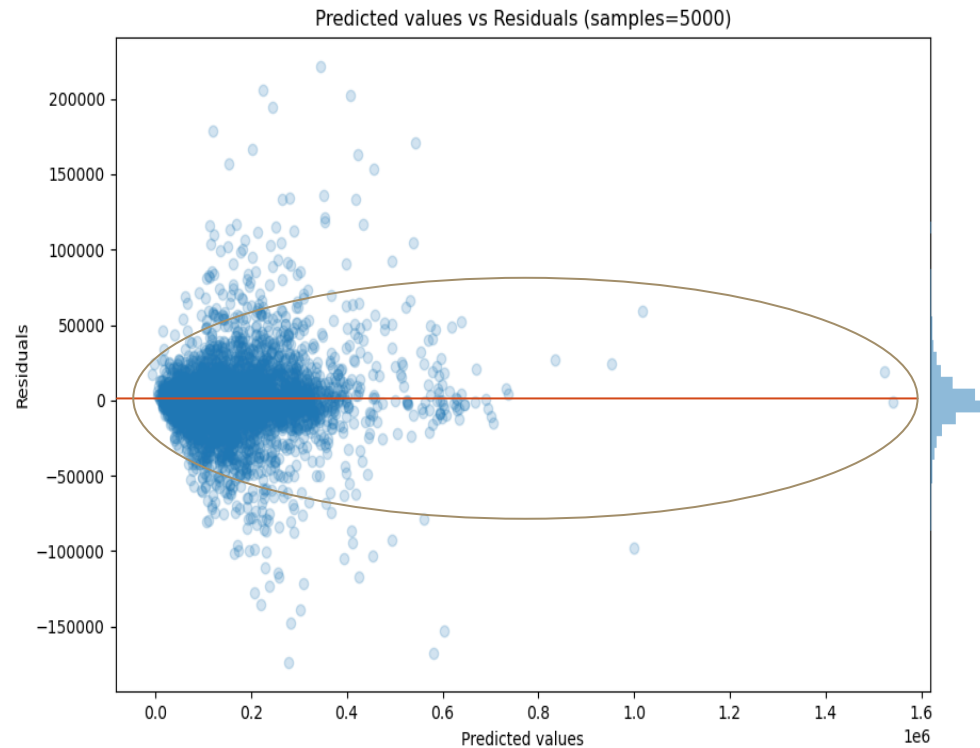


훈련 데이터 최종 결과
MSE:0에 가까울수록 좋은 모델
이다

공식: $MSE = (\text{실 제 값} - \text{예 측 값})^2$
예측력이 가장 좋은 모델

- 1) Xgboost
-> 모델 생성
- 예측력이 안 좋은 모델
- 1) Decision Tree
- 2) Baseline

2.순서- 훈련 및 평가(2021.11.06~2021.12.06)



훈련 데이터 정확도 평가 그래프 출력
1) 제일 좋은 산 점도는 0 주변에
점이 모여 있는 그림이 정확도가
좋다고 판단 할 수 있다
->예측력이 높은 모델이다

2.순서- 모델 생성 및 평가(2021.11.06~2021.12.06)

모델 종류	훈련 데이터 점수 (MAPE)	테스트 데이터 점수 (MAPE)
automl(mljar) 모델:XGBOOST	점수:12.48370113094797 해석:100-12=88점	점수:24.625624613468943 해석:100-24=76점

1.파일 생성

```
> AutoML_1  
> AutoML_classifier
```



2.테스트 데이터에 정확도 평가

24.625624613468943